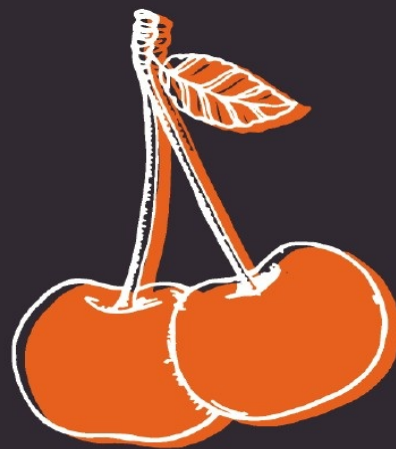


JOSÉ N. ALENCAR

# MANUAL DE MEDICINA BASEADA EM EVIDÊNCIAS

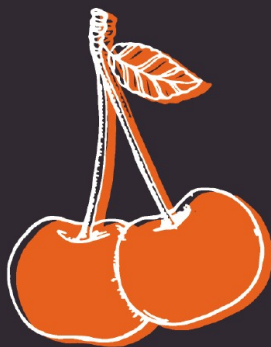
COMO INTERPRETAR  
ARTIGOS CIENTÍFICOS?



JOSÉ N. ALENCAR

# MANUAL DE MEDICINA BASEADA EM EVIDÊNCIAS

COMO INTERPRETAR  
ARTIGOS CIENTÍFICOS?



 sanar

2021

Título: Manual de Medicina Baseada em Evidências

Editor: Guilherme Melo

Projeto gráfico: Bruno Brum

Diagramação: caixadedesign.com

Capa: Mateus Machado

Revisão: Bernardo Machado e Lindsay Viola

Conselho Editorial: Caio Vinicius Menezes Nunes, Paulo Costa Lima, Sandra de Quadros Uzêda e Silvio José Albergaria Da Silva

Dados Internacionais de Catalogação na Publicação (CIP)

Tuxped Serviços Editoriais (São Paulo-SP)

---

A368m

Alencar Neto, José Nunes de (org.).

Manual de Medicina Baseada em Evidências / Organizador: José Nunes de Alencar Neto.– 1. ed.–  
Salvador, BA : Editora Sanar, 2021.

416 p.; 16x23 cm.

Inclui bibliografia.

ISBN 978-65-89822-13-4.

1. Evidências. 2. Manual. 3. Medicina. I. Título. II. Assunto. III. Neto, José Nunes de Alencar.

CDD 610.3

CDU 61(035)

---

#### ÍNDICE PARA CATÁLOGO SISTEMÁTICO

1. Dicionários, manuais e enciclopédias de medicina.

2. Manual de medicina.

#### REFERÊNCIA BIBLIOGRÁFICA

ALENCAR NETO, José Nunes de (org). **Manual de Medicina Baseada em Evidências**. 1. ed. Salvador, BA: Editora Sanar, 2021.

Ficha catalográfica elaborada pelo bibliotecário Pedro Anizio Gomes CRB-8 8846



**Editora Sanar Ltda.**

Rua Alceu Amoroso Lima, 172

Caminho das Árvores

Edf. Salvador Office e Pool, 3º andar.

CEP: 41820-770 – Salvador/BA

Telefone: 0800 337 6262

sanarsaude.com

atendimento@sanar.com

## Como ler este livro?

Este livro foi pensado para ser um manual de leitura de evidência científica para estudantes de Medicina, médicos, jornalistas e até mesmo cidadãos comuns. Ele está organizado em duas seções: na primeira, chamada “Fundamentos”, demonstramos as bases da Medicina Baseada em Evidências (MBE).

O que é MBE e por que existe? Quais as vantagens e as desvantagens de atuar nesse formato? Onde encontrar as melhores evidências? Por que há estudos bons e estudos sem força estatística? Como interpretar aqueles números? Como interpretar um estudo diante do que você já conhecia sobre aquele assunto? Como interpretar um exame diante do paciente?

Os capítulos 1 a 7, portanto, têm o objetivo de ensinar a interpretação de estudos científicos nas áreas de saúde a ponto de deixar o leitor *independente* na interpretação desses dados científicos. Baseado no raciocínio bayesiano (leia o livro para entender), esperamos que o leitor nunca mais seja enganado por terapias ilusórias. O capítulo 8 fala sobre como raciocinar baseado em evidências pode lhe ajudar a atender um paciente e interpretar os seus exames.

A segunda seção, “Especialidades em Evidência”, é um minicompêndio de evidências em diferentes especialidades médicas e até mesmo em jornalismo. Nessa seção, eu solicitei que os autores escrevessem de modo a convencer você, leitor, de que há boas ou más evidências naquela especialidade. O leitor, já munido do conhecimento necessário para interpretar aquilo que está escrito, decidirá se os autores fizeram ou não um bom trabalho nesse sentido, além de ter uma chance de analisar criticamente várias evidências em várias áreas.

Sendo assim, o livro serve a muitos tipos de leitores. A seguir, deixo um guia de leitura para cada um.

### Estudantes de Medicina

Um estudante de Medicina precisa entender que este livro propõe uma mudança de paradigma do ensino e da prática mecanicista (ensinada na sua graduação e cobrada na sua prova de residência) para a prática baseada em evidências científicas. Neste livro, ele compreenderá que as provas de imunologia e farmacologia são importantes, mas não mais importantes do que as pesquisas clínicas populacionais feitas sobre cada assunto daqueles. Sem as pesquisas clínicas, o conhecimento mecanicista não é suficiente para definir condutas, por mais sentido que aquilo fizer quando o seu professor ensinar.

É seu papel mudar a realidade da Medicina do Brasil. Precisamos muito nos desvencilhar desse modelo que leva ao paternalismo médico, que considero muito danoso em nossa sociedade.

Por isso, ler todos os capítulos da seção “fundamentos” é da maior importância. Não se assuste com o capítulo de estatística. Ao contrário do que a cadeira de bioestatística da sua graduação faz, não obrigaremos você a decorar fórmulas, mas a entendê-las.



## Médicos

O médico deve sempre buscar o melhor para os seus pacientes, mas esse melhor deve ser “*comprovadamente melhor*” e não “melhor baseado no que você observa em seus pacientes”. O médico precisa entender que enquanto não houver comprovação científica, também não há maneira de se provar que aquela terapia não é, na verdade, danosa. A consequência não intencional da iliteracia médica é gastarmos energia e dinheiro com ilusões e mentiras. A pandemia de Covid-19 deixou isso escancarado.

Por isso, a leitura de todos os capítulos da seção “Fundamentos” é importante. A seção “Especialidades em Evidências” pode ser um bom pontapé inicial para você conhecer um pouco mais sobre como se faz pensamento crítico em diversas especialidades. Então convido você a ler mesmo aquelas que não sejam a sua.



## Leigos com vontade de aprender

A leitura utilizada neste livro foi cuidadosamente pensada para que seja compreendida por todos. Eu ficarei muito feliz quando receber feedbacks de que leigos (pessoas que não entendem sobre evidências e estatística em Medicina) compreenderam bem este livro. Os capítulos da seção “especialidades em evidência” e o capítulo 16 (Jornalismo baseado em evidências) também são bastante interessantes para a sua erudição.



## Médicos com uma boa base em Medicina baseada em evidências

Este livro possui uma música de fundo que entremeia todos os capítulos: é o raciocínio bayesiano. Este modo de pensar é bastante explorado nos capítulos deste livro, particularmente no seu capítulo específico (o capítulo 5). A leitura deste capítulo pode ser muito interessante para ser o pontapé inicial de uma nova forma de ver o mundo. O capítulo 3 traz uma boa atualização sobre a clássica pirâmide da MBE – que agora é fluida; e os capítulos 6 e 7 trazem a estatística e a interpretação proposta por mim nesse método bayesiano.



## Jornalistas

A divulgação de ciências é muito desafiadora e ninguém melhor que uma jornalista para falar das nuances do seu trabalho. A didática é importante, mas se, pela simplificação didática, perdermos precisão dos dados noticiados, isso pode acabar se transformando em uma *fake news* ou *fake science*. Importante: mesmo que a intenção do jornalista não seja essa.

É por isso que eu penso que o jornalista precisa entender os fundamentos da interpretação de estudos científicos e também o pensamento bayesiano. Eu sei que é interessante publicar uma manchete falando que “ovo faz mal, diz estudo”. Mas, com o pensamento bayesiano, você vai entender que essa notícia é, simplesmente, uma “não notícia” ou uma *fake news*, mesmo que o estudo exista, esteja publicado em uma boa revista e cuja estatística seja convincente.

Recomendo, portanto, a leitura da seção “Fundamentos”, na qual você vai se munir das armas necessárias para interpretar melhor essas evidências, especialmente o capítulo 5 (o que trata sobre essa visão probabilística do mundo) e do brilhante capítulo 16.



**Conheça os trabalhos de José Alencar com a Sanar:**



[linktr.ee/josenalencar](https://linktr.ee/josenalencar)



## Autor e Editor Chefe

### José Nunes de Alencar Neto

Cardiologista. Eletrofisiologista intervencionista pela Universidade Federal de São Paulo (UNIFESP) e Hospital de Santa Cruz (Lisboa, Portugal). Assistente do serviço de Pronto Socorro do Instituto Dante Pazzanese de Cardiologia. Coordenador da Medicina de Emergência da Sanar.

## AUTORES COLABORADORES

### Arn Migowski

Especialização em Saúde Pública (IESC-UFRJ), mestrado em Epidemiologia (IMS-UERJ) e doutorado em Saúde Pública (IMS-UERJ). Prática Clínica Baseada em Evidências pela McMaster University/PROCEP. Consultor da International Agency for Research on Cancer (IARC) da Organização Mundial de Saúde (OMS). Chefe da Divisão de Detecção Precoce de Câncer e Apoio à Organização de Rede do Instituto Nacional de Câncer (INCA).

### Bruno Halpern

Doutor em Ciências Médicas pela Universidade de São Paulo (USP), Especialista em Endocrinologia e Clínica Médica pela USP.

### Bruno Robalinho C. Barbosa

Cardiologista pelo Hospital Agamenon Magalhães (HAM/SUS-PE). Cardiologista intervencionista pela Universidade Federal de São Paulo (UNIFESP). Doutorado em Cardiologia (em curso) pelo InCor/FMUSP/UFPB.

### Cristina van Blarcum de Graaff Mello

Residente de Dermatologia pela Faculdade de Medicina do ABC (FMABC). Especialista em Clínica Médica pela FMABC.

### Edmond Raymond Le Campion

Doutor em Cirurgia pela Faculdade de Medicina da USP (FMUSP). Professor Adjunto de Cirurgia Geral da Universidade Federal de Goiás. Membro Titular do Colégio Brasileiro de Cirurgiões.

## Felipe Nogueira

Doutor em Ciências Médicas pela Universidade do Estado do Rio de Janeiro (UERJ), Mestre e Bacharel em Informática pela PUC-Rio. Divulgador de ciência com contribuições nas revistas Skeptical Inquirer, Skeptic, e Revista Questão de Ciência.

## Guilherme Augusto Teodoro Athayde

Cardiologista pelo Instituto Dante Pazzanese de Cardiologia (SP). Especialista em Arritmias Cardíacas pelo InCor/FMUSP. Doutorado em Cardiologia (em curso) pelo InCor-FMUSP.

## Jonatas Brito de Alencar Neto

Médico Ortopedista especialista em Cirurgia do Joelho e Trauma Ortopedico. Mestre e Doutor em Ortopedia pela Universidade Federal do Ceará (UFC).

## Josikwylkson Costa Brito

Estudante de medicina no Centro Universitário UniFacisa (Campina Grande - PB). CPO da EBM Academy. Integrante e colaborador do grupo Stars, filiado ao Choosing Wisely Brasil.

## Luiza Caires

Jornalista e mestre em Comunicação pela USP. Editora de Ciências do Jornal da USP.

## Márcio Bezerra Gadelha Lopes

Ortopedista pelo Hospital Geral de Fortaleza (HGF), Cirurgião de joelho pela Universidade Federal de São Paulo (UNIFESP), Médico do Esporte pelo Centro de Estudo em Fisiologia do Exercício e Treinamento (São Paulo-SP).

## Otavio Clark

Médico Oncologista pela Universidade Federal de Uberlândia. Doutorado em Medicina pela UNICAMP. Vice-presidente do serviço de Oncologia na Kantar Health, Nova Iorque.

## Paulo Ricardo Criado

Dermatologista. Mestre em Medicina pelo Instituto de Assistência Médica ao Servidor Público Estadual de S. Paulo (IAMSPE), Doutor em Ciências (Dermatologia) pela Faculdade de Medicina da USP (FMUSP). Livre-docente em Dermatologia pela FMUSP. Pesquisador Pleno do Centro Universitário da Faculdade de Medicina do ABC(FMABC).

## Thiago Bosco Mendes

Especialista em Clínica Médica e Endocrinologia pela Faculdade de Medicina da USP.

# Sumário

## Seção 1. Fundamentos

1. Por que precisamos de evidências? Onde encontrá-las? Quando não precisamos delas?
2. Medicina enviesada por evidências
3. A pirâmide fluida da Medicina Baseada em Evidências e o caminho de uma hipótese até a sua confirmação
4. Vieses em estudos científicos
5. A análise bayesiana dos estudos científicos
6. Estatística para quem não gosta de matemática
7. Como interpretar artigos sobre terapias?
8. Como interpretar um teste diagnóstico?

## Seção 2. Especialidades em Evidência

9. Cardiologia baseada em evidências
10. Endocrinologia baseada em evidências
11. Oncologia baseada em evidências
12. Rastreamento de câncer baseado em evidências
13. Cirurgia baseada em evidências
14. Dermatologia baseada em evidências
15. Ortopedia baseada em evidências
16. Jornalismo de saúde baseado em evidências

# Índice

1. Capa
2. Créditos
3. Prefácio
4. Autor e autores colaboradores
5. Como ler este livro?
6. Sumário
7. Seção 1. Fundamentos
  1. 1. Por que precisamos de evidências? Onde encontrá-las? Quando não precisamos delas?
  2. 2. Medicina enviesada por evidências
  3. 3. A pirâmide fluida da Medicina Baseada em Evidências e o caminho de uma hipótese até a sua confirmação
  4. 4. Vieses em estudos científicos

5. 5. A análise bayesiana dos estudos científicos
  6. 6. Estatística para quem não gosta de matemática
  7. 7. Como interpretar artigos sobre terapias?
  8. 8. Como interpretar um teste diagnóstico?
8. Seção 2. Especialidades em Evidência
1. 9. Cardiologia baseada em evidências
  2. 10. Endocrinologia baseada em evidências
  3. 11. Oncologia baseada em evidências
  4. 12. Rastreamento de câncer baseado em evidências
  5. 13. Cirurgia baseada em evidências
  6. 14. Dermatologia baseada em evidências
  7. 15. Ortopedia baseada em evidências
  8. 16. Jornalismo de saúde baseado em evidências

# PREFÁCIO

O Brasil está preparado para praticar Medicina baseada em evidências? A pergunta não é minha, mas do próprio autor deste livro, o médico e professor José Alencar, em uma postagem no Twitter.

A história da Medicina moderna no Brasil deixa bastante a desejar no que diz respeito às evidências. A inclusão – e aceite incontestado da população e da classe médica e científica – de práticas alternativas no Sistema Único de Saúde demonstra uma falta preocupante de pensamento crítico e científico da sociedade. O reconhecimento da homeopatia e da acupuntura como modalidades médicas, com o endosso do Conselho Federal de Medicina, é algo visto em poucos países do mundo ocidental. As práticas existem, claro. Mas existir como especialização médica, com ensino nas escolas de Medicina, e praticada por médicos, é algo bastante particular e não trivial.

A homeopatia, por exemplo, já foi banida da rede pública de saúde da Austrália e do Reino Unido, regulamentada pelo direito do consumidor nos Estados Unidos (a bula deve obrigatoriamente trazer a informação de que o produto não é respaldado em estudos científicos), deixou de ser reembolsada pelo sistema de saúde francês. Espanha e Alemanha lutam para banir a prática em seus países também.

Isso porque o conjunto de evidências que demonstra que a homeopatia não funciona além de um placebo é sólido e consistente. Não há necessidade de mais estudos. Temos revisões sistemáticas e meta-análises robustas que apontam sempre o mesmo resultado: homeopatia não funciona. E nem poderia funcionar, pois não possui nem plausibilidade biológica para isso. Para que a homeopatia funcionasse, as leis da química e da física teriam de estar erradas.

Ao deparar-se com esse parágrafo, talvez o leitor indague: “ok, mas como eu posso saber de tudo isso? Não entendi metade do que está escrito ali. Conheço vários médicos que juram que a homeopatia funciona, que receitam homeopatia, então como eu vou saber o que funciona e o que não funciona?”. Pior: “como eu, jovem estudante de Medicina ou médico, vou ter esse pensamento crítico, se isso faz parte da minha grade curricular? Como vou ler estas revisões e análises, cheias de cálculos esquisitos?”

A resposta é simples: leia o livro de José Alencar e colaboradores. Porque aqui, em cada capítulo, os autores explicam tudo o que você precisa saber para entender o parágrafo sobre homeopatia e outras terapias ineficazes.

Navegando pelo livro, você vai aprender os princípios básicos da Medicina Baseada em Evidências. Vai entender como se testam medicamentos e vacinas. Vai aprender como se faz uma comparação justa, com ensaios clínicos randomizados, e por que eles precisam ser duplos-cegos, e com grupo placebo. O efeito placebo será explicado e você vai perceber por que ele, juntamente com outros fatores de confusão, é um dos melhores amigos da Medicina alternativa e da ciência malfeita.

Vai estudar sobre os diferentes tipos de estudos clínicos, e vai descobrir como somos facilmente enganados por nossos próprios vieses. Para isso, você vai precisar entender o que são vieses de pesquisa, mas não se preocupe, tem tudo isso no livro também.

Para não se deixar enganar facilmente, você também vai precisar entender sobre estatística. Os autores navegam tão bem por esse universo, que mesmo que você seja um daqueles jovens médicos avessos a números, vai ficar espantado ao perceber que desta vez entendeu tudo. Correlações, causas, probabilidade e risco farão parte do seu vocabulário.

Talvez você fique frustrado ao descobrir que muitas das coisas que você achava que funcionavam e fundamentavam a sua prática médica, na verdade, não possuem evidências. Mas essa frustração será logo substituída pela constatação de que a ciência é um processo e que a maneira como esse processo gera evidências é tão fascinante, que o encantamento fará com que você entenda que pensar de forma crítica é um aprendizado, não é intuitivo, mas é extremamente divertido e libertador.

Lembrar do tempo em que você também acreditou em evidências anedóticas e correlações espúrias vai fazer com que você seja um médico didático e paciente ao explicar aos seus pacientes por que não está prescrevendo esse ou aquele tratamento da moda.

E isso vale não só para os médicos, mas para todos os profissionais de saúde e, por que não, para todos os cidadãos. Nosso cotidiano é, afinal, repleto de pequenas escolhas que dependem do que entendemos de ciência, e quando se trata de escolhas em saúde e Medicina, erros e crenças em medicamentos e tratamentos esdrúxulos podem custar vidas.

Jornalistas de saúde também devem aproveitar este livro. Uma notícia sobre saúde, afinal, também precisa ser baseada em evidências, como o último capítulo aponta com clareza, funcionando inclusive como um minimanual de boas práticas do jornalismo de saúde, que certamente vai ajudar muito os jovens em começo de carreira.

A pandemia de Covid-19 trouxe à tona diversos problemas do Brasil. Escancarou a desigualdade social, a corrupção na política, o desmonte da ciência e da educação. O mais marcante, no entanto, foi perceber como falhamos no treinamento dos nossos profissionais de saúde. Não é por acaso que temos tantos médicos prescrevendo medicamentos inúteis ou mesmo perigosos, com a “certeza” de que estão fazendo o melhor para seus pacientes.

Mas se ensinamos pseudociência nas escolas de Medicina, de Saúde Pública, de Nutrição, de Enfermagem, como esperar que nossos jovens profissionais pensem de forma científica? Se ensinamos pensamento mágico, contrariando as leis da Química e da Física e dizendo que bolinhas de açúcar podem curar doenças, desde que a solução borrifada nelas tenha sido sacudida e diluída um número suficiente de vezes, como esperar que pensem de forma crítica?

Além de todos os problemas políticos que agravaram a pandemia no Brasil, a incapacidade de grande parte dos nossos profissionais de saúde de entender o que são evidências científicas e como elas são construídas certamente colaborou muito para nosso alto número de mortes e sofrimento.

O Brasil não está pronto para o livro do José Alencar e colaboradores. Mas Alencar e colaboradores estão prontos para lançar as fundações de uma mudança cultural e educacional necessária. Que este livro seja o primeiro de muitos. A informação de qualidade precisa estar disponível. É nosso papel como cientistas e comunicadores de ciência garantir que a informação de qualidade esteja disponível, em linguagem acessível e de fácil compreensão, e que circule. A informação de qualidade é a base de boas escolhas. Comecem escolhendo este livro.

**Natalia Pasternak**

Presidente do Instituto Questão de Ciência

## SEÇÃO 1

---

# Fundamentos

---

### **Autores**

José Nunes de Alencar Neto

Josikwykson Costa Brito

Felipe Nogueira

### **O que você irá ver neste capítulo**

- › Por que precisamos de evidências? Onde encontrá-las? Quando não precisamos delas?
  - › Medicina enviesada por evidências
- › A pirâmide fluida da Medicina Baseada em Evidências e o caminho de uma hipótese até a sua confirmação
  - › Vieses em estudos científicos
  - › A análise bayesiana dos estudos científicos
- › Estatística para quem não gosta de matemática
  - › Como interpretar artigos sobre terapias?
  - › Como interpretar um teste diagnóstico?



# 1. POR QUE PRECISAMOS DE EVIDÊNCIAS? ONDE ENCONTRÁ-LAS? QUANDO NÃO PRECISAMOS DELAS?

## Introdução

Atuar na filosofia de Medicina Baseada em Evidências (MBE) se confunde com fazer Medicina. Desde a nossa graduação, somos ensinados a entregar aos pacientes a melhor terapia disponível. A MBE só acrescenta que: “faça isso de acordo com os estudos clínicos, porque a simples observação muitas vezes não será suficiente para comprovar o benefício de uma droga”.

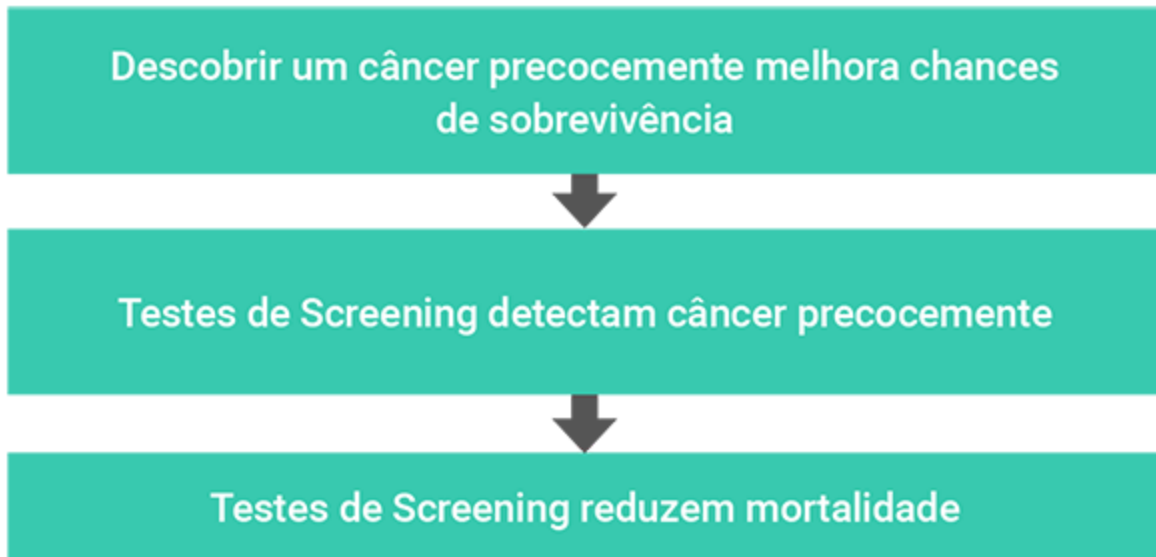
É uma questão de mudança de paradigma: na graduação, por falta de tempo hábil para aprofundar em tudo o que a Medicina tem a ensinar, somos ensinados a pensar de maneira **mecanicista**: “se um fármaco bloqueia a molécula que causava câncer, então o indivíduo não terá mais câncer” (Figura 1). É, de certa forma, chocante o momento em que um médico aprende que não é bem assim que funciona: há muito mais entre as moléculas e as sensações do paciente do que pode imaginar nossa vã Medicina. Esse tipo de lição é um daqueles *game changers*: um momento em que a ficha cai; nesse caso, para fazer o médico perceber que ele não era assim tão habilidoso e que a graduação, por melhor que seja, por falta de tempo, não lhe ensinou exatamente tudo o que ele deveria ter conhecido para entregar o melhor aos pacientes.

Dunning e Kruger descreveram, em 1999, um tipo de viés cognitivo que levou os seus nomes e que explica exatamente este “*game changer*”: o momento em que o médico se percebe ignorante.<sup>1</sup> Até ali, ele era superconfiante em suas condutas e acreditava que a Medicina era quase como uma regra, exceto pelo jargão “cada corpo responde de um jeito”. Ao conhecer a MBE, o médico é imediatamente jogado em um poço de “desesperança”, muitas vezes chegando a se ver inapto a exercer a profissão da maneira como gostaria. É a partir desse evento que o médico vai,

finalmente, fazer aquilo que tanto ouvira antes: “um médico precisa estudar pra sempre”. E, partindo desse poço, o momento mais baixo da curva de confiança, é que ele renovará suas condutas, baseado em evidências científicas (Figura 2).

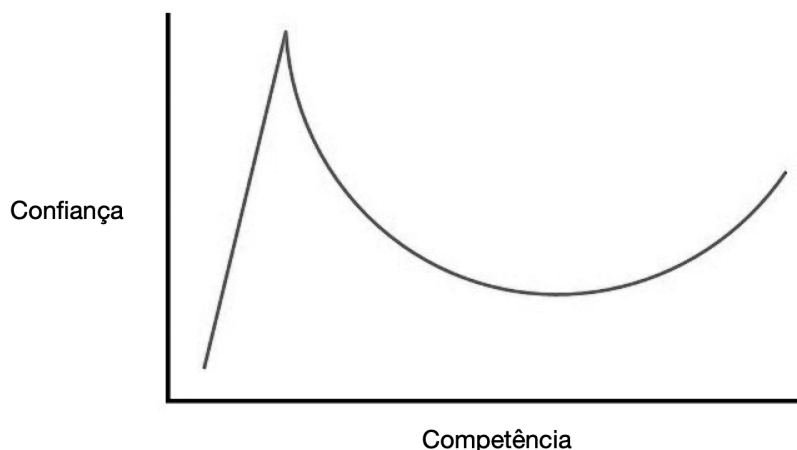
Praticar MBE nada mais é do que praticar boa Medicina.

**Figura 1.** Pensamento mecanicista ensinado na graduação de Medicina.



Pelas aulas de Farmacologia, Fisiopatologia e Anatomia, somos ensinados a pensar desta maneira: se as coisas fazem sentido, devem funcionar.

**Figura 2.** Efeito Dunning-Kruger.



O médico mais inexperiente tende a ser superconfiante. No momento em que ele depara com a Medicina Baseada em Evidências, a confiança tende a cair para um vale, o “vale do desespero”. É nesse momento que ele começa a estudar para adquirir cada vez mais competência.

## O que é medicina baseada em evidências?

Costumo dizer que William Osler, o pai da Medicina Interna moderna, já praticava MBE antes mesmo da existência dela. Para se convencer disso, basta transcrever uma de suas célebres frases: “a Medicina é uma ciência da incerteza, e a arte da probabilidade”. Vamos destrinchar essa frase?

Em primeiro lugar, Osler usa um jogo de palavras para colocar na mesma sentença duas palavras que parecem paradoxais: ciência e incerteza. Bem, elas só parecem paradoxais. Este livro vai demonstrar para você, caro leitor, que a ciência sempre foi o ato de lidar com o incerto. E isso em nada diminui o trabalho dos nossos colegas pesquisadores. Pelo contrário, o torna ainda mais fascinante. A Medicina é uma ciência de incerteza porque, pensando de maneira simplória, “cada corpo reage de maneira diferente”, e com um pensamento mais maduro, “quase nada em Medicina é absolutamente previsível”.

Em segundo lugar, Osler introduz, na mesma frase, duas palavras que são, para muitos, antagônicas: arte e probabilidade. “Arte traz prazer, já a probabilidade é uma ciência Matemática e eu fiz Medicina para não precisar usar calculadoras”, pensa o médico que ainda não desceu ao fundo do poço de Dunning-Kruger. Na prática dessa ciência, são as pesquisas populacionais que definem decisões individuais.

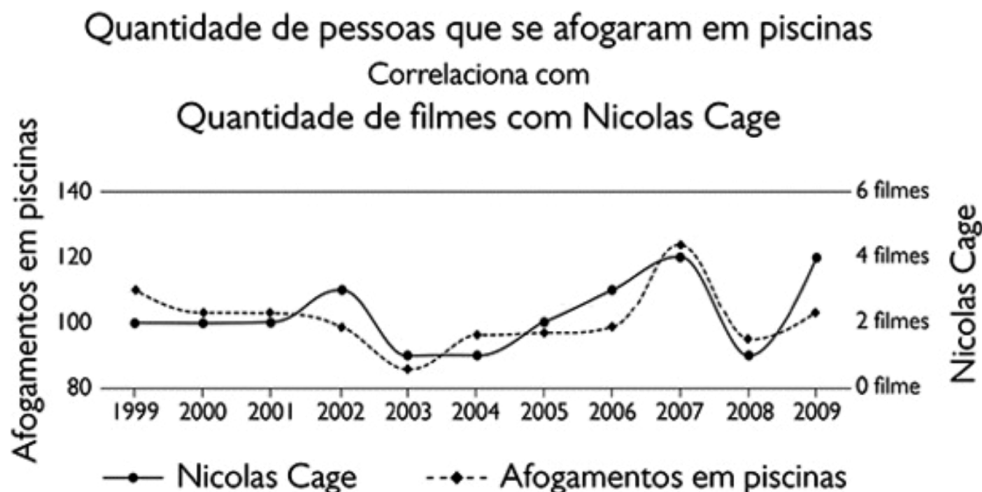
É no fundo do poço de Dunning-Kruger que o médico se depara – e pela primeira vez vê isso de maneira otimista – com o paradoxo da MBE: sendo uma ciência de incerteza, não podemos tratar nossos pacientes como números. Mas, sendo a arte da probabilidade, o médico precisará sim de um pouco (eu prometo que é só um pouco) de matemática para definir se aquele resultado de exame na mão dele é mesmo verdadeiro, ou qual a chance que aquele fármaco tem de salvar a vida do paciente.

O médico precisa aprender um pouco de matemática e estatística para interpretar corretamente os dados, as estatísticas, os gráficos a que é apresentado. Isso porque muitas vezes pode se enganar usando estatística. E, com esse conhecimento, que nem é tão complexo quanto parece (vamos ver no Capítulo 6), o médico percebe que há duas formulações lógicas que impedem que o pensamento mecanicista básico se sobressaia sobre a prática da MBE:

- a. *Post hoc, ergo propter hoc* (depois disso, portanto, por causa disso).
- b. *Cum hoc, ergo propter hoc* (com isso, portanto, por causa disso).

Essas duas falácias de lógica acontecem quando alguém pensa que só porque X aconteceu antes ou ao mesmo tempo de Y, então X levou a Y. E nós sabemos que a vida real não é assim: você tem o hábito de vestir-se antes de sair de casa para trabalhar. Mas vestir-se não é a causa pela qual você trabalha. Você trabalha porque precisa ter dinheiro para se manter. A Figura 3 demonstra um exemplo clássico desse tipo de falácia ao qual o pensamento mecanicista pode levar.

**Figura 3.** Quantidade de pessoas que se afogaram em piscinas correlaciona com quantidade de filmes de Nicolas Cage



A leitura pouco racional desse gráfico pode fazer alguns pensarem que os filmes lançados por Nicolas Cage estão relacionados com o aumento dos casos de afogamentos em piscinas. Mas isso não é plausível, é?

Praticar MBE é praticar Medicina de acordo com a correta e idônea interpretação das evidências que surgem por meio de pesquisas clínicas, porque elas saem do pensamento mecanicista para tentar encontrar correlações clínicas verdadeiras e comprováveis estatisticamente. É aprender a raciocinar criticamente sobre uma hipótese, sobre como uma pesquisa foi desenhada e sobre a veracidade daquele resultado. E depois de tudo isso, considerar os valores e as preferências do paciente, para entregar a melhor terapia disponível, causando o menor dano possível. Como definiu David Sackett, MBE é “o uso consciente, explícito e judicioso das melhores evidências atuais na tomada de decisões sobre o cuidado de pacientes individuais”.<sup>2</sup>

## Importância da medicina baseada em evidências

Não precisamos ir muito longe para imaginar como era a prática médica antes das pesquisas clínicas. Os médicos não tinham evidências para guiar o seu trabalho. A fisiologia e a farmacologia começaram a ganhar impulso maior no último século. Mas isso não é suficiente (como exemplifico mais à frente). O resultado é que os cuidados eram baseados apenas no seu próprio julgamento, na sua experiência (que pode ser enviesada, Capítulo 2) e no que ele aprendeu na universidade.

Nem sempre os médicos conseguem saber o que é melhor para seus pacientes. Por exemplo, um médico, por má formação, por crença, por

preguiça, ou por má-fé, pode achar que a seguinte frase faça sentido: “O remédio x é vendido há anos e nunca fez mal a ninguém”. Entenda, meu caro leitor, que muitas práticas consagradas da Medicina foram, depois de anos, comprovadas como danosas (Tabela 1).

**Tabela 1.** Exemplos de medicamentos/terapias que tiveram sua prática restrita com o advento de melhores evidências científicas.

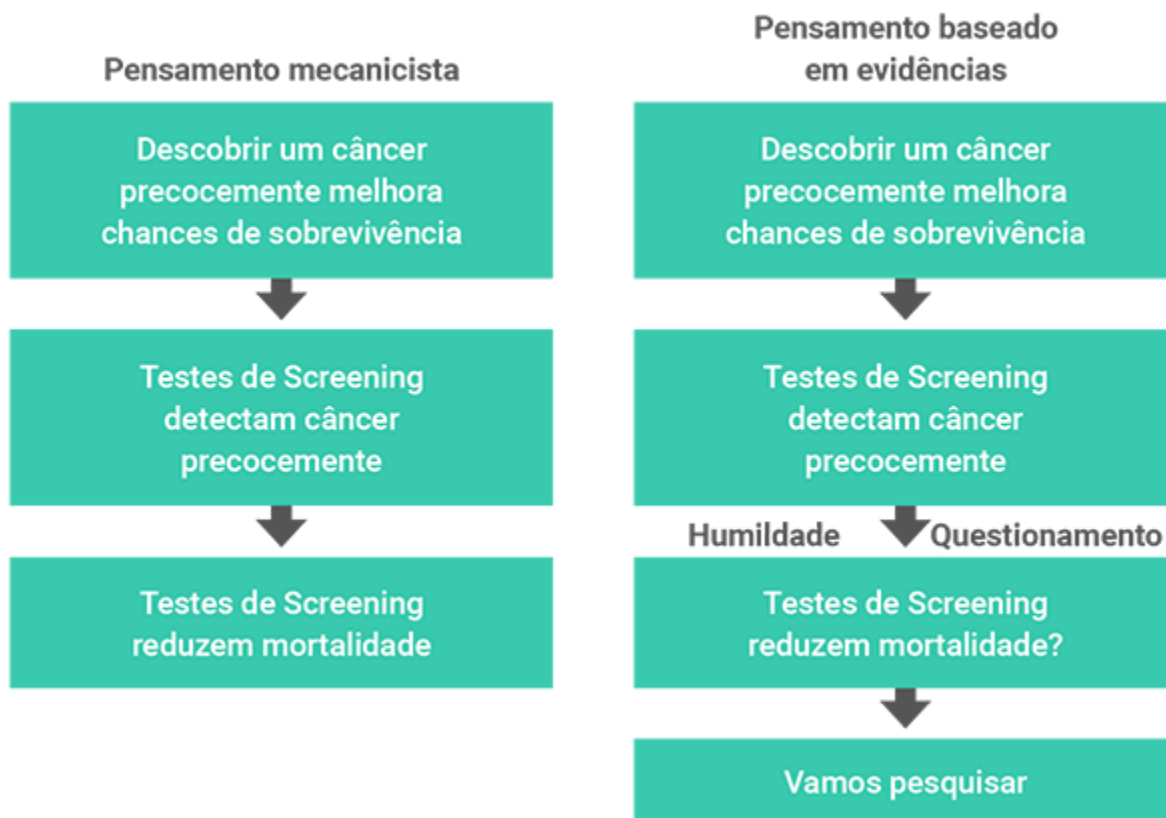
Tempo aproximado em que começou a se usar	Prática que era realizada	Tempo em que foi demonstrada que a prática causava dano	Impacto na prática clínica
Desde 500 antes de Cristo	Sangria.	1820	Interrupção do seu uso.
1957	Talidomida para enjoo em gestantes, que levou ao nascimento de mais de 8 mil bebês malformados.	1960	Interrupção do seu uso para esse nicho de pacientes. Uso restrito na atualidade.
Desde os anos 1980	Flecainida para arritmias ventriculares complexas em pacientes pós-infarto. O CAST trial mostrou que havia aumento de mortalidade entre os usuários dessa classe de drogas.	1989	Interrupção do seu uso para esse fim.
Final dos anos 1990	Inibidores da Cox-2 introduzidos para o tratamento de artrite. Depois foram associados a um maior risco de infarto e acidente vascular encefálico.	2004	Inibidores da Cox-2 tiveram sua prescrição restrita.

O exemplo mais ilustrativo de “*Medical reversal*”, que é quando um medicamento há muito usado se comprova maléfico, é o da flecainida. É um antiarrítmico potente usado em pacientes que haviam sobrevivido a um infarto e tinham arritmias ventriculares complexas. O intuito da prescrição era reduzir a chance (que, até hoje, é muito alta) de morte súbita nessa população. E fazia muito sentido que ela funcionasse, pois os médicos

conseguiam perceber facilmente que, ao prescrever flecainida, o Holter (um exame que mede a quantidade de arritmias diárias) passava a ser normal. O pensamento era: “se normaliza o Holter, o paciente deixa de ter arritmias, logo, tem menos morte súbita”. Ledo engano. Em 1989 foram lançados os resultados preliminares e em 1991 os resultados completos do CAST trial,<sup>3</sup> um estudo que demonstrou que essa classe de drogas aumentava o risco de morte súbita, apesar do Holter. Esse foi considerado como um dos maiores desastres farmacológicos da história, por ter causado dezenas de milhares de mortes apenas nos Estados Unidos.<sup>4</sup>

A transição do pensamento “mecanicista” para o pensamento “baseado em evidências” é urgente no pensamento dos médicos. Há muito mar para se navegar (Figura 4)!

**Figura 4.** Pensamento mecanicista *versus* pensamento baseado em evidências



A transição do modelo de pensamento “mecanicista” para o modelo “baseado em evidências” é urgente em todos os médicos e estudantes de Medicina.

Onde encontrar a melhor evidência científica?

Cerca de dois mil artigos são indexados ao PubMed diariamente, sendo 75 ensaios clínicos randomizados e 11 revisões sistemáticas. Nassim Nicholas Taleb sugere que quanto maior a frequência com que consultamos informações, mais desproporcionalmente propensos estamos a captar ruídos no sinal.<sup>5</sup> Ruídos são informações irrelevantes – que não nos respondem ou enganam – e são muito comuns no mundo das evidências científicas: ao contrário do que você pensava antes de abrir este livro, mais que 99% dos artigos publicados, mesmo no PubMed, são ruídos.

Esses artefatos de sinal são produzidos por:

- a. Estudos muito enviesados.
- b. Estudos sem poder confirmatório.
- c. Estudos manipulados.
- d. Estudos com desfechos irrelevantes.

E são inflados por:

- a. Conflitos de interesses de grandes empresas médicas.
- b. Conflitos de interesses de médicos.
- c. Mídia, em geral, despreparada para a interpretação desse ruído.
- d. Ganho de canetinhas e outros brindes, entregues por representantes de medicamentos.

Parece que estou sendo irônico, mas não. O ruído está em todos os lugares e polui todas as matérias existentes do conhecimento humano. O simples fato de assistir a um noticiário é suficiente para recebermos ruídos e conflitos de interesses.

Para que nos desamarremos da escravidão da irrelevância, há duas coisas que podemos fazer:

- a. Buscar evidências pré-analisadas.
- b. Aprender a diferenciar sinal de ruído.

A alternativa 1, que versa sobre evidências pré-analisadas, é a mais fácil e será tratada nos próximos parágrafos – mas ainda lhe deixa refém do “analisador”: se a pessoa ou instituição que realizou essa pré-análise está, por alguma razão, enviesada (digamos, ganhou de um laboratório de medicamentos uma passagem de avião para Cancún), também enviesada



será sua análise. A única pessoa por quem você deve botar a mão no fogo é você mesmo. A alternativa 2 é um caminho mais difícil e será ensinada até a última página deste livro, lhe entregando a independência para realizar a sua própria análise.

Buscamos evidências científicas em quatro níveis (antigamente eram três, mas a pandemia de Covid-19 trouxe um nível ainda mais baixo de evidência científica). Vejamos, do pior para o melhor:

- a. Pesquisa não revisada por pares: são pesquisas, manipuladas ou não, enviesadas ou não, que não passaram pelo crivo mínimo da revisão por pares, que é quando pesquisadores leem o artigo para identificar e corrigir possíveis vieses, ou recusar sua publicação. Pode ser encontrada em sites como medRxiv.org ou em revistas ditas “predatórias” – em que se paga uma quantia em troca de se obter uma revisão por pares inexistente ou mais amigável.
- b. Pesquisa revisada por pares, mas não pré-analisada: são pesquisas, manipuladas ou não, enviesadas ou não, que, pelo menos, passaram pelo crivo da revisão por pares em uma revista minimamente séria para estar na base de dados PubMed. É nesse repositório, o melhor que há na Medicina, que 99% dos artigos são do tipo “ruído”.
- c. Pesquisa pré-analisada: há *sites* na internet especializados em analisar as evidências para entregar ao leitor apenas o sinal relevante. Eles fazem o que este livro lhe ensinará a fazer: interpretam os artigos. Esses sites são: accesss.org, o McMaster PLUS (Figura 5), entre outros.
- d. Sumários e *guidelines*: sumários são sites que sumarizam as evidências científicas relevantes (UpToDate e DynaMed, por exemplo); e *guidelines* são diretrizes dadas por especialistas para realizar a análise das evidências existentes e separar o ruído do sinal.

**Figura 5.** Tela do site McMaster PLUS .

Ratings		
Discipline Area	Relevance to Practice	Is this News?
Gastroenterology	★★★★★★	★★★★★★

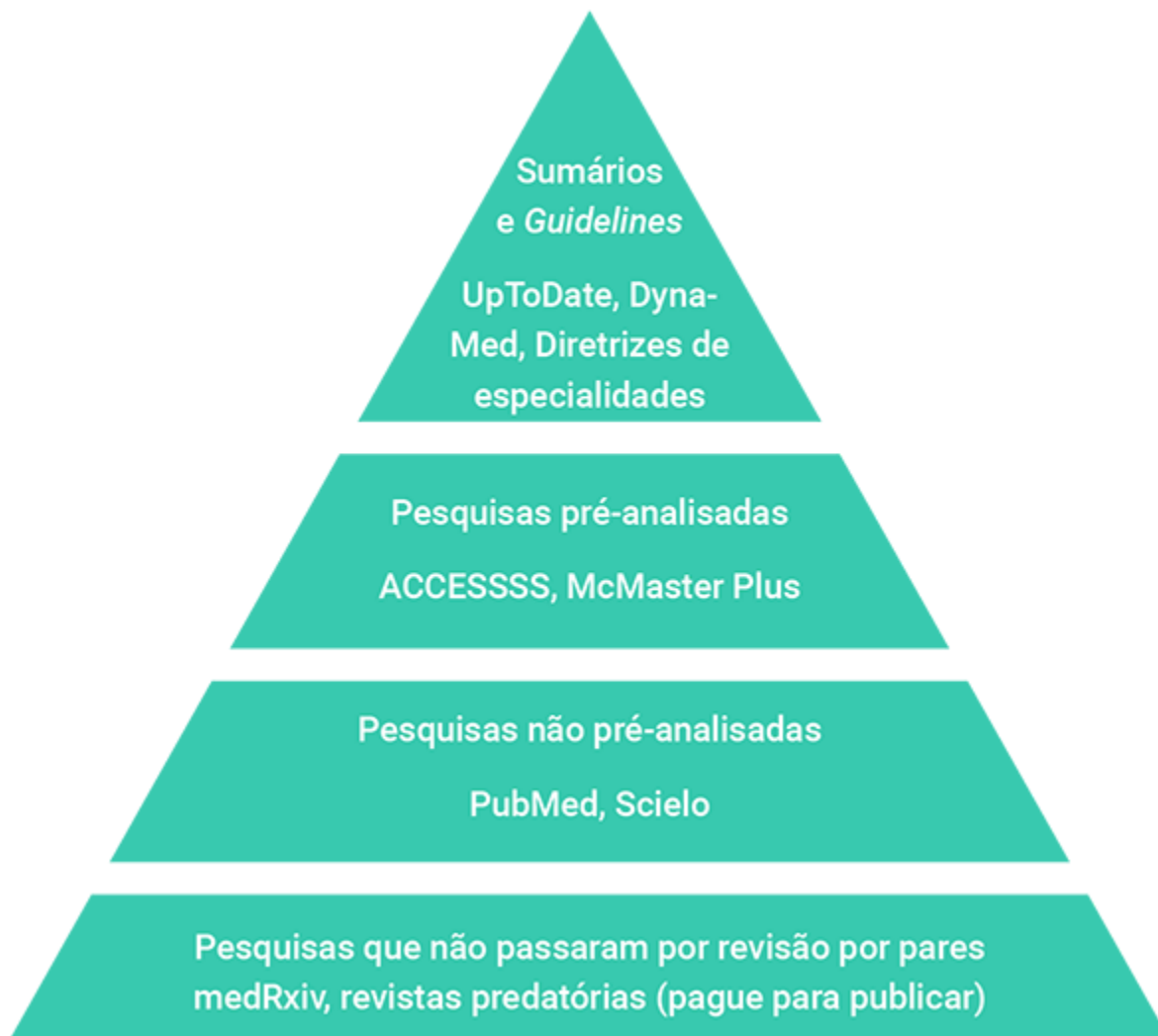
  

Comments from MORE raters
<p>Gastroenterology rater</p> <p>This is an attempted meta-analysis of PBC therapeutics. The quality of published literature limits the ability to perform the planned analysis and no meaningful conclusion is possible.</p>
<p>Gastroenterology rater</p> <p>This network meta-analysis suggests that there is no evidence that any therapy improves outcomes in PBC. This is in contrast to practitioner's experience which indicates that ursodeoxycholic acid therapy slows progression of disease and decreases the likelihood of needing liver transplantation. This meta-analysis suggests the evidence is poor, but this should not be a reason to stop using ursodeoxycholic acid, since there is apparent clinical benefit.</p>

Exemplo de uma análise realizada por especialistas do site McMaster PLUS.

É importante, antes de encerrarmos esta seção, lembrar que até mesmo os sumários e os *guidelines*, o topo da pirâmide de leitura de evidências, podem ser poluídos por ruídos caso haja manipulação ou conflitos de interesses. A Figura 6 traz essa pirâmide – não confunda com a pirâmide fluida da MBE que será vista no Capítulo 3.

**Figura 6.** Pirâmide de leitura de evidências.



Não confunda a pirâmide da leitura de evidências com a pirâmide das evidências em si. Esta traz um guia simples sobre quais os lugares mais confiáveis para se encontrar uma evidência científica de qualidade, com a menor chance de ruídos e conflitos de interesse.

## Quando não precisamos de evidências científicas?

Ao passo de que é razoável que se tenha bastante cautela ao aceitar um novo tratamento à luz das novas evidências, também deve ser assim ao interpretar evidências ditas “negativas”. Um estudo pode ser negativo por má qualidade de pesquisa.<sup>6</sup> A ausência de evidência não é a evidência de ausência. Nas palavras de Gordon Guyatt, “a evidência, forte ou fraca, nunca é suficiente para tomar decisões clínicas”.

Na verdade, todas as críticas à prática da Medicina Baseada em Evidências são provenientes do desconhecimento das suas fundações. Cito

como exemplo o argumento de que “com a introdução da pirâmide de evidências (Capítulo 3), as ciências básicas foram desvalorizadas”. Na verdade, isso é facilmente desmentido com um conhecimento básico sobre a fluidez dessa pirâmide (isto é, mesmo estudos com boas metodologias podem dar respostas enviesadas, e mesmo as metodologias mais “fracas” podem dar justamente a resposta de que precisávamos). Além disso, é importante lembrar que há diversos exemplos de como é falha a extrapolação de medidas direto do laboratório até a beira do leito sem passar pelos passos da MBE.<sup>7</sup>

Tem sido frequente a evocação de uma revisão que demonstra que 8,5% das recomendações de diretrizes norte-americanas de Cardiologia têm o nível A de recomendação (ou seja, são evidências fortes).<sup>8</sup> Apesar de isso ser um fato, ele tem sido usado para diminuir a importância da Medicina Baseada em Evidências. Ora, se não temos assim tantas evidências, significa que temos um longo caminho pela frente, não que devemos abandonar o caminho.

Há uma situação, na prática médica, em que a evidência científica não é tão necessária assim: quando a plausibilidade é grande demais. Só que isso é muito raro: poucas coisas em Medicina são muito, muito plausíveis. Por exemplo, todo o seu livro de farmacologia é plausível, mas não tanto a ponto de justificar que não realizem estudos clínicos com aquelas substâncias (lembre-se da flecainida. Ela está lá em seu livro de Farmacologia).

Plausibilidade é, em palavras simples, o quanto aquela sua ideia faz sentido. É um fator extremamente importante na MBE. Vamos ver no Capítulo 5 que a plausibilidade deve ser o ponto de partida para qualquer hipótese. Em linhas gerais, se a hipótese é muito forte, como a do exemplo que vou dar, mesmo que um estudo negativo seja lançado, ele provavelmente será falso negativo e, portanto, não deve mudar conduta.

Um exemplo de plausibilidade extrema é o uso de paraquedas em quedas de aviões. É muito plausível que um paraquedas reduza mortalidade.<sup>9</sup> Pode ser que não reduza em 100%, porque uma ou outra pessoa podem não conseguir abrir, mas reduz a chance plausivelmente. Outro exemplo, mais recente, é o do uso de máscaras durante a pandemia de Covid-19: não se faz necessário e nem é ético realizar um estudo comparando uso de máscaras versus o não uso, porque a doença é

transmitida por aerossol e o uso de máscaras possui poucos efeitos adversos importantes em comparação ao seu potencial benefício.

## Referências bibliográficas

1. Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Pers Soc Psychol.* 1999 Dec;77(6):1121-34.
2. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ [Internet].* 1996 Jan 13;312(7023):71-2 [[acesso em 8 fev. 2021](#)].
3. Echt DS, Liebson PR, Mitchell LB, Peters RW, Obias-Manno D, Barker AH, Arensberg D, Baker A, Friedman L, Greene HL. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. *N Engl J Med.* 1991 Mar; 324(12):781-8.
4. Moore TJ. *Deadly Medicine: Why Tens of Thousands of Heart Patients Died in America's Worst Drug Disaster [Internet].* New York: Simon & Schuster; 1995. [[acesso em 8 fev. 2021](#)].
5. Taleb NN. *Antifrágil [Internet].* Rio de Janeiro: Best Seller; 2015. [[acesso em 8 fev. 2021](#)].
6. Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ [Internet].* 1995 Aug 19; 311(7003): 485 [[acesso em 8 fev. 2021](#)].
7. Straus S, Haynes B, Glasziou P, Dickersin K, Guyatt G. Misunderstandings, misperceptions, and mistakes. *Evid Based Med.* 2007 Feb;12(1):2-3.
8. Fanaroff AC, Califf RM, Windecker S, Smith SCJ, Lopes RD. Levels of Evidence Supporting American College of Cardiology/American Heart Association and European Society of Cardiology Guidelines, 2008-2018. *JAMA.* 2019 Mar;321(11):1069-80.
9. Correia LC. *O Paradigma do Paraquedas. Medicina Baseada em Evidências.* 2010.

## 2. MEDICINA ENVIESADA POR EVIDÊNCIAS

### Introdução

A Medicina cresce à medida que médicos brilhantes envolvidos com pesquisa clínica e financiados por grandes corporações colaboram com o desenvolvimento de novos fármacos, terapias e exames. As empresas de biotecnologia são estimuladas a criar produtos que sejam, de fato, relevantes para a qualidade de vida e para a sobrevivência de pessoas ao redor do mundo.

Ao mesmo tempo, cresce (e não é de hoje) a preocupação com esse financiamento de pesquisas. É do interesse da empresa que o produto em que ela investiu uma boa quantia de dinheiro seja aprovado e que dê o retorno esperado. É estimado que uma nova droga custe por volta de 800 milhões de dólares para se desenvolver num processo que dura 15 anos.<sup>1</sup> Em 2016, os valores podiam chegar a 2,6 bilhões de dólares.<sup>2</sup>

É aí que nascem os conflitos de interesse. E é aí que pode se enviesar a Medicina.

### Os riscos da interação das empresas de biotecnologia com as pesquisas clínicas

Essa interação pode influenciar indevidamente o julgamento dos profissionais de saúde envolvidos na pesquisa. Essa, aliás, é a razão pela qual o cegamento dos participantes e também dos profissionais é muito importante para que se evitem vieses.

Imagine uma droga em que já foram investidas algumas centenas de milhões de dólares, agora em fase de testes clínicos. O que um pesquisador sério e sem conflitos de interesses faria? Bem, ele faria uma pesquisa de alto nível, evitando ao máximo os vieses, e buscando desfechos clínicos relevantes. Imagine que ele até começou essa pesquisa, mas o resultado foi negativo. A empresa desenvolvedora, não querendo perder todo o dinheiro investido, pode pressionar o investigador a não publicar essa pesquisa. Não

publicar pesquisas negativas é um sério problema chamado “viés de publicação”.<sup>3</sup> Em um estudo, 31% das pesquisas sobre medicamentos antidepressivos nunca foram publicados. Na literatura publicada, 91% dos estudos apresentavam resultados positivos, mas quando se avaliava a totalidade de estudos registrados, apenas 51% tinham resultados positivos.<sup>4</sup> Pior: mais da metade dos estudos clínicos de drogas recentemente aprovadas pela Food and Drug Administration (FDA) não foi publicada 5 anos após a sua aprovação.<sup>5</sup>

E aí ele começa uma nova pesquisa. Só que, dessa vez, ele não procura um nível assim tão alto: “melhor não ‘cegar’ os pacientes e os profissionais”, ele pensa. E também abre mão dos desfechos clinicamente relevantes, começando a procurar desfechos “substitutos”, como chamamos.

Desfechos substitutos serão melhor abordados no Capítulo 7. Eles são preditores ou representativos de um desfecho real e clinicamente significativo. Ou seja, é uma maneira de procurar algo mais simples, mas que represente algo importante. Por exemplo, o valor do colesterol é indicativo de um maior risco de infarto. O leitor precisa ficar atento, contudo, que alguém com colesterol elevado não necessariamente terá um infarto.

A busca por desfechos substitutos é muito vantajosa para a indústria de pesquisas porque economiza tempo e dinheiro: é mais rápido e barato avaliar o valor do colesterol do sangue de alguém do que esperar para ver se ele vai ter um infarto. Só que mais uma vez: ter colesterol elevado no sangue, apesar de denotar maior risco de infarto, não é determinista de um infarto inexorável. Também é irrelevante do ponto de vista da qualidade de vida.<sup>6</sup>

## Os riscos da interação de empresas de biotecnologia com a elaboração de guidelines (diretrizes)

As diretrizes influenciam fortemente nos cuidados ao paciente. Elas são documentos que incluem recomendações guiadas por uma revisão sistemática das evidências e da análise dos benefícios e danos de terapias.<sup>7</sup> Uma pesquisa envolvendo 498 autores de diretrizes norte-americanas de Cardiologia demonstrou que 56% dos autores reportaram algum conflito de interesses.<sup>8</sup>

Um bom exemplo desse problema é o “escândalo” com o estudo EXCEL.<sup>9</sup> Esse foi um estudo randomizado *open-label*, patrocinado pela Abbott Vascular, que tinha como objetivo demonstrar não inferioridade do procedimento de angioplastia primária percutânea (isto é, o implante de *stents*) em comparação à cirurgia de revascularização miocárdica (isto é, a confecção de pontes de mamária ou safena nas coronárias) para lesões do tronco da coronária esquerda. O resultado de não inferioridade foi atingido no estudo original e também na análise após 5 anos. Mas David Taggart, professor de cirurgia cardíaca da Universidade de Oxford e um dos autores principais do EXCEL, anunciou a retirada do seu nome do estudo e informou que as análises haviam sido manipuladas para favorecer o implante de *stents*. A diretriz europeia de Cardiologia e Cirurgia Cardíaca sobre revascularização miocárdica de 2018 deu peso igual à indicação de *stent* ou cirurgia na lesão de tronco,<sup>10</sup> fortemente influenciados pelo peso do EXCEL, mesmo que meta-análise sobre o tema tenham sido discordantes.<sup>11</sup> Após as denúncias de Taggart, a Sociedade Europeia de Cirurgia Cardiorádica também decidiu retirar seu nome da Diretriz, ficando apenas a Sociedade Europeia de Cardiologia como signatária.

## Os riscos do mau uso das evidências em benefício pessoal ou político

Não é fácil interpretar evidências científicas. A análise de artigos científicos e das evidências em conjunto sobre um determinado assunto é tão complexa, que merecem um livro só sobre isso (ainda bem que você comprou um). Vamos ver nos próximos capítulos que para se avaliar a qualidade de um artigo, precisa-se, em primeiro lugar, conhecer a qualidade da ideia (chamamos isso de “plausibilidade”), depois observar a ideia passar por diversos crivos metodológicos diferentes até chegar às metodologias com menor grau de incerteza. Caso o caminho inteiro seja percorrido com ética, humanidade e apreço à boa ciência, é aí que uma nova terapia chegará ao mercado.

Durante a pandemia de Covid-19, médicos, particularmente no Brasil, criaram uma falsa controvérsia sobre a eficácia de drogas pouco plausíveis contra o coronavírus.<sup>12-14</sup> Drogas com comprovação *in vitro*, das quais apenas uma pequena fração (10%) se torna de fato um medicamento, foram depois testadas em estudos enviesados, com desfechos substitutos,



cegamentos inadequados, paciente mal selecionados... Na verdade, foi mais parecido com um show de horrores. A internet e as redes sociais, ansiosas por um tratamento em nome de “remédio”, algo que, no início da pandemia, ainda não existia, acabaram elegendo a Hidroxicloroquina, um antimalárico e a Ivermectina, um antiparasitário, entre muitos outros, como terapia para o novo coronavírus.

O leitor, certamente bem-educado na Medicina como profissão séria e arte das mais belas que existem, percebe que é fraco o raciocínio de leigos que acham que: 1. Remédios, e não o corpo, curam doenças; 2. Medicina se resume a prescrever medicamentos e o papel dos “cuidados”, a mais bela porção da nossa arte milenar, é jogada a escanteio.

Óbvio, leigos e mesmo alguns (muitos) médicos, sem um bom conhecimento ético e de Medicina Baseada em Evidências, acabaram caindo na espiral de *fake news* e *fake sciences*, que não ajudou em absolutamente nada no combate ao vírus. Era até plausível se pensar que pessoas que acreditavam estar recebendo profilaxia antiviral com Ivermectina se achavam no direito de sair da rua sem máscara ou desrespeitar as regras de distanciamento social, pois, em seu pensamento, fortemente influenciado por parte da mídia, estava protegido. Pela Ivermectina. Sim. É bizarro. Eu sei.

Outros médicos e cientistas de várias áreas (entre matemáticos e estatísticos) também se aproveitaram do generalizado desconhecimento básico de como se faz ciência ou mesmo Medicina para ludibriar pessoas (aquelas, famintas por uma cura na imagem de um remédio). O festival de mentiras tomou as sedentas redes sociais, que passou a ser palco de conflito, às vezes entre médicos e políticos. Absurdos do tipo: “fazer estudo com placebo já é antiético, visto o tamanho benefício do uso da Hidroxicloroquina”; ou “médico que não prescreve o Kit Covid (como foi chamado) deve ser denunciado e preso”. Lembrando: até a escrita deste capítulo, esses medicamentos foram comprovadamente ineficazes contra o novo coronavírus.<sup>15</sup>

## Retracted papers – artigos despublicados

Você já parou para pensar o que acontece quando um artigo é despublicado? Será que todos os artigos que o citaram e, portanto, usaram uma informação falsa ou que não existe mais, também serão despublicados?

Não, meu caro leitor. O estrago está feito. Os artigos que o citaram continuam vivos. Para se ter ideia, algumas vezes, pode demorar décadas até que um artigo seja retirado de circulação. Fora esse problema, quanto dinheiro e tempo foi perdido numa ideia que agora não existe mais?

Um artigo pode ser despublicado quando se encontram manipulações metodológicas. Durante a pandemia de Covid-19, diversos (até a redação deste capítulo eram 33) artigos sobre o tema foram despublicados. Dois deles chamaram muito a atenção por terem sido publicados em duas das mais importantes revistas médicas do planeta: *The Lancet* e *The New England Journal of Medicine (NEJM)*. O estudo da Lancet tinha como conclusão mais importante a de que a Hidroxicloroquina aumentava a mortalidade de pacientes com Covid-19.<sup>16</sup> E o estudo do NEJM concluía que inibidores da enzima conversora de angiotensina (iECA) eram seguros para uso durante a pandemia.<sup>17</sup> Os argumentos, usados pelo autor, que é o mesmo, foram que a Surgisphere (empresa norte-americana de análise de dados e educação médica) estava se recusando a enviar aos revisores os dados completos sobre a pesquisa para uma análise mais detalhada. Até a confecção deste capítulo, a Hidroxicloroquina, apesar de não benéfica, não demonstrou malefícios claros na sua prescrição (é neutra e por isso seu uso é desaconselhado). Já os iECA tiveram estudos de moderada qualidade, demonstrando segurança em seu uso durante a pandemia,<sup>18</sup> mesmo com um mecanismo de interação fisiopatológica e farmacológica preocupante.

Pode parecer incrível que duas grandes revistas médicas tenham passado por isso. O próprio *peer-review* ou “revisão por pares” foi posto em xeque. Mas o pior acontece quando o artigo nem chega a ser publicado.

Foi o que aconteceu com a Ivermectina, o tal antiparasitário que caiu nas graças da Medicina Enviesada por Evidências. O estudo que deu o pontapé inicial a toda a *fake science* com seu uso estava hospedado em um site de *pre-prints*, que nada mais são do que artigos que não foram ainda (ou nunca serão) aceitos por nenhuma revista. É isso mesmo. Basta escrever e mandar. Bem, esse estudo foi “despublicado” (entre aspas porque nunca havia sido publicado de verdade), mas sem nenhuma nota: o artigo simplesmente desapareceu. O fantasma desse artigo vive forte na América Latina, em que países, a despeito das recomendações internacionais, chegaram a recomendar por meio do Ministério da Saúde o seu uso.

O que fazer para evitar cair nesse jogo “sujo”?

O primeiro passo você já deu: comprou (espero) um livro de Medicina Baseada em Evidências. O segundo passo é acreditar, como eu acredito, que, apesar dessas falhas, a MBE é o melhor caminho para entregar tratamentos benéficos, sem causar danos, e ao mesmo tempo estimular o avanço científico e tecnológico.

## Referências bibliográficas

1. DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug development costs. *J Health Econ.* 2003 Mar;22(2):151-85.
2. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J Health Econ* [Internet]. 2016;47:20-33 [acesso em 8 fev. 2021]. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0167629616000291>.
3. Dickersin K, Min YI. Publication bias: the problem that won't go away. *Ann N Y Acad Sci.* 1993 Dec;703:135-8.
4. Driessen E, Hollon SD, Bockting CLH, Cuijpers P, Turner EH. Does Publication Bias Inflate the Apparent Efficacy of Psychological Treatment for Major Depressive Disorder? A Systematic Review and Meta-Analysis of US National Institutes of Health-Funded Trials. *PLoS One.* 2015;10(9):e0137864.
5. Lee K, Bacchetti P, Sim I. Publication of clinical trials supporting successful new drug applications: a literature analysis. *PLoS Med.* 2008 Sep;5(9):e191.
6. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med.* 1996 Oct;125(7):605-13.
7. Kredon T, Bernhardsson S, Machingaidze S, Young T, Louw Q, Ochodo E, Grimmer K. Guide to clinical practice guidelines: the current state of play. *Int J Qual Heal Care* [Internet]. 2016 Feb 1;28(1):122-8 [acesso em 8 fev. 2021]. Disponível em: <https://doi.org/10.1093/intqhc/mzv115>.
8. Mendelson TB, Meltzer M, Campbell EG, Caplan AL, Kirkpatrick JN. Conflicts of interest in cardiovascular clinical practice guidelines. *Arch Intern Med.* 2011 Mar;171(6):577-84.
9. Stone GW, Kappetein AP, Sabik JF, Pocock SJ, Morice M-C, Puskas J, et al. Five-Year Outcomes after PCI or CABG for Left Main Coronary Disease. *N Engl J Med* [Internet]. 2019 Sep 28;381(19):1820-30 [acesso em 8 fev. 2021]. Disponível em: <https://doi.org/10.1056/NEJMoa1909406>.
10. Neumann F-J, Sousa-Uva M, Ahlsson A, Alfonso F, Banning AP, Benedetto U, et al. 2018 ESC/EACTS Guidelines on myocardial revascularization. *Eur Heart J* [Internet]. 2019 Jan 7;40(2):87-165 [acesso em 8 fev. 2021]. Disponível em: <https://doi.org/10.1093/eurheartj/ehy394>.

11. Benedetto U, Taggart DP, Sousa-Uva M, Biondi-Zoccai G, Di Franco A, Ohmes LB, et al. New-generation stents compared with coronary bypass surgery for unprotected left main disease: A word of caution. *J Thorac Cardiovasc Surg*. 2018 May;155(5):2013-2019.e16.
12. Pasternak N, Almeida LG de. Ninguém provou que hidroxicloroquina cura Covid-19. São Paulo: Instituto Questão de Ciência; 2020.
13. Dourado D. Forçar adoção da cloroquina contra COVID-19 viola a lei e a ciência. São Paulo: Instituto Questão de Ciência; 2020.
14. Pasternak N. Ivermectina é o novo bezerro de ouro da pandemia. 2020.
15. Mitjà O, Corbacho-Monné M, Ubals M, Alemany A, Suñer C, Tebé C, et al. A Cluster-Randomized Trial of Hydroxychloroquine for Prevention of Covid-19. *N Engl J Med* [Internet]. 2020 Nov 24; Available from: <https://doi.org/10.1056/NEJMoa2021801>
16. Mehra MR, Ruschitzka F, Patel AN. Retraction-Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis. *Lancet (London, England)*. 2020 Jun;395(10240):1820.
17. Mehra MR, Desai SS, Kuy S, Henry TD, Patel AN. Retraction: Cardiovascular Disease, Drug Therapy, and Mortality in Covid-19. *N Engl J Med*. DOI: 10.1056/NEJMoa2007621. *N Engl J Med* [Internet]. 2020 Jun 4;382(26):2582 [acesso em 8 fev. 2021]. Disponível em: <https://doi.org/10.1056/NEJMc2021225>.
18. Lopes RD, Macedo AVS, de Barros e Silva PGM, Moll-Bernardes RJ, Feldman A, D'Andréa Saba Arruda G, et al. Continuing versus suspending angiotensin-converting enzyme inhibitors and angiotensin receptor blockers: Impact on adverse outcomes in hospitalized patients with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) – The BRACE CORONA Trial. *Am Heart J*. 2020 Aug;226:49-59.

# 3. A PIRÂMIDE FLUIDA DA MEDICINA BASEADA EM EVIDÊNCIAS E O CAMINHO DE UMA HIPÓTESE ATÉ A SUA CONFIRMAÇÃO

## Introdução

Os níveis de evidência se referem ao grau de confiabilidade daquela informação. Tradicionalmente, quando se fala de terapia, esses graus de evidência são representados em formato de pirâmide, com os estudos clínicos randomizados e as meta-análises ou revisões sistemáticas no topo. Neste capítulo, vamos aprender a composição dessa pirâmide e, importante, perceber que ela é fluida: algumas vezes uma meta-análise pode sair do topo da pirâmide; outras vezes, um estudo observacional pode subir ao topo.

Aproveito o fato de apresentar a pirâmide para descrever os tipos de estudo em sequência: relatos de casos, pesquisas transversais, estudos de caso-controle, estudos de coorte (ou observacionais), estudos controlados randomizados (ECR), revisões sistemáticas e meta-análises.

Mas antes: será que tudo merece ser testado? Quão boa tem que ser uma ideia para que ela passe por esses testes? Vamos falar dos critérios de Bradford Hill.

## Critérios de Bradford Hill

Em 1965, Sir Austin Bradford Hill publicou nove “critérios” para determinar se associações (ideias) eram causais (não por coincidência).<sup>1</sup> Em outras palavras, se algo que observamos é coincidência ou não. Os critérios de Bradford Hill são o que eu defino como o “pontapé inicial” no nascimento de uma ideia. Quanto mais critérios uma ideia tiver, maior será a sua “probabilidade pré-teste” (em breve, mais detalhes). Repito para você entender bem: não basta ter um critério, ou mesmo os nove. Quanto mais critérios sua ideia tiver, seu time entra em campo mais forte na próxima fase.

Vamos aos critérios:<sup>2</sup>

1. Força de associação: quanto mais forte for a associação, maior a probabilidade de ela ser verdadeira. Por exemplo, quanto mais se observa que muitas pessoas com câncer de pulmão são tabagistas, maior fica a força de associação.
2. Consistência: define-se quando muitos outros estudos, usando populações e metodologias diferentes, demonstram resultados similares. Hill fala da importância disso porque um estudo positivo (dentre outros negativos) não transforma automaticamente algo em verdade.
3. Especificidade: esse critério entrou em desuso com o passar dos anos. Dizia que há maior probabilidade de uma associação verdadeira quando um fator expositor causa uma só doença. No momento em que vivemos existem milhares de sistemas complexos e uma variedade de fatores de risco.<sup>3</sup>
4. Temporalidade: por muitos, considerado o único critério *sine qua non*, isto é, obrigatório.<sup>4</sup> Para que se pense em uma associação entre um agente e a doença (ou entre uma terapia e a melhora ou cura) esse agente deve preceder a doença e a terapia deve preceder a melhora ou cura. O leitor precisa, mais uma vez, ficar atento ao fato de que esse é um critério muito sensível para que uma ideia seja considerada boa (ou seja, poucas ideias boas não respeitam isso). Mas muito pouco específica (ou seja, muitas coincidências também se comportam dessa maneira).
5. Mecanismo de dose-resposta: se o aumento da dose de exposição aumenta a resposta, então é mais provável que se trate de uma associação verdadeira. Obviamente, por diversas razões, isso não será verdadeiro todas as vezes.
6. Plausibilidade biológica: existe alguma evidência que dê suporte fisiológico ou farmacológico para a sua ideia? Atenção: a plausibilidade biológica pode não existir ainda porque não fora descoberta.
7. Coerência: similar à plausibilidade biológica, é quando essa ideia faz sentido levando em consideração todo o conhecimento produzido.
8. Experimentos: se uma ideia nasce de estudos experimentais positivos, então há maior chance de ela ser verdadeira. Perceba que o resultado de um experimento ainda é um critério de Bradford Hill.

9. Analogia: fica mais fácil convencer as pessoas da sua ideia quando uma associação já existe e está bem documentada e você quer provar algo similar.

Para entender esses critérios na prática, vamos fazer um exercício de imaginação. Imagine que você quer comprovar que um antimalárico é um bom tratamento para uma doença viral. E aí você faz um experimento para verificar se o antimalárico tem boa ação *in vitro* contra o vírus. Seu resultado foi positivo. E agora? Já posso vender o medicamento e me candidatar ao Nobel de Medicina? Não!

Com isso, você adquiriu alguns critérios de Bradford Hill e transformou sua ideia em algo plausível. Foram eles: temporalidade, experimentos, plausibilidade biológica e talvez o mecanismo de dose-resposta. Você ainda não tem: consistência, coerência e analogia.

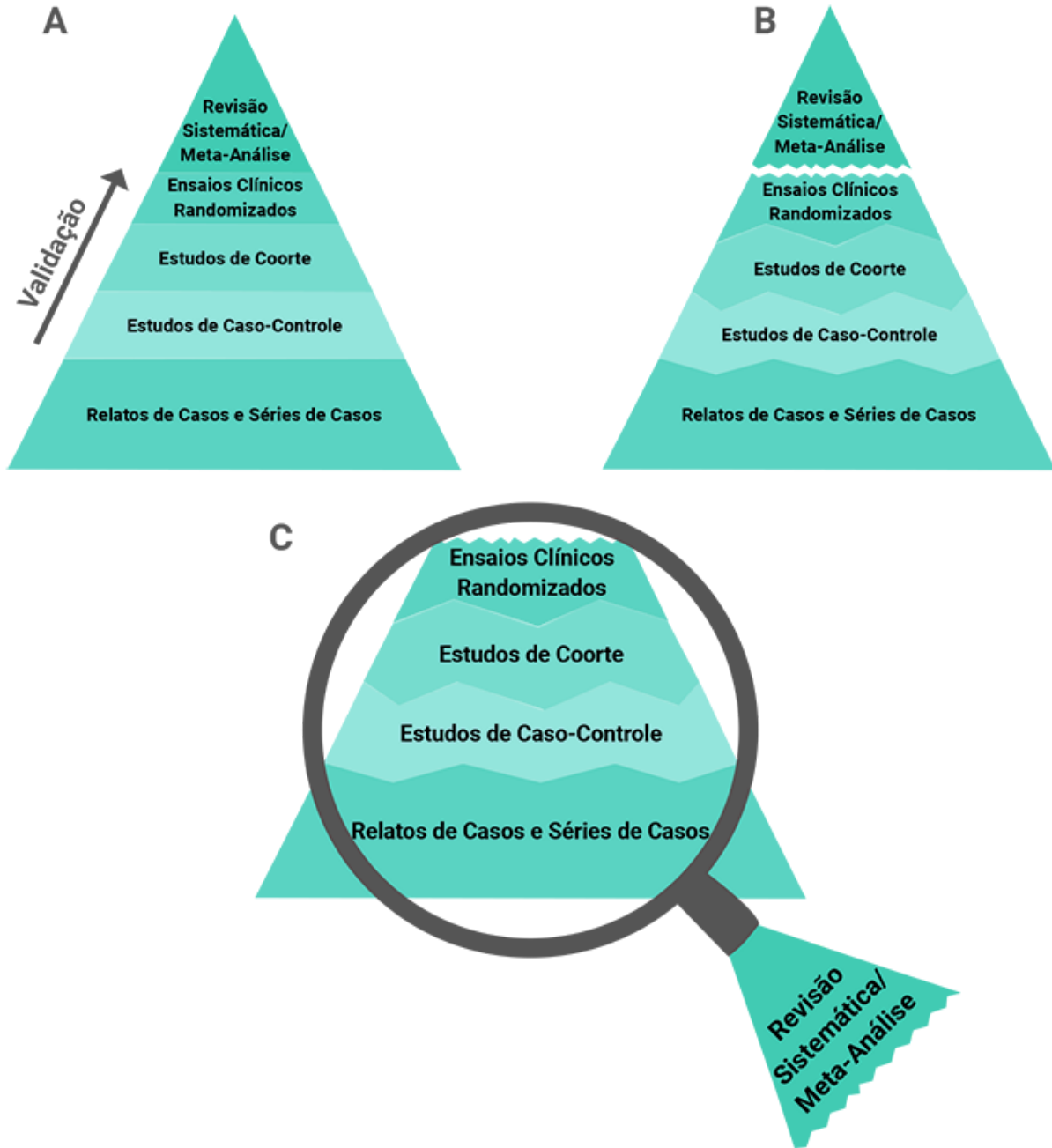
É bem verdade que apenas 12% dos medicamentos testados em estudos clínicos (ou seja, que dão um bom pontapé inicial e depois são levados aos testes mais robustos) acabam sendo comercializados,<sup>5</sup> muitos deles com benefícios apenas marginais. É de se esperar que a quantidade de drogas com boa ação *in vitro* que depois foram descartadas seja ainda menor que 12%.

Agora vamos aos tipos de estudo.

## Alerta para a fluidez da pirâmide se iniciando agora

Agora que vamos falar de cada tipo de estudo científico em Medicina, já fica o alerta para as possíveis vezes em que algo pode sair da pirâmide, ou subir e descer na sua estrutura. Nessa nova versão proposta, as meta-análises e as revisões sistemáticas são retiradas do topo da pirâmide e passam a atuar como lentes para observar a pirâmide (Figura 1).<sup>6</sup>

**Figura 1.**



Em A, observamos a tradicional pirâmide da Medicina Baseada em Evidências (MBE). Em B, observamos a nova maneira de interpretar a pirâmide da MBE: ela é fluida, de modo que estudos de coorte podem, em alguns casos, ser melhores do que estudos randomizados. Outra “novidade” é que as revisões sistemáticas e meta-análises saíram do topo da pirâmide para ocupar uma função de lupa, que observa melhor os diversos tipos de estudos (C).

## Relatos de casos



Um relato de caso descreve a história de um paciente: como ele se apresentou e como evoluiu. Isso pode ser importante para fins de educação médica ou para fazer nascer uma hipótese. Por exemplo, se uma equipe acaba diagnosticando uma rara doença, o relato de caso será interessante para que seus pares conheçam como essa doença se apresenta. Agora, em outro exemplo, se um médico publica uma série de casos demonstrando o sucesso de uma terapia que não era antes prescrita para uma determinada doença, então esse médico acaba de criar uma hipótese: a hipótese de que esse medicamento funciona.

Perceba: esse médico fez uma hipótese nascer a partir da observação clínica, e não experimental. Agora a hipótese dele possui dois critérios de Bradford Hill: temporalidade e talvez mecanismo de dose-resposta. Ele está apto a tentar fazer com que essa ideia passe adiante para estudos maiores.

Isso é importante porque muitos médicos pensam que um relato de caso ou de uma série de casos é suficiente para demonstrar evidência. Muitas vezes, em elevadores de hospital, ouvimos algum médico se vangloriar que um medicamento *off-label* acabou tendo bom benefício ao seu doente, e que a partir dali, passará a usá-lo. Outras vezes, ouvimos: “eu tratei mais de 100 pessoas com uma determinada doença, todos sobreviveram, então meu protocolo funciona”. O médico deve perceber que esse tipo de metodologia não evita vieses:

- a. E se a mortalidade da doença for de 1%? Então, na verdade, só 1 pessoa era esperada morrer nessa sua amostra. O fato de nenhum ter morrido pode ser apenas coincidência.
- b. Como os doentes não foram controlados (ou seja, o médico prescreveu o remédio associado a outros), como afirmar que foi exatamente o seu protocolo que salvou os pacientes? (viés de informação)
- c. E se você selecionou apenas os melhores pacientes porque trabalha em consultório, e os piores pacientes estavam na unidade de terapia intensiva? Ou decidiu receber em seu estudo apenas os pacientes sem comorbidades? (viés de seleção)

## Estudos ecológicos

O observado em um estudo ecológico é uma população específica, como um todo, ou uma comunidade.<sup>7</sup> Dessa observação se concluem taxas de

prevalência e incidência de doenças.

Alguns dos vieses desses estudos (razões pelas quais são apenas geradores de hipóteses) são:

- a. Falácia ecológica: não é correto inferir para um indivíduo algo que se observou para um grupo comunitário (Quadro 1).
- b. Flutuações de incidência e prevalência de doenças com o tempo. Um examinador pouco experiente pode tirar conclusões precipitadas de um fato sendo coincidente ao aumento ou diminuição da prevalência da doença. Por exemplo: casualmente a incidência de câncer de pele foi reduzida em uma população coincidentemente no ano em que se veiculou um vídeo em que um âncora famoso aconselha pessoas a usarem filtro solar.
- c. Migração: poluem a amostra por terem uma genética e uma exposição a fatores ambientais diferentes daquela população que se queria estudar originalmente. Esse é um viés e uma vantagem, ao mesmo tempo, porque o pesquisador pode se aproveitar disso para tentar criar hipóteses sobre esses mesmos fatores ambientais ausentes até certa idade (antes de migrarem).

#### Quadro 1. Émile Durkheim e a falácia ecológica

No seu livro de 1897, chamado *Le suicide*, Émile Durkheim, um sociologista francês, explora as diferenças entre taxas de suicídio entre protestantes e católicos. Observe que aqui se estudam “populações” ou “comunidades” diferentes e se comparam uma contra a outra, um clássico exemplo de estudo ecológico.

Émile concluiu que as taxas de suicídio são maiores em países com mais altas prevalências de protestantes e concluiu que o controle social mais estrito entre católicos levava a uma taxa mais baixa de suicídios. Contudo, essa conclusão não passa pelo crivo do “viés da falácia ecológica”: como nenhuma das regiões era inteiramente protestante ou inteiramente católica, era impossível saber se as maiores taxas de suicídio nesses países eram de fato vinda de protestantes ou católicos. É por isso que se diz que não se pode tirar conclusões a nível individual a partir de estudos ecológicos.

## Pesquisas transversais – observacionais

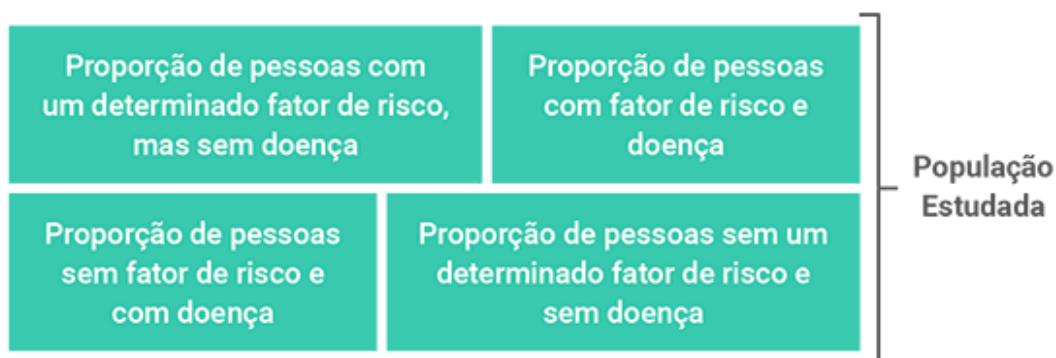
Também chamados “estudos de prevalência”. É aquela pesquisa que lhe pergunta, na praça, qual marca de arroz você prefere. Óbvio, em estudos epidemiológicos se fazem questionamentos mais rebuscados, como “Quantas pessoas são obesas na cidade de São Paulo?”. É como se esse estudo tirasse

uma foto de um grupo de indivíduos em um momento específico. Esses estudos não produzem qualquer intervenção ao indivíduo. Por isso, são também considerados estudos observacionais.

São estudos muito úteis por dar informações sobre o estado de saúde de uma população, como a prevalência de uma doença. Mas não são desenhados para testar hipóteses (Figura 2).<sup>8</sup> O médico deve perceber que esse tipo de metodologia não evita vieses:<sup>9</sup>

- a. De seleção: resulta na escolha de participantes que não representam sistematicamente aquela população que se queria estudar.
- b. De não resposta: pessoas podem se recusar a participar de uma pesquisa e elas podem ter algum motivo para isso. Se o motivo for o mesmo, um tipo específico de população (aquela que por alguma razão não quer participar do estudo) não estará representada na pesquisa.
- c. Do voluntário: similar ao viés da não resposta. Aqui, os que se voluntariam podem ter uma razão específica, razão essa que é compartilhada por um grupo para isso, e acabam sendo super-representados.
- d. De averiguação (*ascertainment bias*): quando o entrevistado não entrega a informação verdadeira na entrevista.
- e. Viés de prevalência/incidência e existência de fatores confundidores (Figura 3): como esses estudos não definem uma sequência temporal de fatos, uma associação encontrada aqui por eles pode ser uma simples coincidência. Por exemplo, se o ato de beber leite é associado, em um estudo como esses, à incidência de úlcera péptica, é porque o leite causa úlcera ou porque os pacientes tendem a beber leite para aliviar sua úlcera?

**Figura 2.**



Representação esquemática de um estudo transversal: uma fotografia da realidade de uma determinada população em termos da existência ou inexistência de fatores de risco e da existência e inexistência de doença ou desfecho.

**Figura 3.**



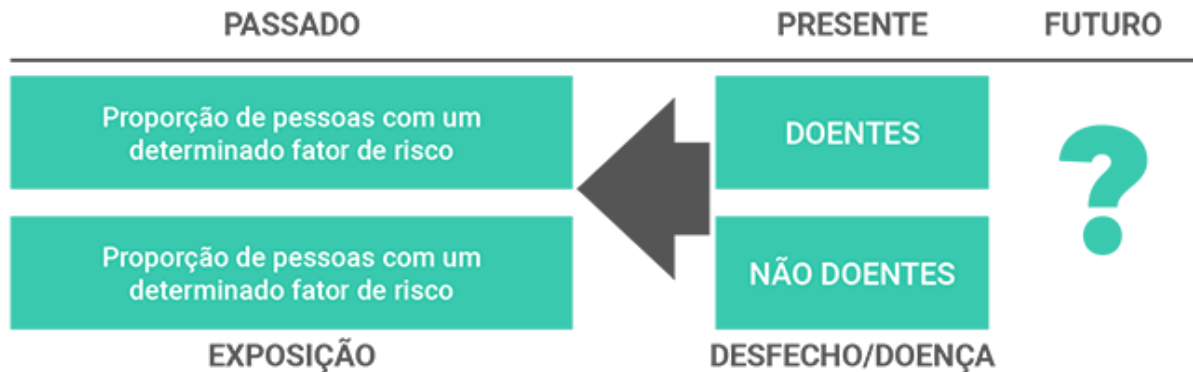
A existência dos fatores confundidores nas pesquisas observacionais (relatos e séries de casos, ecológicas, transversais, caso-controle e coorte) é a principal causa pela qual essas pesquisas são consideradas “geradores de hipóteses”. Os fatores confundidores estarão sempre presentes. Há maneiras de reduzir sua atuação indesejada, mas nem o melhor estatístico do mundo consegue neutralizar todos os fatores confundidores, até porque há fatores confundidores “desconhecidos”, aqueles que não foram descobertos ainda, por exemplo, alguma alteração genética.

## Estudo de caso-controle – observacionais

Servem para avaliar, a partir de um desfecho, se alguns fatores de risco estavam associados. O que se faz é: identificam-se pacientes com determinada doença e observam-se quais fatores de risco estavam presentes no passado desses pacientes. Depois, os pesquisadores comparam com um grupo que não tinha a doença.

Você deve ter percebido que esse estudo é de natureza retrospectiva: os pesquisadores olham para o passado dos doentes (Figura 4). Estudos retrospectivos são, comparativamente aos outros, fáceis de fazer, rápidos e baratos. São importantes, por exemplo, em doenças raras porque estudos prospectivos nessa população seriam lentos; afinal, a doença é rara e poucos doentes entrariam no estudo. O desfecho encontrado ao final dessa comparação é chamado de “*odds ratio*” (OR), ou “**razão de chances**”, em português.

**Figura 4.**



Representação esquemática de um estudo de caso-controle: observa-se o passado de pessoas doentes e não doentes, à procura de diferenças estatisticamente significativas de fatores de risco determináveis.

Um exemplo de estudo de caso-controle seria: analisar doentes com um certo tipo de câncer e também doentes sem esse câncer, pareados na mesma idade e sexo, e olhar para o passado dessas pessoas tentando identificar se houve maior exposição ao uso do contraceptivo oral.

Mas nem tudo são flores: a escolha dos doentes e dos participantes do grupo controle (sem a doença) precisa ser criteriosa. O ideal aqui é que se comparem indivíduos semelhantes: mesma quantidade de homens e mulheres nos dois grupos, idades similares, mesma classe social etc. Uma boa maneira de aumentar o poder estatístico de um estudo como esse é parear um doente com três ou quatro “controles”. Possuem vieses,<sup>10</sup> lógico.

- Recall bias* ou viés de memória: olhar para o passado é, por si só, um viés: temos tendência cognitiva de lembrar de apenas alguns acontecimentos na nossa vida, e outros não.
- Viés de seleção: selecionar pacientes constitucionalmente diferentes entre os dois grupos pode ser um problema: imagine selecionar pacientes hospitalizados em um grupo e atletas no outro.
- Questionários enviesados: é de bom tom que os questionários entregues aos participantes passem por avaliação antes de usada na pesquisa. Se o doente ou controle tiver ideia dos desfechos exatos procurados na pesquisa, ele pode ficar mais enviesado a dar respostas que não daria caso estivesse “cego”.
- Viés de prevalência e seus confundidores já citados anteriormente também podem estar presentes aqui, como na história do café como causador de câncer de pâncreas (Quadro 2).

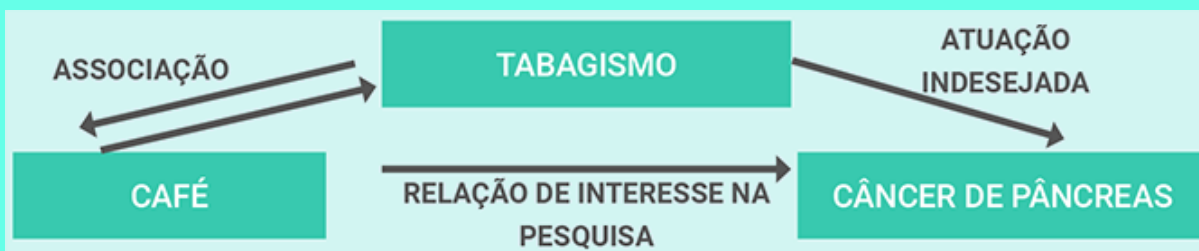
Um excelente exemplo de “fator confundidor”, com pitadas do que eu chamo de “viés da eminência” (quando um pesquisador vem de Harvard ou a publicação foi feita na revista *Lancet* ou *New England*, a pesquisa, automaticamente, ganha ares de confiável), é o do pesquisador Brian MacMahon e seu achado, publicado na *New England Journal of Medicine* (NEJM), em um estudo de caso-controle, que “o café pode ser responsável por uma proporção substancial de casos de câncer de pâncreas nos Estados Unidos”.<sup>17</sup>

Basicamente, o que MacMahon fez foi comparar 369 pacientes com câncer de pâncreas comprovado por biópsia com 644 pacientes sem câncer sobre o seu uso de cigarro, álcool, chá e café. E encontrou uma razão de chances de 1,8 (ou seja, alguém tem 1,8 vez mais chance de ter câncer de pâncreas por beber café) quando se bebiam duas ou três xícaras por dia, e 2,7 se fossem bebidas mais que três xícaras por dia, o que indica até mesmo um dos critérios de Bradford Hill, o “mecanismo de dose-resposta”.

Os problemas encontrados nesse estudo foram:

- Viés de seleção: os pacientes do grupo controle, isto é, os pacientes sem câncer do estudo de MacMahon, eram vistos pelos mesmos médicos da equipe que também internava no hospital os pacientes com câncer de pâncreas, ou seja, gastroenterologistas. Por essa razão, essa era uma população com alta incidência de doença gastrointestinal, por isso, os pacientes eram aconselhados a não beber café. Isso levou à seleção de controles com alta prevalência de doença gastrointestinal e, conseqüentemente, baixa ingestão de café.
- Existência não mensurada de fatores confundidores (Figura do quadro): a ingestão de café tende a ser maior entre indivíduos tabagistas. Como o hábito de fumar não foi mensurado nesse estudo, não há maneira de descartar que não seja o tabagismo o responsável pela maior chance de câncer de pâncreas.

**Figura do quadro:**



O tabagismo entra aqui como fator confundidor.

A lição que fica é a de que em estudos observacionais, não se pode confirmar uma hipótese, visto que os fatores confundidores estarão sempre presentes.

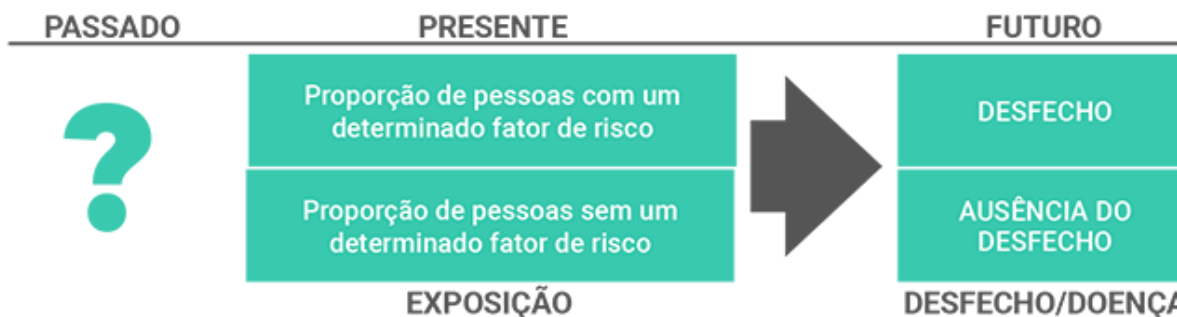
## Estudos de coorte – observacionais

Estudos desse tipo podem ser prospectivos (olham para o futuro) ou retrospectivos (olham para o passado).

Uma pesquisa como essa, prospectiva, observa um grupo (uma coorte) de pessoas no decorrer dos próximos anos. O começo do estudo define o “*baseline*” (as características de base dos indivíduos estudados) e, nesse momento, nenhum participante possui o desfecho a ser estudado (Figura 5). O estudo de coorte mais famoso é o Framingham Heart Study (Quadro 3), em

que se observaram pessoas com determinados fatores de risco (hipertensão, hipercolesterolemia...) e pessoas sem esses fatores de risco para determinar se, no futuro, as pessoas com fatores de risco teriam maior incidência de doença cardiovascular instalada, como infarto agudo do miocárdio e insuficiência cardíaca.

**Figura 5.**



Representação esquemática de um estudo de coorte: observam-se pessoas já expostas a um determinado fator de risco e também pessoas não expostas, à procura de diferenças estatisticamente significativas em desfechos no futuro.

#### **Quadro 3. Framingham Heart Study**

Motivados pelo aumento da incidência de mortes por causas cardiovasculares na primeira metade do século 20, o Serviço Público de Saúde dos Estados Unidos deu início a um estudo longitudinal prospectivo para entender a história natural da doença cardiovascular e seus fatores etiológicos. O Estudo de Framingham foi iniciado em 1948 em Framingham, Massachusetts, uma cidade com economia estável, baixas taxas de emigração e muito próxima de Boston, um forte centro médico americano. Em 1961, uma pesquisa baseada nessa coorte definiu vários fatores de risco importantes: hipertensão arterial, níveis elevados de colesterol, presença de sobrecarga ventricular esquerda ao eletrocardiograma. Atualmente, o estudo de Framingham já estuda a terceira geração de participantes e está, entre outros objetivos, focando no estudo genético.<sup>18</sup>

Os vieses de uma pesquisa como essa são:

- a. Viés de seleção.
- b. São estudos mais caros e mais longos (vide Framingham).
- c. Tamanhos amostrais precisam ser muito grandes para determinar, com menor incerteza, alguma associação naquela população.

Neste momento, alguém pode se perguntar se vale a pena existir um estudo com o seguinte desenho: vamos observar pessoas expostas a um determinado medicamento e também pessoas não expostas ao medicamento. No futuro, observaremos se há uma menor taxa de mortalidade entre os que usaram o medicamento. Assim, podemos comprovar o benefício de uma droga.

Pois é. Estudos observacionais de coorte ou mesmo caso-controle com medicamentos são comuns. Mas, como qualquer observacional já visto, não se pode retirar deles os fatores confundidores.

## Estudos observacionais – um resumo

Estudos observacionais são excelentes para avaliar o impacto de fatores de risco em determinado desfecho. Sobre terapias, contudo, eles apenas geram hipóteses, não as confirmam. E isso, algumas vezes, é tudo que se pode fazer: são estudos baratos e rápidos, em comparação com estudos randomizados e controlados. A razão para a existência dos observacionais é pavimentar o caminho para ensaios clínicos randomizados (ECRs). Perceba no exemplo hipotético demonstrado nas próximas linhas: um médico, por meio do seu pensamento e mecanismo fisiopatológico, tem a ideia de que um determinado medicamento cura uma doença crônica. Esse pensamento possui um critério de Bradford Hill, o de “plausibilidade biológica”. Uma pessoa com conhecimento limitado em Medicina Baseada em Evidências poderia pensar que a publicação de um estudo descrevendo essa tal plausibilidade, ainda mais se em uma revista de alto impacto, seria suficiente para a prescrição do medicamento.

O que essa pessoa não levou em consideração é a existência de fatores confundidores, vieses e coincidências que podem haver caso essa prescrição se inicie, como ele propôs. Em primeiro lugar, o corpo humano se importa muito pouco com sua plausibilidade biológica: as vias metabólicas são tão complexas, que nada garante que seu mecanismo, na vida real, aconteça exatamente como você pensou. Em segundo lugar, se o mecanismo real do medicamento não tiver sido previsto antes, e for maléfico, muitas vidas podem ser perdidas até a comprovação disso.

As “coincidências” em estudos observacionais vêm de diversas fontes: vieses, confundidores, má interpretação das chances.

Os confundidores, quando conhecidos, podem ser evitados pelo desenho do estudo ou pelo ajuste estatístico por estratificação, como restrição da



população estudada e pareamento, ou técnicas de análises de múltiplas variáveis. Calma, não se assuste. Vamos ver agora como isso funciona.<sup>11</sup>

Restrição de populações estudadas, também conhecida como “especificação”: os critérios de inclusão e exclusão dos estudos servem exatamente para estudar uma população menos heterogênea possível e com esses fatores confundidores já excluídos. Por exemplo, se alguém quer saber qual a chance de uma determinada condição cardíaca levar à morte nos próximos cinco anos, é interessante que se excluam pacientes com condições que também levam à morte em cinco anos, como cânceres e outras doenças em estágio terminal, além de pacientes muito idosos.

Pareamento: é a tentativa de incluir, no grupo controle e no grupo “desfecho/doença”, a mesma quantidade de pessoas com as mesmas características, exceto aquela que está sendo estudada. Isso fica progressivamente mais difícil à medida que cresce o número dessas variáveis: selecionar a mesma quantidade de homens e mulheres nos dois grupos pode ser fácil, mas imagine selecionar a mesma quantidade de hipertensos, obesos (com o mesmo percentil de IMC), diabéticos, tabagistas, idade, fração de ejeção ao ecocardiograma, nível de colesterol, valor da hemoglobina glicada, uso de betabloqueadores etc.

Estratificação da análise: é quando se estratificam pacientes de um estudo entre os que possuem as características mais parecidas.

Análise multivariada: são métodos estatísticos que ajustam diversas variáveis simultaneamente. Os métodos de análise mais usados são: modelo de risco proporcional de Cox, o modelo de regressão logística e o modelo de regressão linear. A análise multivariada é mais sujeita a encontrar associações espúrias quanto maior for o número de variáveis ajustadas. Também tem a desvantagem de não conseguir ajustar fatores confundidores não conhecidos.

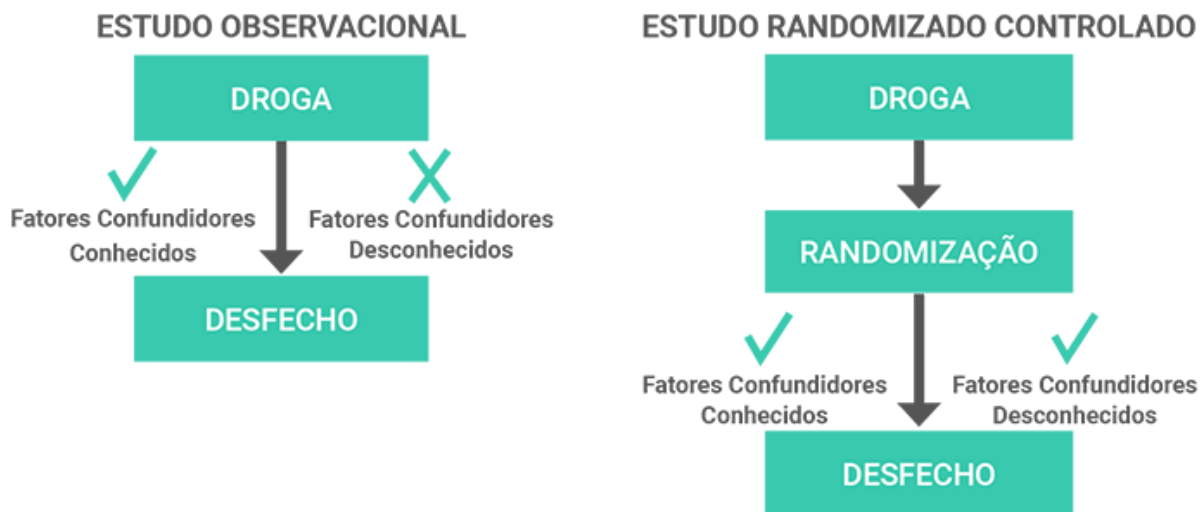
No fim das contas, percebemos que, pela existência desses vieses e fatores confundidores, especialmente os “não conhecidos”, os estudos observacionais, mais uma vez, apenas geram hipóteses.

O que o médico tem que fazer com sua ideia biologicamente plausível, então? O melhor caminho é testar essa ideia em estudos baratos primeiro, talvez uma pesquisa transversal. Caso seja positivo, dois efeitos ocorrerão: sua ideia ganhará probabilidade pré-teste (ainda não está confirmada) e ele, provavelmente, atrairá verba pública ou privada para seguir com sua pesquisa. Com essa verba e com uma probabilidade pré-teste já moderada, ele pode, quem sabe, começar um ECR, o estudo que tem potencial de confirmar hipóteses.

## Estudos clínicos randomizados – o topo da pirâmide, mas nem sempre

No teste de novas terapias, o fato de randomizar pacientes, isto é, colocá-los em um grupo de tratamento ou outro através de sorteio é a principal arma desse tipo de estudo. O sorteio reduz (mas não extingue) a probabilidade de que um fator confundidor atue de maneira indesejada na pesquisa (Figura 6 – não deixe de ler a legenda).

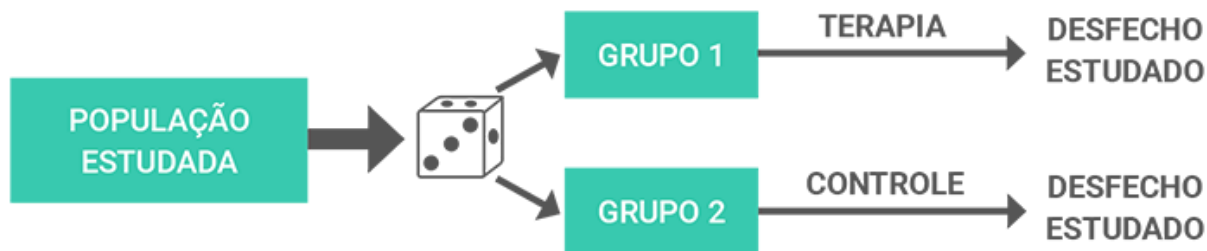
**Figura 6.**



O desenho de um estudo randomizado controlado (ECR), especialmente pela sua randomização, ou seja, a distribuição de pessoas em braços diferentes da pesquisa por meio de sorteio, reduz o efeito indesejado tanto dos fatores conhecidos como desconhecidos. Se há, em um universo de 200 participantes, quatro pessoas com alguma alteração genética que só será descoberta no futuro (portanto, na atualidade ainda é um fator confundidor desconhecido), a maior chance, por sorteio, é que fiquem dois em cada grupo, ou, pelo menos 3 x 1.

É um estudo do tipo “experimental”, em que o pesquisador intervém diretamente no curso da doença de alguns dos participantes da pesquisa (Figura 7). O ECR é capaz de **confirmar hipóteses**, caso seja bem desenhado e tenha a menor taxa de vieses possíveis. Vamos aprender a analisar os estudos randomizados controlados com calma nos Capítulos 6 e 7.

**Figura 7.**



Representação esquemática de um estudo randomizado controlado: por meio de sorteio, pacientes entram no grupo que receberá a terapia e outros entram no grupo que não receberá. Ao final do seguimento (*follow-up*), comparam-se as taxas do desfecho estudado entre os dois grupos.

Há, contudo, preocupações éticas em realizar alguns tipos de pesquisas clínicas em seres humanos. Não é aceitável, por exemplo, expor deliberadamente participantes a danos potenciais. Para reduzir o risco de exposição de participantes a danos, é que existem as já conhecidas fases dos ensaios clínicos randomizados, que são quatro (Tabela 1).<sup>12</sup>

**Tabela 1.**

Fases	Objetivo	Número típico de pacientes
Fase I	Explorar possíveis efeitos tóxicos e determinar a tolerância de uma intervenção ou de uma dose.	10 a 30
Fase II	Determinar se existe efeito terapêutico ou se há chance de que os benefícios superem os riscos.	20 a 50
Fase III	Comparar um novo tratamento à terapia habitual ou um grupo controle ou um placebo (se a terapia ainda não existe).	100 a 5000
Fase IV	Obter informação em escala e de longo prazo sobre os efeitos da droga após sua comercialização.	Muitos milhares

Fases dos ensaios clínicos.<sup>12</sup>

Outro princípio ético que deve ser respeitado é que não se pode deixar um paciente sem a terapia padrão, caso essa exista e comprovadamente tenha benefício clínico. Isto é, uma nova droga será testada contra placebo apenas no caso em que não houver nenhuma terapia comprovada para aquele mesmo fim da droga em estudo. Caso haja, o desenho do estudo deverá ser: terapia padrão + placebo *versus* terapia padrão + nova terapia. Apesar de ser óbvio, um problema emerge disso: o nível de evidências de diversos tratamentos definidos como “padrão ouro” gera debates entre entusiastas da Medicina

Baseada em Evidências sobre se aquela terapia merece realmente esse título. Caso não mereça, mas, mesmo assim, por conflitos de interesses (Capítulo 2), a terapia foi considerada como padrão ouro, então já passa a ser considerado antiético testar esta nova terapia ainda não muito bem comprovada versus placebo, o que atrasa sobremaneira a evolução do conhecimento naquele campo – visto que, nessa hipótese, estaríamos perdendo tempo com uma terapia não comprovada enquanto não estudaríamos potenciais terapias melhores. Um exemplo de um momento em que esse problema foi vencido é o ORBITA *trial* (Quadro 4).<sup>13</sup>

**Quadro 4. O ORBITA trial: stents coronarianos ou placebo para dor anginosa?**

O ORBITA *trial* foi o primeiro estudo cego em quarenta anos a comparar a revascularização com *stent* de artérias coronárias versus placebo em pacientes com doença coronária crônica. Os pacientes foram cegados do seu procedimento (não sabiam se haviam recebido *stent* ou apenas o cateterismo) e os médicos da sala de recuperação também. O resultado foi de que não houve diferença significativa na tolerância a exercícios entre os grupos.

Apesar de que a angina estável, especialmente em casos mais avançados e refratários a medicamentos, seja considerada como uma indicação clássica do implante de *stents*, essa terapêutica era baseada em estudos prévios em que os pacientes não foram cegados da sua terapia, ficando assim, mais predispostos a passar pelo efeito Hawthorne ou pelo efeito placebo.

Alguém pode argumentar de que um estudo desse tipo, quando já há uma terapia “bem estabelecida” (entre aspas porque não havia sido testado com o melhor tipo de estudos clínicos) pode levar um risco de dano ao paciente, por não oferecer a uma parte deles essa tal terapia. Ao mesmo tempo, outra pessoa pode argumentar que se essa terapia “não muito bem estabelecida” continuar sendo dada aos pacientes sem que haja uma releitura das evidências e sem a condução de ensaios clínicos mais robustos, ela é que pode levar a dano à população. Afinal, toda terapia, medicamentosa ou intervencionista, traz consigo um dano.<sup>19</sup>

Como você viu, outra característica importante desse tipo de estudo é a possibilidade de incluir uma terapia do tipo “placebo” na pesquisa, particularmente quando é uma pesquisa cega. O placebo serve para deixar pacientes e pesquisadores “cegos” do potencial de ser bem cuidado (efeito Hawthorne) ou do benefício do placebo em si (efeito placebo) nos resultados subjetivos do estudo. Nem sempre é possível cegar um estudo, nomeadamente nas especialidades cirúrgicas – é difícil simular uma cirurgia, com suas cicatrizes, pontos etc., além de ser obviamente antiético “abrir” um paciente para não entregar terapia alguma.

Resumindo, o estudo é randomizado para evitar (até o máximo possível) confundidores; e (quando possível e ético) testa placebo versus terapia, cegando os pacientes e os pesquisadores para evitar que a avaliação subjetiva de sintomas influencie no resultado final da pesquisa.<sup>14</sup> Os ensaios clínicos

randomizados, por serem a pedra fundamental do teste de novas terapias, serão bem descritos no Capítulo 7.

## Meta-análises e revisões sistemáticas

São resumos bem estruturados das evidências científicas, com objetivos claros e transparentes de avaliar uma quantidade maior de pacientes em estudos similares e para obter um melhor panorama sobre uma determinada terapia. Em outras palavras, um pesquisador analisa vários estudos observacionais e/ou randomizados sobre um mesmo tema e tenta concluir, agora olhando para uma população maior (visto que os pacientes são somados) se a terapia tem benefício.

Estudos desse tipo, como todos os outros, podem ser de boa ou má qualidade. Por isso, na fluidez da pirâmide da MBE, eles foram retirados do topo e passaram a ser usados como uma lupa, para observar com mais rigor as evidências já existentes. Os possíveis vieses encontrados são:

- a. Qualidade de estudos incluídos: as conclusões das meta-análises dependem fortemente da qualidade dos estudos incluídos. Quando se incluem estudos enviesados na pesquisa, também enviesado será o resultado da metanálise (“*garbage in, garbage out*”).
- b. Viés de publicação: ocorre porque a literatura tem uma forte tendência a publicar apenas os resultados positivos. Estudos negativos que não são publicados poderiam modificar o curso de uma metanálise.
- c. Heterogeneidade: estudos podem até ter testado uma mesma terapia, mas para fins diferentes, ou em populações diferentes (uma em pacientes muito idosos e uma em muito jovens, por exemplo – a fraca interseção entre essas populações acaba tornando improvável que essa pesquisa sirva para o mundo real fora de pesquisas).<sup>15</sup>

## Conclusões

O resumo das características mais importantes de cada tipo de estudo está disposto na Tabela 2. O mais importante desse capítulo é fazer o leitor perceber que há um caminho que uma hipótese deve fazer até a sua confirmação. Esse caminho é o seguinte:

- a. Se quer se comprovar a associação de um fator de risco a uma determinada doença ou desfecho, ela deve nascer nos critérios de Bradford Hill e evoluir para estudos observacionais.
- b. Se quer se comprovar o benefício de uma nova terapia, ela deve nascer dos critérios de Bradford Hill (quanto mais, melhor) e depois ganhar os estudos ou ensaios clínicos randomizados, desde a fase I até a fase IV, para que sua probabilidade pré-teste seja posta à prova.

**Tabela 2.**

Desenho do estudo	Características	Limitações
Relatos de casos/séries de casos	<ul style="list-style-type: none"> <li>• Relato escrito de um caso clínico e da sua evolução.</li> <li>• Podem levar à formulação de hipóteses através da observação.</li> <li>• Podem reportar eventos raros.</li> <li>• Educativos.</li> </ul>	<ul style="list-style-type: none"> <li>• Não têm força suficiente para confirmar associação estatística, portanto, não confirma hipóteses.</li> <li>• Podem refletir a característica daquele paciente ou grupo de pacientes, não sendo necessariamente extrapolável para o mundo real.</li> </ul>
Estudos ecológicos	<ul style="list-style-type: none"> <li>• Examina dados de saúde e cruza com informações geográficas para obter rápidas informações sobre as características de uma doença na população daquele local (bairro, cidade, estado, país ou continente).</li> </ul>	<ul style="list-style-type: none"> <li>• Resultados são baseados na média da população e, portanto, não necessariamente se aplicam a todos os indivíduos do país.</li> <li>• Diferenças nos protocolos de diagnóstico e tratamento podem ser responsáveis por diferenças geográficas das características das doenças.</li> </ul>
Pesquisas transversais	<ul style="list-style-type: none"> <li>• Dão um panorama da doença e da exposição a fatores de risco (ou terapias) ao mesmo tempo.</li> <li>• São úteis para determinar prevalência de uma doença ou dos seus desfechos.</li> </ul>	<ul style="list-style-type: none"> <li>• Não conseguem determinar se a exposição ao fator de risco estudado ou à terapia em questão precedeu a doença ou o desfecho (falta “temporalidade”, um critério <i>sine qua non</i> de Bradford Hill).</li> </ul>

Desenho do estudo	Características	Limitações
Estudos de caso-controle	<ul style="list-style-type: none"> <li>• Observam o pesquisado de agora, olhando para o seu passado, portanto, útil para tentar averiguar razões para ele estar doente ou com o desfecho de agora.</li> <li>• Relativamente rápidos e baratos.</li> <li>• Bons para doenças crônicas que levam muito tempo para determinar seus desfechos.</li> <li>• Bons também para doenças raras.</li> <li>• Podem examinar múltiplos fatores de risco ou terapias para uma única doença.</li> </ul>	<ul style="list-style-type: none"> <li>• Suscetíveis aos vieses de seleção e de memória.</li> <li>• Relações temporais entre fatores de risco (ou terapias) e doenças (ou desfechos) são, às vezes, dificilmente estabelecidas.</li> <li>• São ineficientes para fatores de risco raros.</li> </ul>
Estudos de coorte	<ul style="list-style-type: none"> <li>• Observam o pesquisado de agora para seu futuro, portanto, útil para averiguar se um fator de risco levará a uma doença no futuro.</li> <li>• Excelente para estudar fatores de risco raros.</li> <li>• Consegue estudar múltiplos desfechos desses fatores de risco.</li> <li>• Permite determinar a incidência de uma doença.</li> <li>• Relações temporais entre exposição ao fator de risco (ou a alguma terapia) e doença (ou desfecho) podem ser estabelecidas.</li> </ul>	<ul style="list-style-type: none"> <li>• Podem ser longos e caros.</li> <li>• Perdas de <i>follow-up</i> (isto é, pacientes que saem da pesquisa) podem impactar na validade dos resultados.</li> <li>• Não conseguem confirmar hipóteses em casos em que se estuda o efeito de terapias.</li> </ul>

Desenho do estudo	Características	Limitações
Estudos randomizados controlados	<ul style="list-style-type: none"> <li>• Selecionam pesquisados com características predefinidas e uma doença em comum para dar uma terapia a metade deles, dando placebo ou outra terapia à outra metade.</li> <li>• Podem entregar a mais forte evidência da existência de alguma relação de causa-efeito entre exposição (ou terapia) e doença (ou desfecho).</li> <li>• Randomização minimiza potenciais vieses e impacto de fatores confundidores.</li> <li>• Padrão-ouro no estudo de terapias.</li> </ul>	<ul style="list-style-type: none"> <li>• Podem ser (e costumam ser) caros e longos.</li> <li>• Várias considerações éticas precisam ser feitas e revistas durante a realização de um estudo como esse, particularmente quando submete pessoas a uma nova terapia.</li> <li>• Pode ser impraticável quando se necessite de amostras e <i>follow-up</i> (acompanhamento) muito longos para se determinar um resultado significante.</li> <li>• Apesar de ser “padrão-ouro”, ainda é suscetível a diversos vieses que serão vistos no Capítulo 4.</li> </ul>

Características e limitações dos tipos de artigos científicos que apresentamos neste capítulo.

A pirâmide da Medicina Baseada em Evidências é importante como um norte, mas o leitor deve entender a sua fluidez: muitas vezes, a melhor evidência que conseguimos ter virá de estudos observacionais. Outras vezes, ensaios clínicos randomizados serão tão enviesados, que não servirão como o topo da pirâmide.

## Referências bibliográficas

1. Hill AB. The environment and disease: association or causation? Proc R Soc Med. 1965 May;58(5):295-300.
2. Howick J, Glasziou P, Aronson JK. The evolution of evidence hierarchies: what can Bradford Hill’s “guidelines for causation” contribute? J R Soc Med [Internet]. 2009 May;102(5):186-94 [[acesso](#) em: 18 mar. 2021].
3. Fedak KM, Bernal A, Capshaw ZA, Gross S. Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology. Emerg Themes Epidemiol [Internet]. 2015 Sep 30;12:14 [[acesso](#) em: 18 mar. 2021].
4. Rothman KJ, Greenland S. Causation and causal inference in epidemiology. Am J Public Health. 2005;95 Suppl 1:S144-50.
5. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: New estimates of R&D costs. J Health Econ [Internet]. 2016;47:20-33 [[acesso](#) em: 18 mar. 2021].



6. Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. *Evid Based Med* [Internet]. 2016/06/23. 2016 Aug;21(4):125-7 [[acesso](#) em: 18 mar. 2021].
7. Sedgwick P. Ecological studies: advantages and disadvantages. *BMJ Br Med J* [Internet]. 2014 May 2;348:g2979 [[acesso](#) em: 18 mar. 2021].
8. Sedgwick P. Cross sectional studies: advantages and disadvantages. *BMJ Br Med J* [Internet]. 2014 Mar 26;348:g2276 [[acesso](#) em: 18 mar. 2021].
9. Sedgwick P. Bias in observational study designs: cross sectional studies. *BMJ*. 2015 Mar;350:h1286.
10. Kopec JA, Esdaile JM. Bias in case-control studies. A review. *J Epidemiol Community Health* [Internet]. 1990 Sep;44(3):179-86 [[acesso](#) em: 18 mar. 2021].
11. Jepsen P, Johnsen SP, Gillman MW, Sørensen HT. Interpretation of observational studies. *Heart* [Internet]. 2004 Aug;90(8):956-60 [[acesso](#) em: 18 mar. 2021].
12. Kenneth S. Design of Randomized Controlled Trials. *Circulation* [Internet]. 2007 Mar 6;115(9):1164-9 [[acesso](#) em: 18 mar. 2021].
13. Al-Lamee R, Thompson D, Dehbi H-M, Sen S, Tang K, Davies J, Keeble T, Mielewczik M, Kaprielian R, Malik IS, Nijjer SS, Petraco R, Cook C, Ahmad Y, Howard J, Baker C, Sharp A, Gerber R, Talwar S, Assomull R, Mayet J, Wensel R, Collier D, Shun-Shin M, Thom SA, Davies JE, Francis DP. Percutaneous coronary intervention in stable angina (ORBITA): a double-blind, randomised controlled trial. *Lancet* (London, England). 2018 Jan;391(10115):31-40.
14. Kendall JM. Designing a research project: randomised controlled trials and their principles. *Emerg Med J* [Internet]. 2003 Mar;20(2):164-8 [[acesso](#) em: 18 mar. 2021].
15. Greco T, Zangrillo A, Biondi-Zoccai G, Landoni G. Meta-analysis: pitfalls and hints. *Hear lung Vessel* [Internet]. 2013;5(4):219-25 [[acesso](#) em: 18 mar. 2021]. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/24364016>.
16. Wakefield AJ, Murch SH, Anthony A, Linnell J, Casson DM, Malik M, Berelowitz M, Dhillon AP, Thomson MA, Harvey P, Valentine A, Davies SE, Walker-Smith JA. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children (RETRACTED ARTICLE). *Lancet* (London, England). 1998 Feb;351(9103):637-41.
17. MacMahon B, Yen S, Trichopoulos D, Warren K, Nardi G. Coffee and cancer of the pancreas. *N Engl J Med*. 1981 Mar;304(11):630-3.
18. Andersson C, Johnson AD, Benjamin EJ, Levy D, Vasan RS. 70-year legacy of the Framingham Heart Study. *Nat Rev Cardiol* [Internet]. 2019;16(11):687-98 [[acesso](#) em: 18 mar. 2021].
19. Prasad V, Cifu AS. The Necessity of Sham Controls. Vol. 132, *The American journal of medicine*. United States; 2019. p. e29–30.

## 4. VIESES EM ESTUDOS CIENTÍFICOS

### Introdução

Você acordou hoje em sua linda casa das Ilhas Maldivas. Vai até a janela, da qual se vê o Leste, e observa algumas nuvens escuras típicas de uma tempestade tropical; depois, volta-se à janela que aponta ao Oeste, dá um suspiro e admira um lindo horizonte azul refletindo os raios de um sol brilhante. Pensa: “Hoje, será um ótimo dia para mergulhar com os tubarões”.

Você veste sua roupa de mergulhador, despede-se das pessoas da casa, abre a porta, anda 5 metros em direção ao sol e retorna: uma tempestade acaba de chegar. Você até olhou para as nuvens escuras no lado Leste, mas o lindo do Sol do Oeste era tão estonteante, que ignorou qualquer possibilidade de chuva e **enviesou** seu pensamento. Você viu apenas o que quis ver para seu propósito de diversão e acabou tendo uma interpretação errada do clima. Apesar de ser uma situação hipotética (talvez, nem tanto assim), o nosso dia a dia é cheio de circunstâncias parecidas. O raciocínio humano é repleto dessas armadilhas, que acometem não só nossas intuições pessoais, mas nossos melhores métodos de conhecimento, como veremos a seguir.

A intuição humana é enviesada e cheia de armadilhas, e a ciência surge com a finalidade de minimizá-las. Dado nosso contexto atual, de importante dependência do conhecimento, esse esforço é extremamente relevante, até mesmo porque a própria ciência, sendo construída por humanos, é sujeita a esses erros, chamados de vieses. A definição formal de viés é “qualquer processo em qualquer estágio de inferência que tende a provocar resultados ou conclusões que diferem sistematicamente da verdade”.<sup>1</sup> Conhecê-los é imprescindível para os evitarmos e fomentarmos uma boa prática científica.

Neste capítulo, você vai conhecer os vieses mais comuns e relevantes da pesquisa científica. Alerta: os próximos parágrafos podem conter humor não baseado em evidências.

### Viés do franco-atirador texano

Talvez, você já tenha ouvido falar deste viés, mas não com esses termos. Trata-se de uma analogia descrevendo um atirador que desfere vários

disparos contra um objeto plano e, só após ter feito vários buracos, desenha cuidadosamente linhas ao redor deles, como se essas fossem o verdadeiro alvo (Figura 1). O exemplo ilustra o problema de análises *a posteriori*, ao invés de *a priori*, que seria equivalente ao atirador ter desenhado os alvos antes de atirar, método realmente capaz de detectar a real eficácia dos tiros.<sup>2</sup>

**Figura 1.**



A análise feita *a posteriori* é muito bem ilustrada pelo viés do franco-atirador texano. A atitude de procurar por associações estatísticas espúrias apenas depois que o pesquisador já possui os dados é uma fonte inesgotável de vieses e tão confiável quanto desenhar o alvo apenas após o tiro.

Mas, na prática, o que seria uma análise *a posteriori*? Seria equivalente a ficar o dia inteiro olhando para uma tabela do Excel com vários dados coletados buscando uma relação entre eles. Todos nós conhecemos quem faça isso. Se você não conhece, talvez, você seja esse indivíduo. A má notícia é que, apesar do esforço em contribuir com a ciência, essa não é a melhor maneira de testar hipóteses. Colher dados sem nenhum propósito para só posteriormente (ou *a posteriori*) testar ou formular uma hipótese ou “aproveitar” dados que já foram colhidos para um outro propósito são atitudes que aumentam muito as chances de encontrar uma relação espúria, produto do acaso.

## Escolha de cerejas (cherry-picking)

Não se classifica exatamente como um viés de pesquisa, mas merece menção honrosa, visto que o autor deste livro acredita que é a fonte principal da maioria dos vieses que serão discutidos adiante (e também do franco atirador texano, que você acabou de ler).

A atitude se define como uma falácia de atenção e ocorre quando um pesquisador escolhe seletivamente os dados que mais lhe convém naquele conjunto de dados apresentados e deixa de lado aqueles que negam a sua hipótese ou narrativa (Figura 2). Além do terreno de pesquisa, esse viés é muito comum no cotidiano, por exemplo, quando um jornal decide apresentar apenas notícias negativas sobre uma pessoa ou quando um médico decide compartilhar apenas estudos positivos sobre um medicamento.

**Figura 2.**



## Viés de seleção

O viés de seleção é aquele que acontece durante a escolha da amostra, levando a uma falha de representação legítima da população. É bastante

abrangente e incorpora vários outros vieses também caracterizados como tal. David Sackett expôs 22 subtipos em seu famoso artigo *Bias in Analytic Research* (está achando que a vida de pesquisador é fácil?).<sup>1</sup>

Para interpretar a existência do viés, o leitor deve buscar informações contidas na famosa “Tabela 1” (Figura 3). A primeira tabela de estudos observacionais ou ensaios clínicos randomizados costuma conter informações sobre as pessoas estudadas: quantos eram hipertensos? Qual a média de idade? Quantos eram homens ou mulheres?

### **Figura 3.**

	NGT group (n=697)	GDM group (n=177)	P values
Age (years)	32.73±5.10	33.44±5.28	0.111
Race/Ethnicity			
Caucasian	470 (67.4)	117 (66.1)	0.032
Hispanic	207 (29.7)	53 (29.9)	
Others	20 (2.9)	7 (4.1)	
Family history			
Type 2 diabetes	138 (19.8)	42 (23.7)	0.450
MetS (>2 components)	145 (20.8)	38 (21.5)	
History			
GDM	15 (2.2)	10 (5.7)	0.043
Miscarriage	225 (32.3)	53 (29.9)	
Educational status			
Elementary education	53 (7.6)	25 (14.2)	0.030
Secondary school	167 (24.0)	45 (25.4)	
University degree	471 (67.6)	106 (59.9)	
UNK	6 (0.9)	2 (1.2)	
Employment	539 (77.3)	135 (76.3)	0.844
Number of pregnancies			
Primiparous	301 (43.2)	77 (43.5)	0.857
Second pregnancy	225 (32.3)	56 (31.6)	
>2 pregnancies	171 (24.5)	44 (25.9)	
Smoker			
Never	384 (55.1)	93 (52.5)	0.786
Current	55 (7.9)	17 (9.6)	
Gestational age (weeks) at entry	12.1±0.5	12.1±0.7	0.261
Body weight (kg)			
Prepregnancy	59.7±10.2	65.1±12.0	0.0001
At entry	61.8±10.5	67.2±12.3	0.0001
Weight gain at entry	2.0±2.9	2.2±2.9	0.508
BMI (kg/m <sup>2</sup> )			
Prepregnancy	22.6±3.5	24.9±4.2	0.0001
At baseline	23.4±3.6	25.7±4.4	0.0001
Systolic BP (mm Hg)	106±11	111±10	0.001
Diastolic BP (mm Hg)	65±10	67±8	0.063
FBG (mg/dL)	81±6	83±6	0.0001
HbA <sub>1c</sub> % (mmol/mol)	5.2±0.3 (33±0.9)	5.0±0.3 (31±0.9)	0.165
Cholesterol (mg/dL)	173±29	177±31	0.288
Triglycerides (mg/dL)	80±41	89±37	0.034
MEDAS score	4.90±1.73	4.89±1.56	0.931
Nutrition Score	0.37±3.2	0.27±3.16	0.699
Physical Activity Score ≥0	80 (12.4)	19 (10.8)	0.694

Data are mean±SD or number (%).

Physical Activity Score ≥0: (1) walking daily (>5 days/week); score 0: at least 30 min; score +1: if >60 min; score -1: if <30 min. (2) Climbing stairs (floors/day, >5 days a week): score 0, between 4 and 16; score +1, >16; score -1: <4.

BMI, body mass index;BP, blood pressure;FBG, fasting blood glucose;GDM, gestational diabetes mellitus;HbA<sub>1c</sub>, hemoglobin A<sub>1c</sub>;MEDAS, Mediterranean Diet Adherence Screener;MetS, metabolic syndrome;NGT, normal glucose tolerance;UNK, unknown.

Imagem com o exemplo de uma tabela de *baseline*, a famosa “tabela 1”. Observe que aqui estão descritos dados demográficos e características clínicas que podem atuar como fatores confundidores em uma pesquisa. Entenda que cada pesquisa pode incluir dados diferentes na Tabela 1 porque as várias doenças possuem fatores de risco e confundidores diferentes. Perceba também que os valores de *p* estão dispostos na última coluna dessa tabela, uma das maneiras encontradas para tentar definir significância estatística nessa diferença (que nesse caso, o ideal é que não exista) e passando a ideia de que o sorteio realizado para alocação dos pacientes no estudo não foi suficiente para homogeneizar o IMC pré-gravidez (em inglês, BMI *prepregnancy*) das participantes desse estudo, havendo um valor de *p* significativo nessa comparação. Se esse fator confundidor for muito fortemente relevante para o desfecho estudado, essa diferença pode enviesar o resultado a ponto de ele ficar desacreditado. Há dois problemas nesta interpretação (a de adicionar valor de *p* à “Tabela 1”): (a) a randomização é, por definição, um processo realizado ao acaso (sorteio) e, por isso, o valor de *p* é inútil; (b) mesmo as diferenças não significativas no *baseline* podem gerar vieses importantes nos desfechos estudados.<sup>53</sup>

Essas informações são importantes para que se avalie se aquela população estudada reflete a vida real (ou se, ao contrário, são mais novos e de melhor prognóstico que o basal de uma doença...) e, também, se participantes dos dois grupos têm características demográficas e de saúde similares.

## Viés de alocação

O objetivo de ensaios clínicos é comparar intervenção com o placebo (ou outra intervenção) a partir de uma amostra de uma certa população dividida, respectivamente, em grupo-intervenção e grupo-controle. Uma simples divisão da amostra, porém, não é o suficiente. Os integrantes precisam ser alocados de maneira que as características entre os grupos sejam as mais semelhantes possíveis, à exceção da intervenção em estudo. Em outras palavras, os participantes precisam ser **randomizados**. O viés da alocação surge quando o pesquisador tem conhecimento prévio do grupo ao qual o paciente foi alocado, gerando uma interferência e levando a uma diferença sistemática entre os dois grupos do ensaio clínico.<sup>3-5</sup>

Uma análise de 280 estudos (entre ensaios clínicos e meta-análises) demonstrou que 80% deles tiveram ocultação de alocação pouco clara. Esse viés pode ser responsável por um aumento das estimativas de tamanho do efeito da ordem de 7% a 40%.<sup>6</sup>

Uma alocação ideal inclui a randomização, prevenindo diferenças sistemáticas entre os grupos, e a ocultação dos resultados dela, evitando alterações voluntárias posteriores, como a troca de pacientes entre os grupos ou a exclusão de pacientes em virtude de suas particularidades.<sup>5, 7</sup>

## Viés de atrito (ou viés de exclusão)

A palavra “atrito” implica o processo de perda de indivíduos durante um estudo que usualmente acontece como fruto do acaso, por fatores como inelegibilidade, violação de protocolo, desfecho precoce ou perda de contribuição. Quando a perda é desproporcional entre os grupos (>5%), há possibilidade de haver viés de atrito.<sup>8,9</sup>

## Viés de acesso ao diagnóstico

Certas pessoas, em virtude de discrepâncias temporais, espaciais ou econômicas, têm acesso maior a testes diagnósticos do que outras. Isso pode gerar uma falsa sensação de que uma doença é mais prevalente em um determinado grupo de pessoas em comparação a outro grupo que, por acaso ou por tais discrepâncias, não esteja sendo testado na mesma proporção.<sup>1,10</sup>

## Viés de suspeita diagnóstica

Saber sobre o histórico de um paciente pode induzir a uma investigação mais ou menos intensa acerca de um determinado diagnóstico.<sup>1</sup> Esse é um viés perigoso e difícil de lidar, pois pode gerar tanto sub quanto sobrediagnóstico, a depender da subjetividade que for lançada por um paciente no seu médico. Um bom exemplo é a coleta diária de exames de sangue e eletrólitos em pacientes internados imaginando que, apenas por estarem no hospital (independentemente de suas patologias), as chances de desenvolverem distúrbios sanguíneos ou hidroeletrólíticos é maior.

## Viés do associado

Refere-se à escolha de um grupo de pesquisa com características sistematicamente diferentes do resto da população.<sup>1</sup> Um exemplo clássico é o **efeito do trabalhador saudável**, presente em estudos de coorte ocupacional que comparam participantes empregados com a população geral. O problema é que pessoas que não trabalham podem não estar trabalhando exatamente por serem menos saudáveis (p. ex., aposentados com mais idade e comorbidades). Logo, ao tentar analisar desfechos como mortalidade em pessoas empregadas (mais saudáveis), o número será naturalmente menor do que na população em geral.<sup>11,12</sup>



## Viés do voluntário

Também conhecido como **viés de não resposta**, é um fenômeno presente nos mais variados tipos de pesquisa, mas é mais evidente em questionários. Nesse caso, surge quando há uma diferença relevante entre quem responde e quem não responde à pesquisa.<sup>1, 13</sup>

Pense no exemplo hipotético de um questionário sobre o uso de drogas em estudantes universitários. Caso a pesquisa seja aberta, os voluntários que a respondem podem ser diferentes dos que não respondem, pois os que não respondem podem, por exemplo, ser usuários e ter medo de ser expostos.

## Viés da taxa de admissão

Também conhecido como **viés ou paradoxo de Berkson**, em homenagem ao primeiro que o descreveu, em 1946.<sup>14</sup> Refere-se a conclusões espúrias em estudos observacionais com população hospitalar ao se tentar relacionar duas condições que ocorrem de maneira independente, mas, pela metodologia do estudo, casos em que nenhuma das condições ocorre foram excluídos da análise.<sup>1, 14</sup>

Um exemplo disso é o clássico estudo do MacMahon, que concluía que a ingestão de café era um fator de risco para câncer de pâncreas. O que se descobriu depois é que um dos fatores que levou MacMahon a fazer essa conclusão espúria foi o fato de que o grupo controle (isto é, os pacientes sem câncer observados nesse estudo para comparação) continha pacientes internados pela equipe de gastroenterologia do hospital. Por essa razão, essa era uma população com alta incidência de doença gastrointestinal, o que fazia os médicos aconselharem a não beber café. Naturalmente, pois, haveria uma diferença significativa no consumo de café entre os dois grupos de comparação.<sup>15</sup>

Um exemplo mais atual pôde ser visto na busca por grupos de risco durante a pandemia de Covid-19. Muitas das análises que definiram esses grupos podem ter sido enviesadas pelo desenho do estudo se, por acaso, tiveram metodologia similar. Comparar pacientes hospitalizados por Covid-19 *versus* hospitalizados por outras causas exclui os pacientes não hospitalizados por Covid-19 e pelas outras causas (Figura 4).

**Figura 4.**



O paradoxo de Berkson ocorre quando, através da metodologia de um estudo, se excluem pacientes que não possuem as duas condições estudadas. Muito comum em estudos de pacientes hospitalizados, porque são excluídos os pacientes não hospitalizados (portanto, de melhor prognóstico) de ambas as condições estudadas, restando apenas aqueles com pior prognóstico e facilitando a criação de conclusões espúrias sobre essa comparação.

Um viés que segue a mesma linha é a **heurística de colisão**. Existe uma variável (chamada, nesse caso, de *variável de colisão*) comum entre pessoas hospitalizadas por qualquer razão, que é o fato de estar hospitalizado. Quando estudos são condicionados por essa variável (hospitalizados), há grandes chances de associações espúrias.<sup>16</sup>

## Vieses de prevalência-incidência e sobrevivência

Esse merece um certo destaque, não exatamente por sua definição, mas por ter sido gerado num artigo feito pelo cientista Jerzy Neyman (razão pelo qual também pode ser chamado de **viés de Neyman**) intitulado como “Estatística – a serva de todas as ciências”.<sup>17</sup> Ora, é isso que temos tentado demonstrar por aqui.

Em um estudo observacional, casos incidentes fornecem dados mais confiáveis que prevalentes, pois é possível acompanhar a doença desde o início, sem que os indivíduos tenham passado por tratamentos ou progredido com sua doença. Pacientes com doenças antigas com progressão de gravidade terão uma maior severidade, podendo não ser hábeis a entrarem no estudo.<sup>18, 19</sup> Quando a doença leva à morte, pode-se originar o **viés de sobrevivência**: apenas indivíduos que sobrevivem são incluídos, *subestimando* a gravidade da doença.<sup>20, 21</sup> Por sua vez, pacientes com doenças antigas que já passaram por tratamento podem já estar melhores ao ponto de não serem incluídos no estudo, reservando a amostra a casos mais severos e, desse modo, *superestimando* a gravidade.<sup>19</sup>

Um exemplo clássico do viés de sobrevivência, que você já deve ter ouvido falar, é o trabalho da equipe do estatístico Abraham Wald durante a

Segunda Guerra Mundial. Após analisar os pontos mais atingidos por projéteis de aviões que voltavam de batalhas, ele concluiu que essas estruturas deveriam ser reforçadas. No entanto, os aviões continuaram sendo abatidos na mesma frequência. O motivo é que os aviões atingidos nas estruturas que Wald observara não caíam e, por isso, voltavam, ou seja, eram sobreviventes. Aviões atingidos em pontos vitais caíam no campo de batalha, então, não eram incluídos na amostra estudada (Figura 5).<sup>22</sup>

**Figura 5.**



Viés do sobrevivente. Na Segunda Guerra Mundial, os militares aliados quiseram reforçar as áreas em que os aviões mais frequentemente eram atingidos para melhorar a sua robustez nos embates contra os nazistas. Mas o matemático húngaro Abraham Wald apontou que, na verdade, como esses pontos foram estudados de aviões que conseguiram retornar à base (sobreviventes), os dados dos aviões que, de fato, sofreram danos graves estão faltando na amostra, pois não foram colhidos. Segundo ele, os pontos com menos balas é que deveriam ser reforçados.

## Viés do tratamento de sobreviventes

Apesar de ter nome parecido e também ser aplicado a estudos observacionais, adquire um sentido distinto do já mencionado *viés de sobrevivência*. Um sujeito que sobrevive por mais tempo e fez uso de um medicamento pode ter facilmente a sobrevida associada ao seu consumo. No entanto, o próprio fato de haver mais tempo de vida predispõe ao uso mais longo e frequente do medicamento. A sobrevida, portanto, ao invés de

consequência, é a causa, podendo gerar inferências erradas acerca de tratamentos que, na realidade, são ineficazes.<sup>23</sup>

Como exemplo, considere um certo tipo de câncer cuja taxa de progressão varia entre extremamente lenta e extremamente rápida, mas a Medicina ainda não sabe disso. Os pacientes com progressão extremamente rápida sequer tiveram tempo de procurar o serviço e morreram muito antes, então, os que receberão o tratamento serão os que, naturalmente, teriam maior sobrevida. Seria melhor eliminar esse viés caso existisse um estudo controlado e randomizado, porém, em um estudo observacional (especialmente, caso-controle), a chance de uma conclusão equivocada é maior.

## Viés de averiguação

Chamado também de **viés de amostragem**, refere-se a uma “escolha de amostra que favorece determinados desfechos”.<sup>24</sup> Uma ótima ilustração do viés pelo Departamento de Matemática da Universidade do Texas fala sobre pesquisas de *marketing* por telefone cuja amostra tenha sido escolhida por uma lista telefônica de determinada área. Esses trabalhos vão deixar escapar consumidores sem telefone, fora da área ou que não desejam participar, atingindo apenas um tipo muito específico e selecionado de consumidor e, possivelmente, gerando conclusões não extrapoláveis à população como um todo.<sup>24</sup>

O viés de averiguação pode, certas vezes, ser tratado como um *viés de detecção*, sendo definido como uma “distorção sistemática no acesso às medidas de desfecho por investigadores ou participantes por eles estarem cientes da alocação do tratamento”.<sup>25</sup> Percebe-se, portanto, que o viés de averiguação pode apresentar dois significados diferentes, um relacionado à escolha da amostra e outro ao cegamento.

É importante não confundir *viés de amostragem* com erro de amostragem. Enquanto o primeiro termo trata de um erro sistemático, o segundo define a natural imprecisão de uma amostra em relação à população em função de seu menor tamanho.<sup>26</sup>

## Viés de aferição

Também é chamado de **viés de informação** (ambos os nomes usados preferencialmente por epidemiologistas), ou como **viés de detecção** (mais

usado por pesquisadores de estudos clínicos, ou trialistas).<sup>27</sup> Trata-se de erros nas medidas ou classificações de desfechos, riscos de desfecho, *status* de intervenção ou exposições, causando valores inapropriados nas medidas de associação e efeito.<sup>8, 27-29</sup>

Pode ser classificado em **não randômico (diferencial)** – quando as informações são coletadas de maneira discrepante entre os grupos de intervenção e controle, gerando resultados irreais – e **randômico (não diferencial)** – quando há coleta errônea em ambos os grupos nas mesmas proporções, ocultando resultados reais e inclinando-se à hipótese-nula.<sup>29-30</sup> Na prática científica, o tipo randômico causa menos problemas – motivo pelo qual é chamado de *clean dirt* (“sujeira limpa”) – que o não randômico – denominado de *dirty dirt* (“sujeira suja”).<sup>28-29</sup>

Assim como ocorre com o viés de seleção, o viés de aferição também pode ser subdividido em muitos outros vieses. Então, prepare-se que lá vem mais uma série de tópicos.

## Viés de memória

Você lembra do que comeu exatamente 6 dias atrás? Provavelmente, não, muito menos eu. No entanto, caso 6 dias atrás, você estivesse na Turquia almoçando no restaurante de Nusret Gökçe com o dinheiro da Mega da Virada, lembraria. Para o garçom, porém, você seria apenas mais um cliente.

A memória depende de muitos fatores, sejam eles relacionados ao sujeito, (idade, nível de escolaridade e/ou *status* socioeconômico) ou ao evento de exposição (intervalo entre ocorrência e pergunta, frequência, duração, vivacidade, significado e/ou aceitabilidade social). Estudos que dependem da memória (retrospectivos, como caso-controle ou coortes históricos) podem ser facilmente enviesados, particularmente, quando tratam de exposições raras ou são feitos com grupos-controle originados da comunidade (e não de nosocômios).<sup>1, 31</sup>

Ainda, o viés pode afetar pesquisadores, quando se fazem mais perguntas aos casos que aos controles, por esperarem alguma resposta.<sup>32</sup>

## Viés do observador

Apesar de termos instrumentos capazes de descrever o mundo de maneira objetiva e formal, muitas vezes, somos dependentes da interpretação humana. Quando a interpretação dos fenômenos por um certo observador difere dos fenômenos reais, surge este viés. Além disso, por ter um caráter

subjetivo, as interpretações podem variar entre diversos observadores (**interobservador**) ou mesmo em apenas um que olhe para um mesmo fenômeno (**intraobservador**).<sup>8, 33</sup>

Variações inter e intraobservadores, por exemplo, estão presentes em uma ausculta pulmonar. Estudos que precisem dos dados desse exame são propensos a apresentarem o viés do observador, em especial, se os examinadores não forem cegos.

## Viés de expectativa

Quando investigadores coletam dados com uma expectativa prévia dos seus resultados, eles podem ser distorcidos pelo viés da expectativa.<sup>1</sup> Precisa falar, de novo, que o cegamento minimiza isso? Imagine um ensaio clínico controlado, randomizado e não cego que busca saber a eficácia de flebotônicos para melhorar sintomas de insuficiência venosa crônica. Investigadores que souberem quais pacientes receberam o medicamento podem, naturalmente, supervalorizar os relatos dos indivíduos desse grupo, gerando um resultado que não condiz com a realidade.

## Viés de verificação

Ocorre em estudos de testes diagnósticos e rastreamento. Esses estudos comparam um teste experimental a um teste que seja considerado como “referência” ou padrão-ouro. A partir dessa comparação, determina-se a sensibilidade e a especificidade do novo teste.

O delineamento em questão pode favorecer o viés de verificação, de duas formas possíveis: apenas alguns participantes recebem o teste de referência (**parcial**) ou indivíduos recebem testes de referência diferentes (**diferencial**). Certas vezes, é financeiramente inaplicável usar testes de referência em todos (pelo seu alto custo) e até mesmo antiético, pois muitos são invasivos.<sup>8</sup>

Para exemplificar, imagine que um estudo testa a acurácia do d-dímero para diagnóstico de trombose venosa profunda (TVP). Pacientes com d-dímero positivo fariam ultrassonografia (exame de referência 1), enquanto os pacientes que possuíam d-dímero negativo seriam encaminhados para uma consulta de rotina 3 meses após (exame de referência 2). Assim, pacientes com TVP que, por acaso, tiveram d-dímero falso negativo podem ter passado despercebidos porque, após 3 meses, a doença pode ter se resolvido espontaneamente (história natural da doença). Um estudo com uma

metodologia como essa subestima os falsos negativos, portanto, superestima a sensibilidade.<sup>34</sup>

## Efeito Hawthorne

Esse foi um termo cunhado pelo alemão Henry Landsberger após criticar um estudo feito nos anos 20 do século XX por Elton Mayo em uma companhia elétrica de Hawthorne, Illinois. O estudo de Mayo analisava a performance dos trabalhadores do estabelecimento, e Landsberger percebeu que a produtividade alta se devia à presença do estudo; após seu término, ela diminuiu.<sup>8, 35-36</sup>

O Efeito Hawthorne é definido como “uma mudança de comportamento como resposta motivacional a interesse, cuidado e atenção recebidos através da observação e da avaliação”.<sup>36</sup> Tanto os pacientes – dos grupos de intervenção e controle – quanto cuidadores podem modificar seu desempenho (positiva ou negativamente) apenas por saberem que estão sendo parte de um estudo. Contudo, quando a mudança parte do grupo-controle, recomenda-se a denominação de **efeito John Henry**.<sup>37</sup>

## Viés de resposta

Apesar de parecer uma antítese ao viés de não resposta, esse viés adquire um sentido diferente. O *viés da resposta* ocorre quando os sujeitos respondem de maneira irrealista às perguntas, em outras palavras, mentem (não necessariamente de forma intencional). Por sua vez, o *viés de não resposta* surge quando há uma diferença relevante entre as pessoas que respondem e as pessoas que não respondem ao questionário.<sup>13</sup>

Como ilustração, pense novamente no questionário sobre o uso de drogas em estudantes universitários. Sujeitos que usam drogas podem escolher responder a pesquisa, mas mesmo sabendo do anonimato, omitir informações por medo de serem expostos (*viés de resposta*). Igualmente, os usuários podem escolher não participar da pesquisa pelo mesmo medo, reservando-a aos não usuários (*viés da não resposta*).

Neste tópico, também vale citar o **viés da obsequiosidade**, que se refere a uma alteração voluntária das respostas na direção do desejo do pesquisador.<sup>1</sup>

## Viés da medida insensível

Quando tratamos da dualidade hipótese-nula *versus* hipótese-alternativa, podemos nos deparar com os erros tipo I (falso-positivo) e II (falso-negativo). Um dos motivos do erro tipo II é o viés de medida insensível, quando, em função da falta de acurácia de um teste, somos incapazes de encontrar diferenças clinicamente significantes nos estudos.<sup>1</sup>

## Vieses de relato

Problema particularmente relevante nos estudos de mais alta qualidade, o viés de relato faz referência a um grupo de vieses que interfere diretamente na exposição ou ocultação de trabalhos (em revisões sistemáticas ou meta-análises) ou de resultados (em ensaios clínicos) por conta de sua natureza e direção.<sup>8, 38</sup>

## Viés de publicação

Ocorre quando pesquisas não publicadas não são incluídas em revisões pelo simples fato de não serem conhecidas e o fato de não ter havido publicação tem relação direta com as conclusões produzidas (Figura 6).<sup>39-40</sup>

**Figura 6.**



Viés de publicação.

Estudos com resultados significativos têm mais chances de ser publicados, por exemplo.<sup>40</sup> Então, quando um pesquisador faz um compilado dos trabalhos para sua meta-análise, é provável que ache mais resultados que



se distanciam da hipótese nula (uma vez que os que se aproximam não foram publicados).

## Viés do relato de desfecho

Acontece quando trialistas escolhem voluntariamente quais resultados devem ou não ser expostos em sua pesquisa.<sup>38</sup> Geralmente, o fenômeno acontece quando os desfechos são indesejáveis ou não cumprem com as expectativas do pesquisador, o que pode gerar, por exemplo, um subrelato de efeitos adversos ou, se envolver desfechos primários, resultados incompletos.<sup>38, 40</sup>

De todo o grupo dos vieses de relato, os de publicação e relato de desfecho merecem um maior destaque, mas quase todo o grupamento pode ser resumido em uma citação do grupo *Cochrane* (grifo nosso):

[Estudos com] resultados positivos, estatisticamente significantes, que indicam que uma intervenção funciona, têm mais chance de serem publicados [**viés de publicação**], mais chance de serem publicados rapidamente [**viés de atraso**], mais chances de serem publicados em inglês [**viés de linguagem**], mais chances de serem publicados mais de uma vez [**viés de múltipla publicação**], mais chances de serem publicados em jornais de alto impacto [**viés de localização**] e, com base no último ponto, mais chances de serem citados por outros [**viés de citação**].<sup>39</sup>

## Viés cronológico

Já falamos do **viés de prevalência-incidência**, que acontece quando selecionamos casos prevalentes ao invés de incidentes e geramos um erro sistemático. Se indivíduos forem escolhidos por incidência, podem ser acompanhados desde o diagnóstico, enquanto se por prevalência, é possível que haja uma maior progressão de gravidade ou uso de tratamentos prévios que ofuscam o pesquisador.

Essa situação é análoga ao que acontece no **viés cronológico**. O recrutamento de indivíduos para um estudo demanda tempo. Durante isso, mudanças acerca de definições ou tratamentos podem ocorrer e, caso eles sejam utilizados como condicionantes para a apuração dos indivíduos, a metodologia do estudo pode ser comprometida.<sup>41, 42</sup>

## Vieses de intervenção

Também conhecido como **viés de exposição**, refere-se a um grupo de vieses que gera exposições ou intervenções desiguais entre braços de um estudo (em geral, controle e intervenção).<sup>43</sup>

## Viés de cointervenção

Quando se faz um ensaio clínico, o intuito é isolar o máximo de variáveis possível para manter apenas aquela de interesse. Intervenções paralelas desproporcionais em um dos braços do estudo são variáveis extremamente importantes, que podem alterar a frequência do desfecho medido.<sup>43, 44</sup>

Sackett faz uma ilustração exemplar. Imagine-se na década de 1980 sendo médico-assistente de um paciente hipertenso que, após três ataques isquêmicos transitórios, teve detectadas pressão ligeiramente alterada e 80% de obstrução na artéria carótida esquerda. Por conta disso, o sujeito, após consentimento, foi selecionado para um grupo controle de um ensaio clínico não cego de endarterectomia de carótidas. Você teve conhecimento disso e, temendo que ele pudesse ter um final desagradável pela falta de tratamento, fez questão de controlar seus níveis pressóricos de maneira mais rigorosa. A mudança de comportamento, se seguida nos mesmos moldes pelos médicos assistentes dos pacientes do grupo-controle, provocaria uma diminuição do risco. Dessa forma, subestimaria os possíveis efeitos do procedimento medido. Da mesma maneira, pacientes que passaram pela cirurgia podem ter uma motivação a mais para cuidarem melhor de sua saúde, sendo mais aderentes às terapias com antiagregantes plaquetários ou anti-hipertensivos (partindo do pressuposto que esses medicamentos melhoram desfechos), superestimando o efeito do procedimento.<sup>44</sup>

Percebam que o problema principal está sempre na cointervenção ser aplicada de maneira diferente nos dois braços do estudo. Portanto, se o trialista for capaz de organizar os grupos de modo que cointervenções sejam proporcionais (e a principal forma de fazer isso é por meio do cegamento), é improvável que o estudo seja vítima do viés em questão.

## Viés de contaminação

Viés não prevenido pela randomização, ocorre quando pacientes do grupo-controle recebem a intervenção em estudo (ou seja, o controle torna-se contaminado) por motivos variados. Esse processo desvia a conclusão em

direção à hipótese-nula, uma vez que o risco do grupo-controle se aproximará do risco do grupo intervenção diminuído (reduzindo, pois, alguma possível diferença entre os dois braços em estudo).<sup>43, 45-46</sup>

O crossover, que será comentado no capítulo 7, é uma forma de viés de contaminação. No estudo CABANA, por exemplo, que estudou ablação de fibrilação atrial *versus* medicamentos em um acompanhamento de 4 anos para os desfechos: morte, AVC debilitante, sangramento maior ou parada cardíaca; houve um crossover de 27,5% do braço terapia farmacológica para o braço intervenção (no caso, ablação de fibrilação atrial) e 9,2% do braço intervenção para a terapia farmacológica.<sup>47</sup>

Uma outra situação que favorece esse viés é o não cegamento. Algumas vezes, pacientes entram em estudos experimentais apenas para ter acesso a algum “tratamento revolucionário” buscando alguma solução para seu problema. Saber que estão no grupo-controle acaba lhes frustrando e motivando a buscar uma solução.

## Viés de proficiência

Quando há uma diferença entre os grupos em virtude da diferença de habilidades ou de treinamento dos técnicos que estão aplicando a intervenção.<sup>43</sup> Por exemplo, se você fizer um estudo comparando a acurácia diagnóstica em exames de imagem por radiologistas e médicos generalistas, o resultado vai claramente favorecer os radiologistas, profissionais com maior treinamento.

## Viés de aderência

Quando, além da intervenção, o grau de aderência dos indivíduos ao estudo influencia no desfecho. Um exemplo dado por Sackett: “são os pacientes com alto risco de doença coronariana que abandonam os programas de exercício”.<sup>1, 43</sup>

## Viés do tempo de intervenção

Quando não só a intervenção impacta o desfecho, mas também o momento em que ela foi realizada.<sup>54</sup> Por exemplo, testar apenas a eficácia da desfibrilação isoladamente em uma parada cardiorrespiratória poderia gerar resultados insatisfatórios, uma vez que, se feita de maneira precoce, o impacto é bem maior que quando realizada tardiamente.

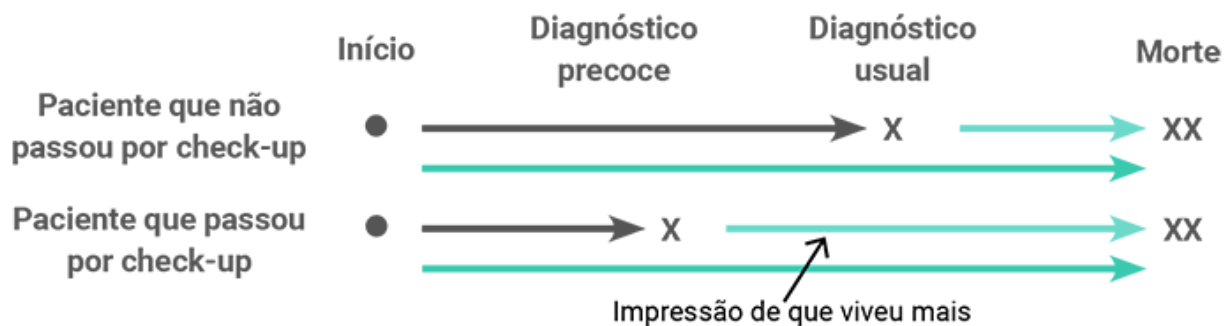
## Vieses inerentes ao rastreamento

Embora possam ser encaixados acima, existem alguns vieses que decidi abordar separadamente por serem exclusivos da prática do rastreamento de doenças. Eles serão mais detalhados no capítulo 13.

### Viés do tempo ganho (lead time bias)

O rastreamento, para funcionar, precisa antecipar o diagnóstico. O viés do tempo ganho ocorre quando o momento do diagnóstico é antecipado sem que a vida dos rastreados seja prolongada (Figura 7). Em outras palavras, os rastreados apenas vivem mais tempo com o diagnóstico, aparentando que viveram mais do que aqueles diagnosticados clinicamente.<sup>48</sup>

**Figura 7.**



*Lead time bias* ou viés do tempo de antecipação: mesmo que um *check-up* ou um rastreamento de doença não seja verdadeiramente impactante na quantidade de dias vividos, a impressão que passa, ao paciente e ao médico, é que foi.

### Viés de duração (length time bias)

As pessoas rastreadas, por definição, não têm sintomas daquela doença sendo pesquisada. Em comparação, quando ocorre o diagnóstico *clínico*, as pessoas têm sintomas e, justamente por isso, estão mais graves e com mais risco de morte (Figura 8). Ou seja, o grupo cuja doença é detectada por rastreamento terá um melhor prognóstico do aqueles que foram diagnosticados clinicamente, mesmo quando o rastreamento é completamente ineficaz.<sup>48</sup>

**Figura 8.**



*Length time bias* ou viés do tempo de duração. O rastreamento de doenças tende a identificar aquelas mais indolentes, que já teriam um comportamento mais benigno. Ao compararmos os prognósticos de pacientes rastreados com não rastreados, teremos a impressão que o rastreamento é benéfico, mesmo quando não é.

## Rob-2

Os artigos mais confiáveis não são aqueles com ausência de vieses (inexistentes), mas os que tentam minimizá-los (ou, ao menos, os que são honestos ao apresentá-los). Para simplificar essa análise, a Cochrane Library desenvolveu uma ferramenta capaz de avaliar o risco da presença dos vieses mais importantes, chamada de **Rob-2**.

Usada em ensaios clínicos randomizados, a Rob-2 é uma evolução de uma primeira versão publicada em 2008. Para utilizá-la, os pesquisadores precisam selecionar os resultados mais importantes do estudo em questão e especificar se este foi delineado com análise “por protocolo” ou por “intenção de tratar”. Após isso, terão acesso a cinco domínios, cada qual devendo ser explorado com uma série de perguntas objetivas, julgamento do tamanho do risco (alto, com algumas preocupações ou baixo) a partir de algoritmos pré-determinados e de opções para justificar as escolhas e tentar sugerir as direções dos vieses (se distanciam ou aproximam da hipótese-nula).<sup>49</sup>

Após analisar cada um dos domínios, o pesquisador deverá dar um julgamento completo do estudo (Tabela 1), classificando-o em baixo risco, com algumas preocupações e alto risco. É importante ressaltar que, se determinado domínio tem certo nível de viés, necessariamente, o estudo como um todo terá um desvio ao menos da mesma gravidade.

**Tabela 1.**

<b>Baixo risco</b>	Todos os domínios foram interpretados como sendo de baixo risco.
<b>Com algumas preocupações</b>	Houve algumas preocupações em ao menos um domínio.

<b>Alto risco</b>	Houve alto risco em ao menos um viés ou algumas preocupações em múltiplos domínios de maneira a diminuir a confiança do pesquisador.
-------------------	--

Tabela 1. Risco geral de vieses segundo a ferramenta ROB-2.<sup>33</sup>

Além da Rob-2, outras ferramentas foram desenvolvidas pela *Cochrane* para acessar riscos em outros tipos de estudos, como a **ROB-ME** (que busca por viés de relato em sínteses de evidências),<sup>50</sup> a **ROBINS-I** (usada em estudos intervencionais não randomizados)<sup>51</sup> e a **robvis** (usada em revisões sistemáticas).<sup>52</sup>

## Referências bibliográficas

1. Sackett DL. Bias in analytic research. *J Chronic Dis.* 1979;32(1):51-63.
2. Grufferman S. Clustering and aggregation of exposures in Hodgkin's disease. *Cancer.* 1977 Apr;39(4 Suppl):1829-33.
3. Attia A. Evidence-Based Medicine Corner Bias in RCTs: confounders, selection bias and allocation concealment. *Middle East Fertil Soc J.* 2005 Jan;10(3):258-61.
4. Sedgwick P. Selection bias versus allocation bias. *BMJ Br Med J.* 2013 May;346:f3345.
5. Nunan D, Heneghan C, Spencer EA. Catalogue of bias: allocation bias. *BMJ Evidence-Based Med.* 2018 Feb;23(1):20 LP-21.
6. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA.* 1995 Feb;273(5):408-12.
7. Paludan-Müller A, Teindl Laursen DR, Hróbjartsson A. Mechanisms and direction of allocation bias in randomised clinical trials. *BMC Med Res Methodol.* 2016;16(1):133.
8. Porta M. *A Dictionary of Epidemiology.* 6a. Oxford University Press; 2014. 376 p.
9. Tierney JF, Stewart LA. Investigating patient exclusion bias in meta-analysis. *Int J Epidemiol.* 2005 Feb;34(1):79-87.
10. Hennessy TW, Ballard DJ, DeRemee RA, Chu CP, Melton LJ 3rd. The influence of diagnostic access bias on the epidemiology of sarcoidosis: a population-based study in Rochester, Minnesota, 1935-1984. *J Clin Epidemiol.* 1988;41(6):565-70.
11. McMichael AJ. Standardized mortality ratios and the "healthy worker effect": Scratching beneath the surface. *J Occup Med.* 1976 Mar;18(3):165-8.
12. Kirkeleit J, Riise T, Bjørge T, Christiani DC. The Healthy Worker Effect in Cancer Incidence Studies. *Am J Epidemiol.* 2013 Jun;177(11):1218-24.
13. Sedgwick P. Non-response bias versus response bias. *BMJ Br Med J.* 2014 Apr;348:g2573.

14. Berkson J. Limitations of the Application of Fourfold Table Analysis to Hospital Data.\*;†. *Int J Epidemiol.* 2014 Apr;43(2):511-5.
15. MacMahon B, Yen S, Trichopoulos D, Warren K, Nardi G. Coffee and cancer of the pancreas. *N Engl J Med.* 1981 Mar;304(11):630-3.
16. Catalogue of bias collaboration. Lee H, Aronson JK ND. Collider Bias. *Catalogue of Bias.* 2019.
17. Neyman J. Statistics—Servant of All Science. *Science (80-).* 1955 Sep;122(3166):401 LP-406.
18. Magnus M. *Essentials of Infectious Disease Epidemiology.* Jones & Bartlett Learning; 2007. 227 p.
19. Norman DLS& GR. *PDQ Epidemiology.* 3. ed. McGraw-Hill Europe; 2009. 208 p.
20. Anderson CD, Nalls MA, Biffi A, Rost NS, Greenberg SM, Singleton AB, Meschia JF, Rosand J. The effect of survival bias on case-control genetic association studies of highly lethal diseases. *Circ Cardiovasc Genet.* 2011 Apr;4(2):188-96.
21. Ho AM-H, Dion PW, Yeung JHH, Joynt GM, Lee A, Ng CSH, Chang A, So FL, Cheung CW. Simulation of survivorship bias in observational studies on plasma to red blood cell ratios in massive transfusion for trauma. *Br J Surg.* 2012 Jan;99 Suppl 1:132-9.
22. Mangel M, Samaniego FJ. Abraham Wald's Work on Aircraft Survivability. *J Am Stat Assoc.* 1984 Jun;79(386):259-67.
23. Glesby MJMD, Hoover DR. Survivor Treatment Selection Bias in Observational Studies: Examples from the AIDS Literature. *Ann Intern Med.* 1996 Jun;124(11):999-1005.
24. Smith MK. Biased Sampling and Extrapolation. Departamento de Matemática da Universidade do Texas. 2012.
25. Sedgwick P. Bias in clinical trials. *BMJ.* 2011 Jul;343:d4176.
26. Sedgwick P. What is sampling error? *BMJ Br Med J.* 2012 Jun;344:e4285.
27. Mansournia MA, Higgins JPT, Sterne JAC, Hernán MA. Biases in Randomized Trials: A Conversation Between Trialists and Epidemiologists. *Epidemiology.* 2017 Jan;28(1):54-9.
28. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ WV (editors), editor. *Cochrane Handbook for Systematic Reviews of Interventions.* 2. ed. Chichester (UK): John Wiley & Sons; 2019.
29. Page LA, Henderson M. Appraising the evidence: what is measurement bias? *Evid Based Ment Heal.* 2008 May;11(2):36 LP-37.
30. Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet (London, England).* 2002 Jan;359(9302):248-52.
31. Coughlin SS. Recall bias in epidemiologic studies. *J Clin Epidemiol.* 1990;43(1):87-91.
32. Morgenstern J. Recall Bias. *First10EM blog.* 2018.
33. Mahtani K, Spencer EA, Brassey J, Heneghan C. Catalogue of bias: observer bias. *BMJ Evidence-Based Med.* 2018 Feb;23(1):23 LP-24.

34. O'Sullivan JW, Banerjee A, Heneghan C, Pluddemann A. Verification bias. *BMJ Evidence-Based Med.* 2018 Apr;23(2):54 LP-55.
35. Singletary R. Henry Landsberger 1926-2017. Departamento de Sociologia da Universidade de Carolina do Norte. 2017.
36. Sedgwick P, Greenwood N. Understanding the Hawthorne effect. *BMJ Br Med J.* 2015 Sep;351:h4672.
37. Deshpande S, Ostapenko L, Englesbe M. Chapter 13 – Threats to Experimental Hygiene. In: Markovac J, Kleinman M, Englesbe MBT-M and SP, editors. Academic Press; 2018. p. 121-9.
38. Sedgwick P. Meta-analysis: testing for reporting bias. *BMJ.* 2015 Jan;350:g7857.
39. Reporting Biases. *Cochrane Methods Bias.*
40. Sedgwick P. What is publication bias in a meta-analysis? *BMJ.* 2015 Aug;351:h4419.
41. Matts JP, McHugh RB. Analysis of accrual randomized clinical trials with balanced groups in strata. *J Chronic Dis.* 1978;31(12):725-40.
42. Rückbeil MV, Hilgers R-D, Heussen N. Randomization in survival studies: An evaluation method that takes into account selection and chronological bias. *PLoS One.* 2019;14(6):e0217946.
43. Krishna R, Maithreyi R SK. Research bias: a review for medical students. *J Clin Diagnostic Res.* 2010;4:2320-4.
44. Sackett DL. Clinician-trialist rounds: 5. Cointervention bias – how to diagnose it in their trial and prevent it in yours. *Clin Trials.* 2011 Aug;8(4):440-2.
45. Murray DM. Statistical models appropriate for designs often used in group-randomized trials. Vol. 20, *Statistics in medicine.* England; 2001. p. 1.373-85.
46. Simmons N, Donnell D, Ou S-S, Celentano DD, Aramrattana A, Davis-Vogel A, Metzger D, Latkin C. Assessment of contamination and misclassification biases in a randomized controlled trial of a social network peer education intervention to reduce HIV risk behaviors among drug users and risk partners in Philadelphia, PA and Chiang Mai, Thailand. *AIDS Behav.* 2015 Oct;19(10):1.818-27.
47. Packer DL, Mark DB, Robb RA, Monahan KH, Bahnson TD, Poole JE, et al. Effect of Catheter Ablation vs Antiarrhythmic Drug Therapy on Mortality, Stroke, Bleeding, and Cardiac Arrest Among Patients With Atrial Fibrillation: The CABANA Randomized Clinical Trial. *JAMA.* 2019 Apr;321(13):1261-74.
48. Raffle AE, Gray J. M. *Screening: Evidence and Practice.* Oxford: Oxford University Press; 2007.
49. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ.* 2019 Aug;366:l4898.
50. Page M, Sterne J, Higgins J O behalf of the development group for R-M. The ROB-ME (Risk Of Bias due to Missing Evidence) tool: a new tool for assessing reporting biases in



evidence syntheses. In: Abstracts of the 25th Cochrane Colloquium. Edinburgh, UK: Cochrane Database of Systematic Reviews; 2018.

51. Sterne JAC, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016 Oct;355:i4919.
52. McGuinness LA, Higgins JPT. Risk-of-bias VISualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. *Res Synth Methods*. 2020 Apr;n/a(n/a).
53. Altman DG. Comparability of randomised groups. *J R Stat Soc Ser D (The Stat)*. 1985;34(1):125-36.

# 5. A ANÁLISE BAYESIANA DOS ESTUDOS CIENTÍFICOS

## Introdução

Imagine a seguinte situação: você aprende, na graduação, que determinado remédio está contraindicado em uma doença porque o mecanismo farmacológico dela tem um efeito específico que não combina bem com a fisiopatologia dessa doença. Depois, já graduado, você se vê criticando os colegas que usam o medicamento por não conhecerem o que você pensa ser um mecanismo farmacológico tão básico, que aprendeu na graduação.

Mas, um certo dia, você abre a sua caixa de e-mails, e toma conhecimento de um novo estudo, o qual afirma que, testado na população, aquele medicamento, na verdade, contrariando o mecanismo farmacológico, faz bem.

Como agir diante dessa situação? Devemos esquecer o mecanismo farmacológico e fisiopatológico? Devemos acreditar 100% no novo artigo? A resposta correta é: devemos fazer uma análise “bayesiana” sobre esse tema.

## O que é essa tal análise bayesiana e o que isso tem a ver com o livro?

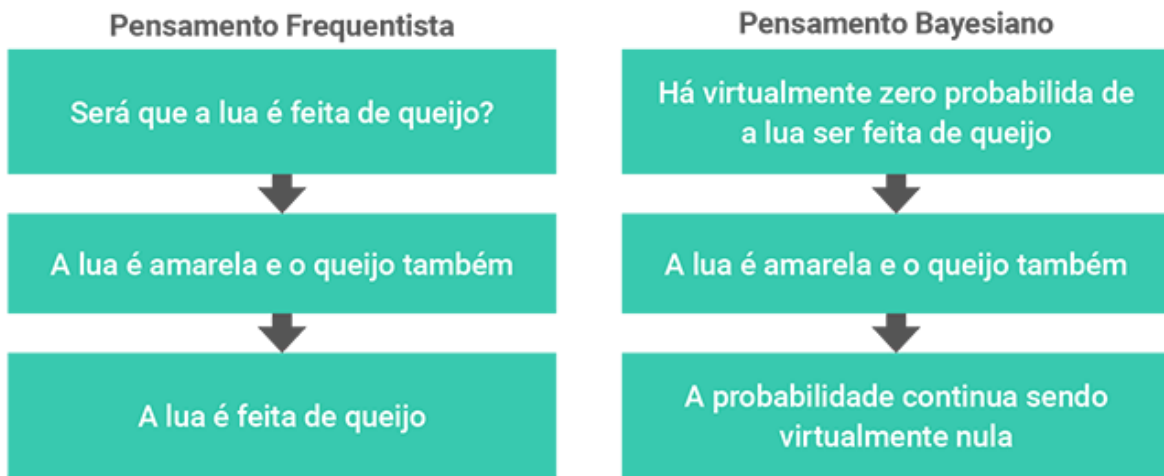
O pensamento bayesiano é a aplicação de uma fórmula para atualizar a evidência científica sob a luz das novidades: é a probabilidade condicional. Sendo assim, sempre que quisermos aprender algo e estivermos usando o pensamento bayesiano, nós precisamos nos perguntar o que já se conhece sobre o tema para usar aquilo como “ponto de partida” ou “probabilidade prévia”.<sup>1</sup>

Isso é oposto ao que se escreve em notícias de jornal, ou se faz pensar durante a nossa graduação, quando ouvimos que “um novo estudo afirmou que o que se fazia antes era errado”. Esse pensamento, denominado “frequentista”, é diferente do pensamento “mecanicista” (aquele em que o médico se baseia apenas em mecanismos fisiopatológicos para tomar as suas conclusões), já citado no Capítulo 1, porque os frequentistas até leem pesquisas clínicas, mas, usualmente, interpretam superficialmente seus achados: “se um novo estudo negou o anterior, preciso acreditar no mais novo porque é mais novo” ou, no máximo, tentar definir qual dos estudos foi metodologicamente superior?

O pensamento bayesiano não apenas tenta definir qual foi o estudo metodologicamente superior, mas vai além: ela se pergunta “como esse novo estudo, com as vantagens e as desvantagens da metodologia escolhida pelos pesquisadores, adiciona à literatura já existente?”. O bayesiano acredita na probabilidade condicional.

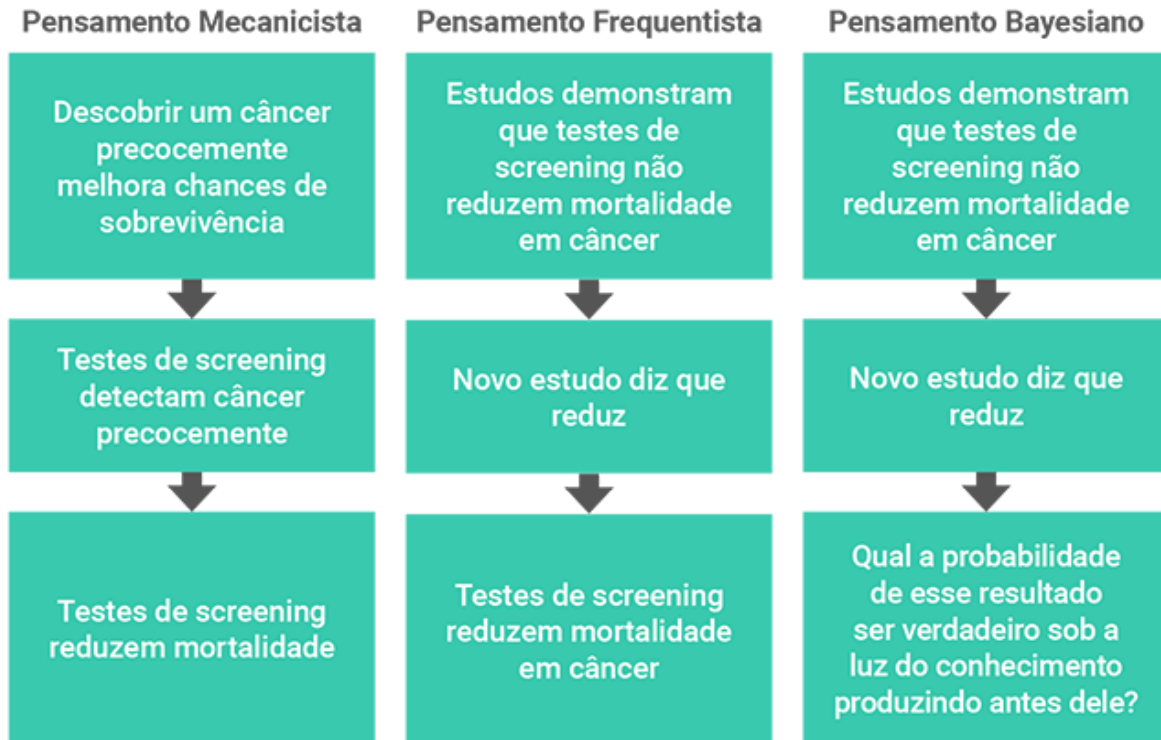
Uma comparação básica entre pensamento frequentista e pensamento bayesiano é demonstrada na Figura 1. A Figura 2 demonstra uma comparação entre um pensamento mecanicista, frequentista e bayesiano em Medicina.

**Figura 1.** Comparação entre modelo frequentista e modelo bayesiano de interpretação de evidências.



Perceba que o modelo frequentista tende a aceitar a nova evidência como verdadeira, enquanto o modelo bayesiano usa o conhecimento prévio sobre aquele tema (probabilidade pré-teste quase nula de que a lua seja feita de queijo) para definir se a probabilidade pós-teste deve ser aceita ou recusada.

**Figura 2.** Diferença entre os pensamentos mecanicista, frequentista e bayesiano.



Diferença entre o pensamento mecanicista, habitualmente encontrado em nossas graduações, quando se supervaloriza o pensamento fisiopatológico e farmacológico e se subvalorizam as pesquisas clínicas; o pensamento frequentista, em que um estudo, mesmo que com metodologia bem analisada, é suficiente para mudar o pensamento vigente; e o pensamento bayesiano, em que se analisa o novo estudo sob a luz do conhecimento prévio sobre aquele tema.

## Teorema de Bayes em poucas palavras e sem firulas

Imagine que você acordou em 1992 (era pré-internet) com vontade de fazer um piquenique. Mas abriu a cortina do seu apartamento e observou que estava um dia nublado. O que você decide fazer? Usar o seu conhecimento sobre o clima da sua cidade para estimar a chance de chuva, dado que está nublado.

O seu raciocínio foi:

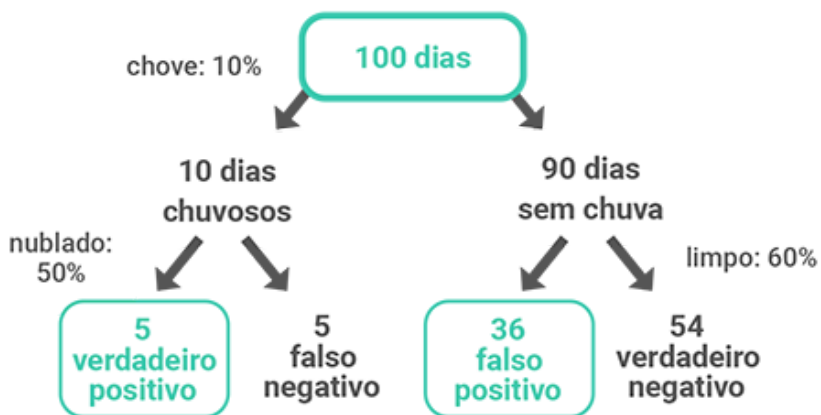
- Qual a chance de chover, visto que está nublado?
- Em 10% dos dias, chove em algum momento;
- Às vezes o dia amanhece nublado, mas não chove. Em 50% dos dias que amanhece nublado, chove.
- Também é verdade que em 60% dos dias que não chove, o dia amanhece limpo.

Perceba aqui a mudança de paradigma do pensamento frequentista para o bayesiano. É possível que o frequentista se contentasse com a informação de

que chove em 10% dos dias e, portanto, assumir 10% como a chance de chover hoje. Mas o frequentista entra em contradições e paradoxos facilmente porque se em 50% dos dias que amanhece nublado, chove, então deve haver 50% de probabilidade de chuva hoje. E aí? 10% ou 50%? O bayesiano interpreta todos esses dados ao mesmo tempo porque ele sabe que a resposta estará entre eles.

Veja como esse problema pode ser resolvido usando o teorema de Bayes (Figura 3).

**Figura 3.**



$$\text{Probabilidade de ser verdadeiro positivo} = \frac{\text{Verdadeiro positivo}}{\text{Todos positivos}}$$

$$\text{Probabilidade de ser verdadeiro positivo} = \frac{5}{5 + 36}$$

$$\text{Probabilidade de ser verdadeiro positivo} = 12\%$$

Construção do raciocínio bayesiano para responder qual a chance de chuva no dia que está nublado, sabendo que a chance global de chuva é de 10%, a chance de chover quando está nublado é de 50% e a chance de não chover quando não está nublado é de 60%.

## Onde isso entra na análise de artigos científicos?

De várias maneiras. A iniciar, pela própria formulação e tentativa de confirmação de uma hipótese. Conforme você leu no Capítulo 3, para que uma hipótese sobre uma nova terapia ganhe força, o ideal é que ela passe primeiro pelo crivo dos critérios de Bradford Hill.

Então vamos lá. Um cientista elabora a hipótese que o medicamento “Evidenciolol” trata uma doença. Ele se baseia no mecanismo fisiopatológico (plausibilidade biológica, um dos critérios de Bradford Hill) para a elaboração dessa ideia. Para um mecanicista, a questão acabaria aí: o mecanicista pensa que apenas a plausibilidade já justifica o uso de uma terapia.

O frequentista e o bayesiano, contudo, sabem que “entre as moléculas e a clínica do paciente há mais do que a vã Medicina pode imaginar”. E, então, aguardarão por estudos científicos clínicos, de preferência metodologicamente bem fundamentados, para tirar suas conclusões.

O estudo finalmente é lançado: foi um estudo observacional retrospectivo em que se percebeu que os pacientes que usaram o medicamento tinham uma menor chance de morrer do que os pacientes que não usaram. O estudo obteve um  $p < 5\%$ , que significa que há menos de 5% de chance de falso positivo **neste estudo**.

O frequentista interpretará esse dado como verdadeiro, com 95% de probabilidade (visto que o valor de  $p$  foi  $< 5\%$ ), até que venha um novo estudo que negue o primeiro (desde que ele não caia no “viés de ancoragem”, que é ancorar-se na primeira impressão e fechar-se para novidades).

O Quadro 1 traz um spoiler do que são *power* e valor de  $p$ . E o bayesiano vai muito além dessa interpretação, ele analisa os novos dados sob a luz dos prévios.

**Quadro 1. Para nunca mais esquecer: *power* e valor de  $p$ , erro tipo 1 e tipo 2.**

Concordamos que a literatura científica podia ser mais palatável. Por isso, vamos resumir breve e didaticamente esses termos e deixar, aqui, uma analogia que pode ajudar você a entender de uma vez por todas essas questões.

Erro tipo 1 é quando o pesquisador encontra um resultado falso positivo. Erro tipo 2 é quando o pesquisador encontra um resultado falso negativo.

Imagine que 100 pessoas com tromboembolismo pulmonar realizam uma tomografia de tórax com contraste. Em 10% delas, a tomografia foi erroneamente normal, portanto, falso negativas. Se isso fosse um estudo científico, se chamaria “erro tipo 2”. Se em 10% houve erro, significa que em 90% houve acerto. A capacidade de um teste ser verdadeiro-positivo é chamada de “sensibilidade” em Medicina, mas em pesquisa se chama “*power*”. O *power* normalmente buscado em pesquisas médicas é de 0,8 ou 80% (ou seja, se aceita um máximo de 20% de falsos negativos ou erros do tipo 2).

Agora imagine que 100 pessoas sem tromboembolismo pulmonar realizaram uma tomografia de tórax com contraste. Em 10% delas, a tomografia erroneamente detectou tromboembolismo pulmonar, um resultado falso positivo. Se isso fosse um estudo científico, se chamaria “erro tipo 1” ou erro  $\alpha$ . Se em 10% houve erro, então em 90% houve acerto. A capacidade de um teste ser verdadeiro negativo é chamada de “especificidade” em Medicina. Em pesquisas, é chamado de  $1 - \alpha$ .

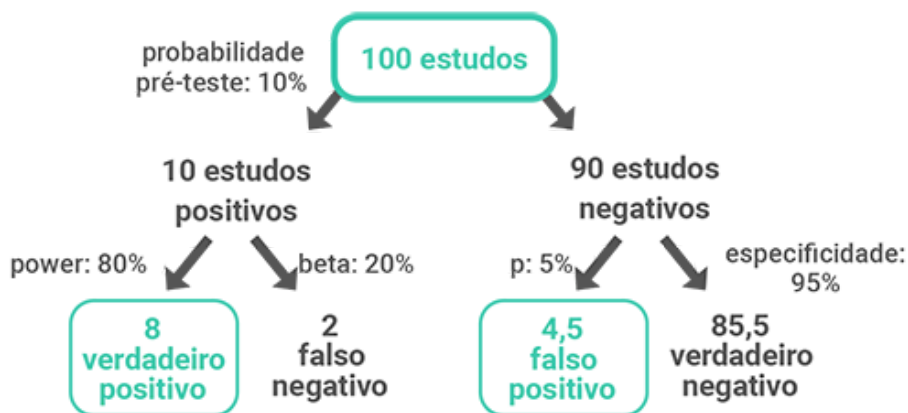
Agora vamos usar os mesmos conceitos na análise de artigos científicos. Suponha que você deseja descobrir qual entre dois antidiabéticos tem maior eficácia no tratamento do diabetes. Para isso, você realiza um estudo e os seguintes cenários podem se apresentar:

- Você conclui que os antidiabéticos são igualmente eficazes e eles, de fato, são. Você aceitou a hipótese nula (a de que não há diferença) e, como ela é verdadeira na vida real, nenhum erro foi cometido.
- Você conclui que um antidiabético é melhor que o outro, mas na realidade eles são igualmente eficazes. Você, equivocadamente, cometeu um erro do tipo 1. Em outras palavras, um falso positivo.
- Você conclui que os tratamentos são igualmente eficazes, mas, na vida real, um dos tratamentos é mesmo mais eficaz. Você cometeu um erro do tipo 2, que é o falso negativo.
- Você conclui que um dos antidiabéticos é mais eficaz e, na vida real, isso está correto. Você rejeitou corretamente a hipótese nula (de que não havia diferença). Parabéns, você não cometeu nenhum erro!

## Indo muito além do valor de p

A maioria das pessoas que lê artigos médicos e, até mesmo muitas das que fazem boas análises de artigos científicos, é frequentista. É basicamente assim: se o estudo for metodologicamente bom, se aceita o valor de p como a chance máxima de falso positivo daquele achado. Se o valor de p for  $< \alpha$  (uma convenção que será entendida melhor no Capítulo 6), então é  $> 95\%$  a probabilidade de ser um estudo verdadeiro. O bayesiano entende que “em um mundo de probabilidades, não há certezas, mas diferentes graus de incerteza” e usa o mesmo gráfico já demonstrado para fazer a sua interpretação da probabilidade pós-teste de o artigo ser verdadeiro usando uma probabilidade pré-teste de 10%, um *power* de 80% e um p de 0,05 ou 5% (valores tradicionalmente perseguidos em Medicina) (Figura 4).

Figura 4.



$$\text{Probabilidade de ser verdadeiro positivo} = \frac{\text{Verdadeiro positivo}}{\text{Todos positivos}}$$

$$\text{Probabilidade de ser verdadeiro positivo} = \frac{8}{8 + 4,5}$$

$$\text{Probabilidade de ser verdadeiro positivo} = 64\%$$

Aquele que sabe que o mundo é feito de probabilidades e incertezas. Para interpretar corretamente o valor de p (a possibilidade máxima admitida de falso positivo naquele estudo), deve se levar em consideração

que esse dado, sem levar em consideração os dados do mundo real, é meramente empírico. Então, o bayesiano parte da probabilidade pré-teste e verifica qual a chance de o estudo ser verdadeiro positivo agora que obteve um p de 0,05.

O Quadro 2 ensina, da maneira mais fácil que você já viu, como é formado o teorema de Bayes e quais seus componentes. Não deixe de lê-lo.

**Quadro 2. Facilitando mais uma: entendendo o teorema de Bayes de uma vez por todas.**

“Ah, lá vem o José falar em fórmula”. Olha, eu sou do time que acha que as fórmulas matemáticas poderiam ser ensinadas de um jeito mais didático. Então vamos ver se eu, um simples médico, consigo atingir esse objetivo lhe explicando o teorema de Bayes.

Você já viu nas Figuras 3 e 4 qual é o fluxograma que eu desenho para obter valor de probabilidade pós-teste. Vamos usar esse mesmo fluxograma para ensinar como funciona a fórmula.

**Figura do Quadro 2. O melhor que eu pude fazer para tentar lhe ensinar o teorema de Bayes.**

**Pergunta: qual a chance de meu estudo ser verdadeiro positivo?**

- Passo 1: veja quais os positivos entre os verdadeiros. Isso se consegue assim (pontilhado cinza):
  - » Multiplique o power (ou sensibilidade) pela probabilidade pré-teste.
- Passo 2: veja quantos são positivos entre os falsos. Isso se consegue assim (pontilhado laranja):
  - » Multiplique o  $\alpha$  (ou 1 - especificidade) pelos falsos (ou 1 - probabilidade pré-teste).
- Passo 3: veja quantos são verdadeiros. Isso se consegue assim (retângulos laranjas):
  - » Divida os verdadeiros positivos pela soma dos verdadeiros + falsos positivos.
- Passo 4: Jogue tudo em uma fórmula:
  - » Divida o passo 1 pela soma do passo 1 com o passo 2.

Probabilidade de ser verdadeiro positivo =  $\frac{\text{Verdadeiro positivo}}{\text{Todos os positivos}}$

Fórmula =  $\frac{\text{Probabilidade pré-teste} \times \text{power}}{\text{Probabilidade pré-teste} \times \text{power} + [(1 - \text{esp}) \times (1 - \text{probabilidade pré-teste})]}$

A sorte é que você não vai precisar usar essa fórmula na vida real. Por exemplo, pode usar um site como o Medical Teste Calculator ([calculator.testingwisely.com](http://calculator.testingwisely.com)). Lá você escreve o valor da probabilidade pré-teste, a sensibilidade (que será o *power* da pesquisa escrito na seção “metodologia”) e a especificidade (1 -  $\alpha$ ); e o site dá o valor da probabilidade pós-teste. Pode ser usado em exames diagnósticos também (você vai aprender no Capítulo 8).

E como saber a probabilidade pré-teste? Bem, eu falei que o mundo é feito de probabilidades e incertezas. Então, a análise bayesiana de artigos, muitas vezes, começa de um dado subjetivo, que numericamente não existe. Mas não se preocupe, pois eu vou lhe dar umas dicas nos próximos capítulos.



Na Figura 4 você viu algo que, até então, era inimaginável acontecer: usando um teorema, eu transformei a probabilidade de a **pesquisa** ser falsa de, no máximo, 5% em uma probabilidade de o **achado** ser falso de 36% (100-64%). Observe que deixei em negrito duas palavras importantes que preciso que você tenha atenção: o valor de  $p$  fala sobre a chance de aquela pesquisa ser falso positiva, mas não fala sobre a chance de aquele achado ser verdadeiro no mundo real. Observe na Tabela 1 o quanto de probabilidade pós-teste se consegue cruzando valores de  $p$  com probabilidade pré-teste usando sempre 80% de *power*. É o uso da probabilidade condicional.

**Tabela 1.**

$p$	Pré	Pós	Pré	Pós	Pré	Pós	Pré	Pós	Pré	Pós	Pré	Pós
0,1	5%	30%	10%	47%	20%	67%	30%	77%	40%	84%	50%	89%
0,05	5%	46%	10%	64%	20%	80%	30%	87%	40%	91%	50%	94%
0,01	5%	81%	10%	90%	20%	95%	30%	97%	40%	98%	50%	99%
0,005	5%	89%	10%	95%	20%	98%	30%	99%	40%	>99%	50%	>99%
0,001	5%	98%	10%	99%	20%	>99%	30%	>99%	40%	>99%	50%	>99%
0,0001	5%	>99%	10%	>99%	20%	>99%	30%	>99%	40%	>99%	50%	>99%
0,00001	5%	>99%	10%	>99%	20%	>99%	30%	>99%	40%	>99%	50%	>99%

Probabilidade pós-teste após cruzar o valor de  $p$  com a probabilidade pré-teste de uma pesquisa com 80% de *power* (sensibilidade). Se a pesquisa tinha apenas 5% de probabilidade pré-teste, mesmo com um valor de  $p$  de 0,05, a chance de esse resultado ser real não é de 95%, mas de 46%. Mais uma consequência disso é a de que estudos muito plausíveis continuam com alta probabilidade pós-teste mesmo depois de estudos em que o valor de  $p$  é “negativo” (em Medicina  $> 0,05$ ): por exemplo: 50% de probabilidade pré-teste com  $p$  de 0,1 ainda tem 89% de chance de ser verdadeiro.

## Indo muito mais além do valor de $p$

Vamos ainda mais longe. Já vimos que o *power* do estudo é  $1 - \beta$  (ou  $1 -$  erro tipo 2) e que isso pode ser lido como a “sensibilidade” do teste: a chance de ele encontrar relações verdadeiras. Normalmente se aceita um *power* de 80% (ou seja, um  $\beta$  máximo de 20%). Vimos que também que  $p$  é o valor máximo de erro tipo 1 que se aceita na maioria dos estudos em Medicina: 0,05 ou 5% (dá no mesmo, matemáticos de plantão). É, em analogia a exames clínicos, a especificidade: a chance de que se encontre um falso positivo na pesquisa.

O que acontece se a pesquisa contiver muitos vieses (aqueles que vimos no Capítulo 4)? Os erros do tipo 1 e do tipo 2 serão potencializados e o *power* e a especificidade da pesquisa se reduzirão quanto mais importantes forem os vieses (vide Tabela 2). A potencialização dos erros tipo 1 e 2 também ocorrerá

quanto mais pesquisas existirem sobre um tema porque aumenta a chance de que um resultado significativo ocorra ao acaso.

Levando esses fatos em conta, e colocando tudo em cálculos formidáveis, John Ioannidis afirma ser possível provar matematicamente que “a maioria dos estudos científicos é falsa” em seu clássico e seminal artigo.<sup>2</sup> A Tabela 2, adaptada desse artigo, traz diversos tipos de pesquisas (daquelas que vimos no Capítulo 3), sua probabilidade pré-teste e sua probabilidade pós-teste, agora contando com os possíveis vieses.

**Tabela 2.**

Power (1 - $\beta$ )	Pré	Viés	Exemplo prático	Pós
80%	50%	10%	Ensaio clínico randomizado com <i>power</i> adequado, pouca incidência de viés e uma probabilidade pré-teste de 50%	85%
95%	66%	30%	Meta-análise confirmatória usando estudos de boa qualidade de uma boa hipótese (probabilidade pré-teste de 66%).	85%
80%	25%	40%	Meta-análise de pequenos estudos inconclusivos de uma hipótese moderada (25% de probabilidade pré-teste).	41%
20%	16%	20%	Ensaio clínico de fase I ou II com pouco <i>power</i> , mas com boa metodologia de uma hipótese em ascensão (16% de probabilidade pré-teste).	23%

Probabilidade pós-teste de várias combinações de estudos com seus *powers* estimados, vieses e probabilidades pré-teste. O valor de probabilidade pré-teste indica a chance de aquele resultado ser um verdadeiro positivo.

A Tabela 3 traz um breve exercício de imaginação para analisar novamente a probabilidade pós-teste, mas, agora, tendo em conta a incidência dos vieses nas pesquisas. Compare como o valor cai em comparação aos da tabela 1.

**Tabela 3.**

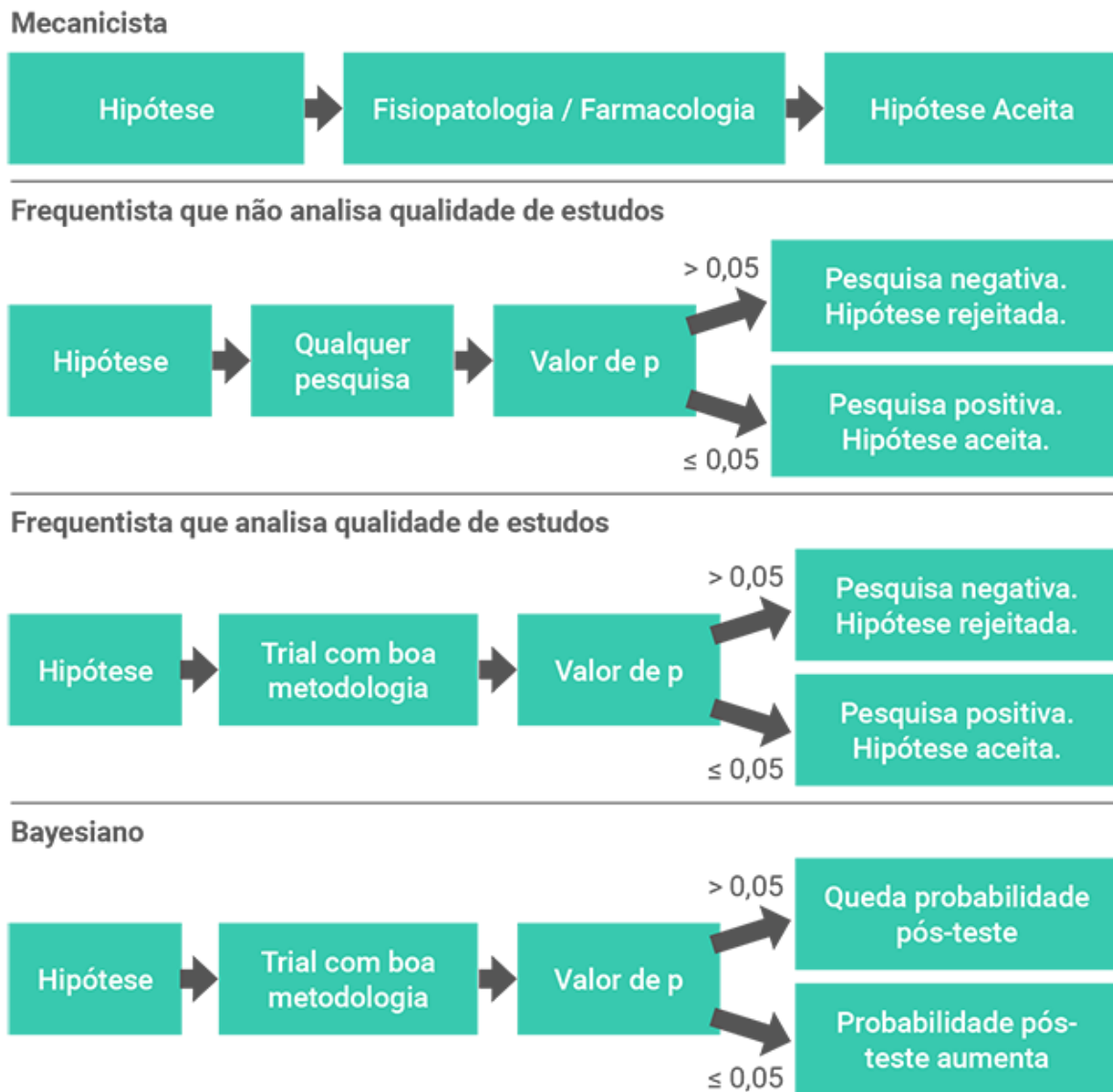
p	Pré	Viés	Pós	Viés	Pós	Pré	Viés	Pós	Viés	Pós
0,1	10%	10%	30%	30%	14%	50%	10%	79%	30%	60%
0,05	10%	10%	36%	30%	16%	50%	10%	83%	30%	63%
0,01	10%	10%	42%	30%	17%	50%	10%	87%	30%	65%
0,005	10%	10%	43%	30%	17%	50%	10%	87%	30%	65%
0,001	10%	10%	44%	30%	17%	50%	10%	88%	30%	65%

Probabilidade pós-teste após cruzar o valor de p com a probabilidade pré-teste de uma pesquisa e também acrescentando o peso de viés que reduzirá em 10% ou em 30% o *power* e a especificidade da pesquisa (incidindo igualmente nas duas para fins de cálculos). Observe que mesmo uma pesquisa com p de 0,05 ou menor, mas com 10% de vieses, chega a uma probabilidade pós-teste de apenas 36%, tendo ainda 64% de chance de resultados falsos positivos.

# Como incorporar o pensamento bayesiano na análise de pesquisas?

Como vimos, o pensamento bayesiano consiste em interpretar um novo achado sob a luz do conhecimento prévio, e não apenas o mecanismo fisiopatológico/farmacológico ou a qualidade da metodologia da pesquisa (Figura 5).

**Figura 5.**



Diferenças de interpretação de uma evidência por um mecanicista, um frequentista e um bayesiano.

O exercício que você leitor vai fazer é o seguinte: ao ler o resultado de uma pesquisa do tipo “novo medicamento se comprova eficaz contra cardiomiopatia hipertrófica”, os passos que você vai tomar serão:

- a. Procure saber qual a probabilidade pré-teste de aquilo funcionar de verdade. Leia na introdução do artigo qual a plausibilidade para seu uso, se há coerência, consistência... Poderá ser necessário fazer uma breve revisão de literatura sobre o tema para saber se há controvérsias em seu mecanismo.
- b. Estime uma probabilidade pré-teste: se a hipótese for muito convincente (ou tiver muitos critérios de Bradford Hill e outras pesquisas apontando para a mesma direção – consistência), pense em algo próximo de 50%. Se não for muito convincente, mas o estudo já passou pelas fases I e II e agora está na fase III, ou estudos observacionais apontaram benefício, use algo em torno de 20%. Se a pesquisa caiu de paraquedas e não tem nada plausível ou consistente, considere algo em torno de 1% a 5%.
- c. Analise o artigo criticamente (você verá os fundamentos dessa análise no Capítulo 7). Observe o *power* e o valor de *p* encontrados. Observe também a quantidade de vieses e, também, qual a magnitude que eles podem ter tido no resultado final da pesquisa.
- d. Para obter o valor da probabilidade pós-teste, use o teorema de Bayes ensinado no Quadro 2.
- e. Caso não queira fazer fórmulas sempre que ler um artigo, pelo menos lembre que você deve ir muito além do valor de *p* na sua análise crítica, e isso significa não só captar os pormenores (Capítulo 7) e os vieses (Capítulo 4) da pesquisa, mas realizar uma análise daquele resultado partindo do conhecimento prévio sobre aquilo (por exemplo, usando os critérios de Bradford Hill).

## Referências bibliográficas

1. Bittl JA, He Y. Bayesian Analysis: A Practical Approach to Interpret Clinical Trials and Create Clinical Practice Guidelines. *Circ Cardiovasc Qual Outcomes*. 2017 Aug;10(8).
2. Ioannidis JPA. Why most published research findings are false. *PLoS Med*. 2005 Aug;2(8):e124.

# 6. ESTATÍSTICA PARA QUEM NÃO GOSTA DE MATEMÁTICA

## Introdução

Uma das razões pelas quais nós, no auge da nossa ignorância adolescente, escolhemos Medicina ou outra profissão de saúde como a nossa atividade laboral eterna é a doce ilusão de que: a) após o vestibular, não precisaremos mais estudar tanto; b) teremos prestígio e um bom salário; c) não precisaremos estudar matemática ou física.

Eu, José Alencar, um dos autores deste capítulo, eletrofisiologista e adepto à Medicina Baseada em Evidências, posso dizer que as frases contidas nas letras (a) a (c) do parágrafo anterior não passam, de fato, de uma ilusão. E aqui estamos nós, mais uma vez tentando falar de matemática para quem quer distância de cálculos.

Entendendo nossa dificuldade (“nossa” porque eu também me encaixo nesse perfil), decidi criar um capítulo muito objetivo para demonstrar da maneira mais fácil possível os conceitos e os cálculos mínimos necessários para a boa leitura de artigos científicos. Eu os descrevi da maneira como eu entenderia – e se eu posso entender, acredite, você também pode.

E, para a escrita deste capítulo, contei com a preciosa ajuda do Felipe Nogueira, um daqueles caras que não têm qualquer problema com números, pelo contrário, lida tão bem com eles quanto nós achamos que lidamos com pacientes.

## Os primeiros passos – entendendo porcentagens

Ninguém consegue entender estatística sem antes entender bem a “porcentagem”. É assim: divide-se o número de pessoas com determinado desfecho pelo total de pessoas.

Por exemplo, 6 pessoas num total de 50 acabaram sendo infectadas por um novo vírus. A porcentagem de pessoas infectadas é de  $6/50 = 0,12$ . O número fracionado 0,12 significa 12%. É só multiplicar por cem para encontrar uma “porcentagem”. Assim, 0,06 significa 6%. E 0,002 significa 0,2%.

Um adendo importante é o de que a literatura em inglês não usa vírgulas para separar os números decimais, eles usam pontos. Então, ao ler um bom artigo em inglês você lerá 0.34 como 34%.

A atenção que eu quero que você, leitor, tenha é que o dado “porcentagem” pode esconder informações valiosas. Por exemplo, se uma nova doença pandêmica tem potencial de matar 2% da população, esse número pode parecer pequeno e a doença pode parecer não ser tão mortal assim. Mas, 2% de 209 milhões de pessoas, que é a população do Brasil, significa que essa doença tem o potencial de matar quase 4 milhões e 200 mil pessoas, o que seria força suficiente para dizimar as cidades de Manaus e Curitiba.

Mais um exemplo de como as porcentagens, isoladamente, podem enganar: imagine que um determinado medicamento reduz a evolução para um desfecho indesejado em 33%. Parece bom. Isso com certeza vai estampar as propagandas do medicamento e muitos médicos ganharão canetas de cortesia para prescrevê-lo a fim de reduzir a chance de seu paciente ter esse desfecho indesejado. Mas a informação “33% de redução”, na verdade, não diz muita coisa. Você deve se perguntar: reduziu de quanto pra quanto? Se a redução foi de 0,03% a 0,02%, houve uma redução relativa de 33%. Porém, o risco de 3 desfechos a cada 10 mil pacientes foi reduzido para 2 desfechos a cada 10 mil pacientes. Uma redução tão ínfima, que talvez do ponto de vista do custo para a saúde pública não justifique a sua caneta nova.

Esses são apenas alguns exemplos em que se aplica a clássica frase: “existem três tipos de mentiras: as mentiras, as mentiras sujas e as estatísticas”. Não confundam a precisão dos números com exatidão: se eu disser que 32,3% das pessoas que tomaram um hipotético medicamento foram curadas enquanto apenas 12,4% das pessoas que não o tomaram se curaram, alguém pode interpretar esse dado como muito confiável pela sua precisão e exatidão, mas só há precisão aqui. As pessoas que tomaram esse medicamento podem, talvez, ter melhores condições financeiras (por isso o compraram) ou, quem sabe, um acesso a melhores farmácias, médicos e hospitais. Uma terceira hipótese seria a de que a maioria dessas 32,3% usou também um outro medicamento e ele sim é que foi benéfico. Ainda: o grupo em que se percebeu maior porcentagem de cura era de pessoas mais jovens que possuía menos comorbidades... O médico precisa estar preparado para perceber quais fatores confundidores podem ter sido responsáveis pelo

resultado que está lendo. Esse é um exercício que não deve ser feito nem só uma ou duas vezes, mas a cada pesquisa que você ler.

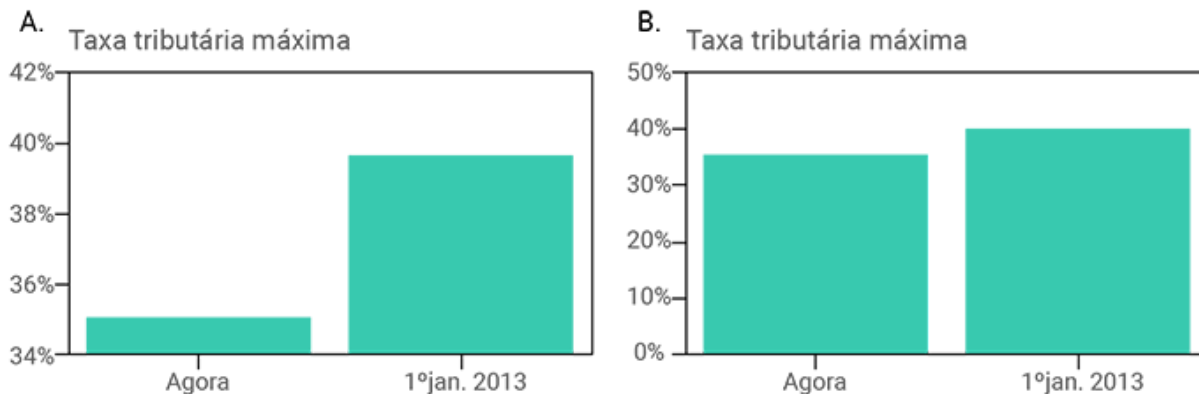
Também desconfie de como os dados foram obtidos: se um jornal anuncia a proporção de suicídios entre pessoas de determinado grupo minoritário, desconfie, porque há enorme dificuldade legista em se determinar algumas mortes como suicídio. E também porque, depois de morto, faltam dados que confirmem que aquela pessoa participava realmente de um grupo minoritário. Se uma estatística é composta de chutes, simplificações, equívocos, a sua conclusão é muito provavelmente falsa, apesar de precisa.

## Você não se complica pelo que você não sabe, mas pelo que você sabe com certeza, só que não é verdade

Frase de Mark Twain. Boa, não é? Antes de passarmos para o conhecimento das estatísticas, vamos ver algumas das maneiras com as quais as estatísticas podem nos enganar. Uma leitura adicional recomendada é do livro *O guia contra mentiras*, de Daniel Levitin.<sup>1</sup>

Começando pelos próprios percentuais, que parecem tão simples. Uma campanha de candidato pode modificar a escala em que a porcentagem das intenções de votos aparece na televisão (por exemplo, de 10 em vez de 50) para dar a impressão de que seu crescimento de 3 para 4 pontos percentuais foi relevante. Se ele quiser ir além, pode demonstrar em seu programa que teve um aumento de 33% nas intenções de votos (1 ponto a mais partindo dos 3 prévios é  $1/3$ , que dá 0,33 ou 33%), um número chamativo, mas ainda irrelevante. A Figura 1 traz mais uma ilustração desse exemplo.

### **Figura 1.**



Ao reduzir a escala de porcentagem de 50% para 8% (perceba que a escala do eixo x na figura A vai de 34% a 42%), se passa a impressão rápida de que a taxa tributária vai aumentar cinco a seis vezes. Em B, usando uma escala de 50% no eixo x (vai de 0 a 50), o leitor pode perceber o real aumento da taxa tributária de maneira mais realista.<sup>1</sup>

Outra forma bastante comum e fácil de mentir com gráficos é exibir dados cumulativos. Note que dados cumulativos, como o nome sugere, só aumentam. Essa tática foi utilizada em 2013 por Tim Cook, presidente da Apple, em uma de suas apresentações para mostrar o sucesso de vendas do iPhone. Mesmo com o título corretamente descrevendo que se tratava de vendas cumulativas, o gráfico não mostrava que as vendas caíram no período anterior à apresentação. Como outro exemplo, no início da pandemia da Covid-19, muitos criticavam a postura do governo federal americano pela falta de testes na população. Para mostrar que o governo havia aumentado o número de testes e que isso não era mais um problema, o governo americano apresentou um gráfico com dados cumulativos.

Em muitos dados nas pesquisas médicas e, especialmente aquelas que são veiculadas na mídia, falta plausibilidade. E isso é algo que você pode (e deve) fazer sozinho, desde que seja racional. É só pensar um pouco na informação antes de divulgar. Por exemplo, a informação amplamente divulgada nos veículos jornalísticos de que “150 mil meninas e mulheres jovens morrem anualmente por anorexia nos Estados Unidos” é extremamente implausível e impossível porque o número anual de mortes por qualquer causa em mulheres de 15 a 44 anos nos Estados Unidos é de 55 mil. Ora, se o dado “morte por qualquer causa” é a mortalidade global dessa população, como pode ser menor do que a morte por uma só causa?

## Medidas de tendência



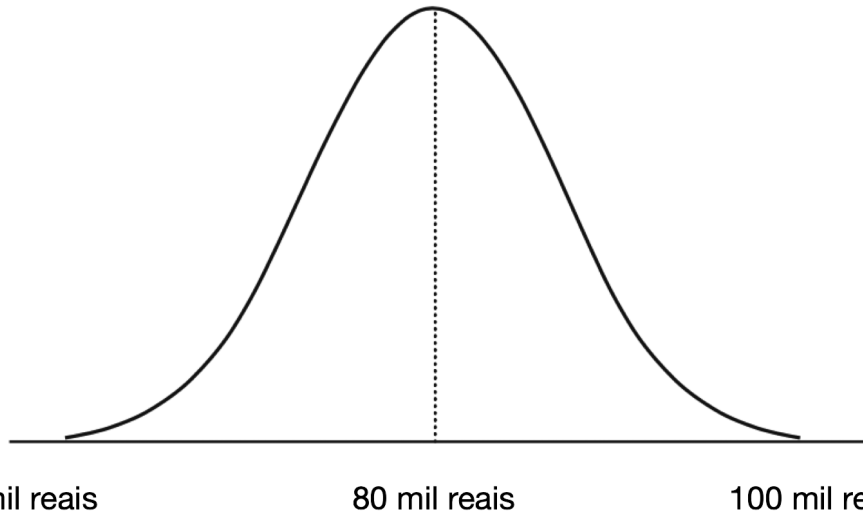
Tão simples, que se aprende na escola. O que não se aprende é que, na verdade, existem três medidas de tendência e é comum que elas entreguem resultados diferentes, então é óbvio que você precisa conhecer melhor esses cálculos e também aprender a evitar seus vieses.

Pois é. É que a palavra “média” acabou ganhando a fama, mas há várias maneiras de calcular a tendência central de um conjunto de dados. As mais conhecidas são a média, a mediana e a moda. Agora sim você lembrou do que estamos falando! Seu professor de epidemiologia falou isso em sala.

Vamos entender os cálculos e as importâncias de cada um. A primeira, como falamos, é a média. É calculada somando todos os dados observados e depois dividindo pelo número de dados observados. É uma média aritmética simples.

O problema da média simples é que ela não conta muito bem com *outliers*. Imagine que, em uma sala com 100 pessoas, a média da fortuna seja de 35 milhões de reais. Fabuloso, não? Mas não te contaram que essa sala está o Silvio Santos e mais 99 mendigos paupérrimos. Silvio Santos, nesse caso, é um *outlier*, alguém que atrapalha a média. Se você, pensando que nessa sala estariam muitas pessoas ricas, mandasse seus melhores vendedores, eles teriam que se digladiar para tentar convencer o Silvio Santos a comprar seu produto. Ao concluir algo sobre indivíduos baseado no seu conhecimento sobre o grupo, você teria caído em um viés chamado falácia ecológica (já comentado no Capítulo 3). A média, portanto, funciona melhor em dados distribuídos “normalmente” – quando os dados observados são distribuídos igualmente em cada lado do ponto médio (Figura 2).

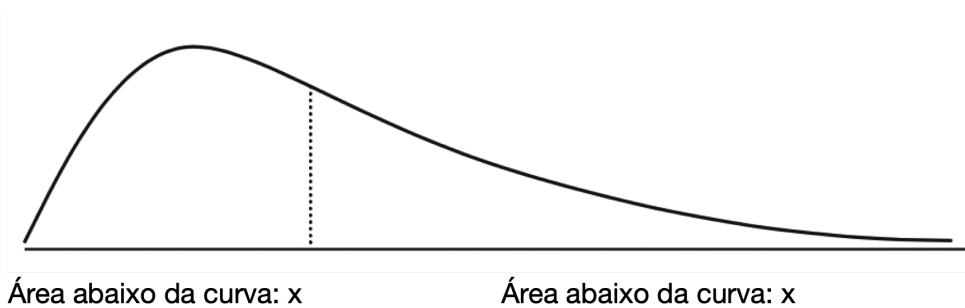
**Figura 2.**



Distribuição “normal”. Aqui, os dados observados são muito similares de um lado e do outro do ponto médio: observe que a variação é de 20 mil reais para cada lado. Oitenta mil reais é a média aritmética, caso seja calculada. A linha tracejada demonstra a média.

Mediana é o ponto em que metade dos dados observados são menores e a outra metade é composta por números maiores. Imagine que você está estudando sete pacientes e eles possuem 48, 54, 55, 59, 61, 62 e 98 anos. A mediana desses valores é 59 anos, a idade abaixo da qual estão metade das pessoas e acima da qual está a outra metade. Perceba que a média aritmética de idade seria 62.<sup>2</sup> A mediana lida melhor com outliers, apresentando, por exemplo, um gráfico como o da Figura 3.

**Figura 3.**

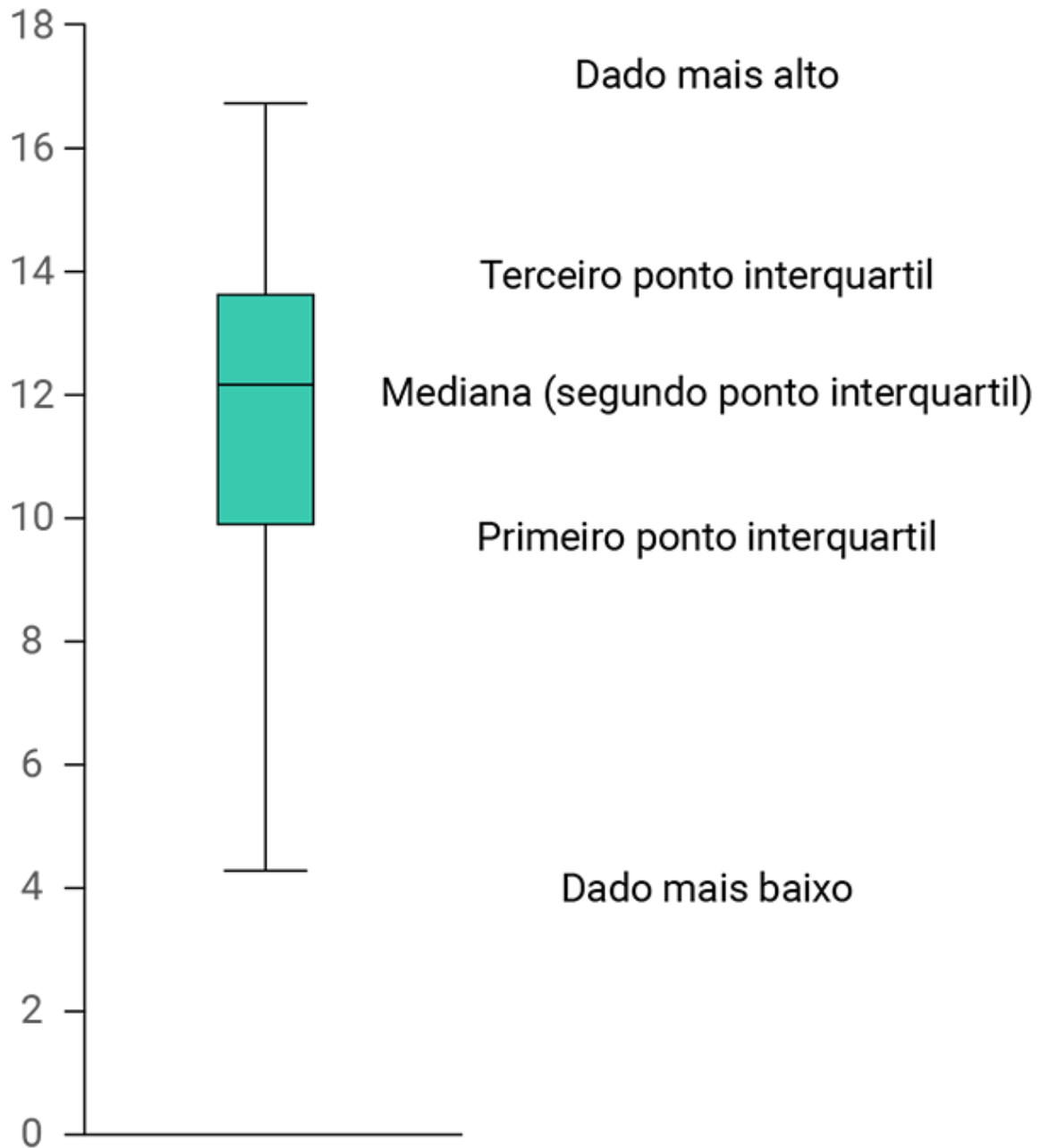


A mediana divide a amostra em dois e o ponto médio é exatamente o valor abaixo do qual estão metade dos dados e acima do qual está a outra metade. Por isso, representado no gráfico como áreas iguais abaixo da curva. A linha tracejada demonstra a mediana.

A mediana pode, ainda, ser apresentada no formato de “**distância interquartil**”. Quartil é um quarto, portanto 25% da amostra. Assim, dividindo por quatro, os estatísticos nos mostram o 1º e o 3º pontos

interquartis. O primeiro ponto interquartil divide o 1º interquartil do segundo, e o terceiro ponto interquartil divide o terceiro interquartil do quarto. O interquartil possui 50% da amostra e os gráficos são representados como na Figura 4.

**Figura 4.**

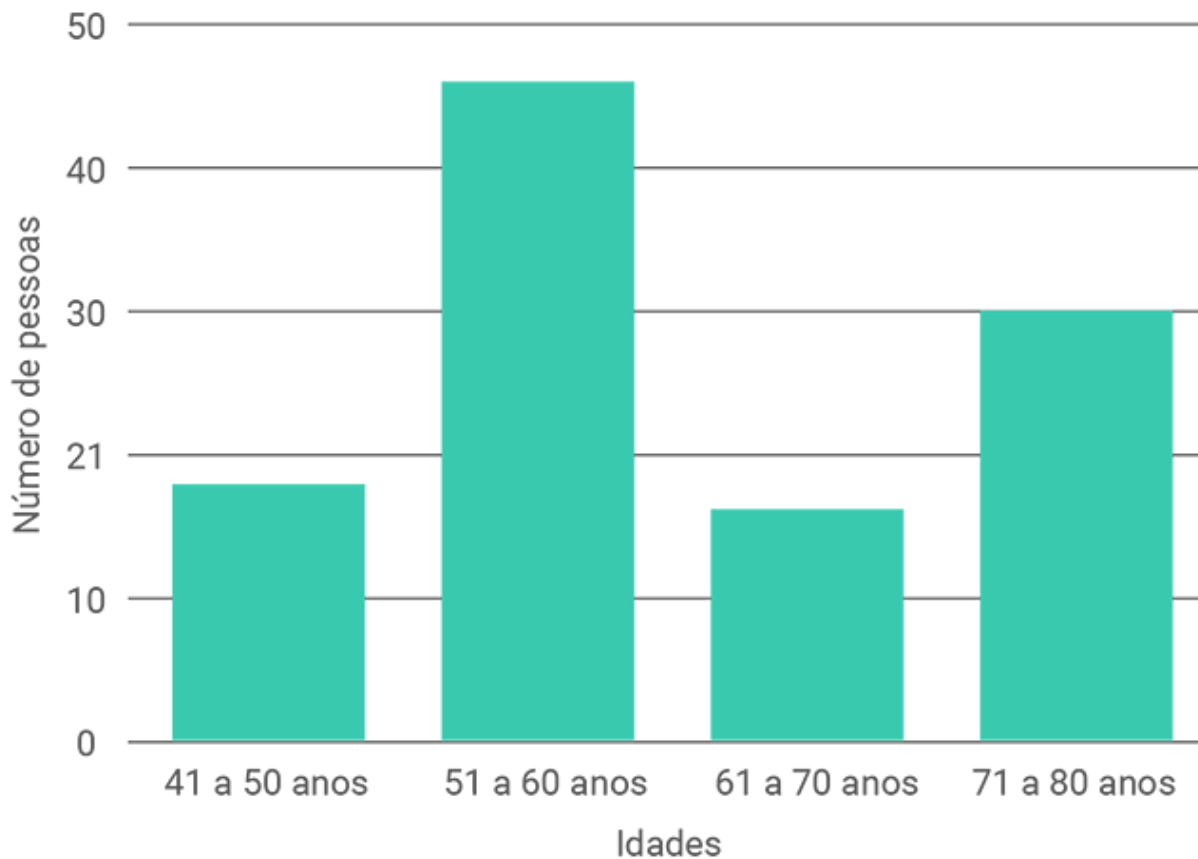


Pontos interquartis, muito similares à mediana, só que em vez de dividir metade dos dados acima e abaixo, ele divide em quatro quartos e apresenta os dados correspondentes destacando os interquartis

do meio: segundo e terceiro. O primeiro ponto interquartil dá início ao segundo interquartil e o terceiro ponto interquartil termina o terceiro interquartil.

A terceira forma de calcularmos o centro é a “moda”, que é o valor mais comumente encontrado em uma amostra, como na Figura 5.

**Figura 5.**



Nessa pesquisa, a moda é ter entre 51 e 60 anos, porque foi o resultado mais encontrado. Alguém pode, inclusive, dizer que houve um comportamento “bimodal” dos dados dessa pesquisa, porque a idade dos 71 a 80 anos também apresentou uma prevalência digna de nota.

Interessante perceber que, em média, os seres humanos do planeta possuem aproximadamente 1 testículo, a mediana é zero e a moda também (porque há mais mulheres que homens no mundo).

## Medidas de dispersão

A primeira medida de dispersão, isto é, como os dados se dispersam na amostra, já foi apresentada na seção anterior: as distâncias interquartis.

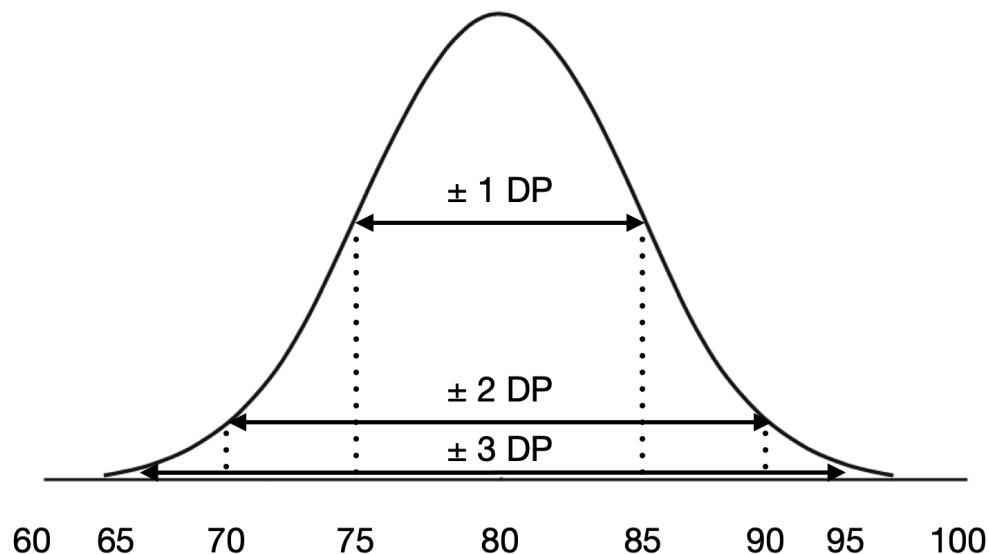
Agora vamos conhecer mais dois: o desvio padrão e os intervalos de confiança.

**Desvio padrão** deve ser usado em amostras cujos dados estão distribuídos “normalmente”, como o gráfico da Figura 2, que é também o momento mais adequado para se usar a média aritmética, lembra? Parece com o intervalo interquartil, mas em vez de dividir em quartis, ele divide da seguinte forma:

- a. Um desvio padrão ( $\pm 1$  DP) inclui 68,2% dos valores, sendo 34,1% de cada lado da média.
- b.  $\pm 2$  DP incluem 95,4% dos valores, metade de cada lado.
- c.  $\pm 3$  DP incluem 99,7% dos valores.

Esses números são universais. Desvios padrões são sempre assim (Figura 6). E, como você percebeu, funciona bem com distribuições “normais”.<sup>3</sup> Uma dica para saber se está diante de uma distribuição normal quando lhe apresentarem apenas números é: confira se  $\pm 2$  DP ainda traz um resultado contido em um intervalo lógico. Imagine que a média de dias hospitalizado por uma determinada doença é 10 e o desvio padrão é 8. Isso significaria o absurdo de que pessoas receberam alta seis dias antes mesmo de serem internados, porque  $10 - (2 \times DP) = 10 - (2 \times 8) = -6$ .

**Figura 6.**



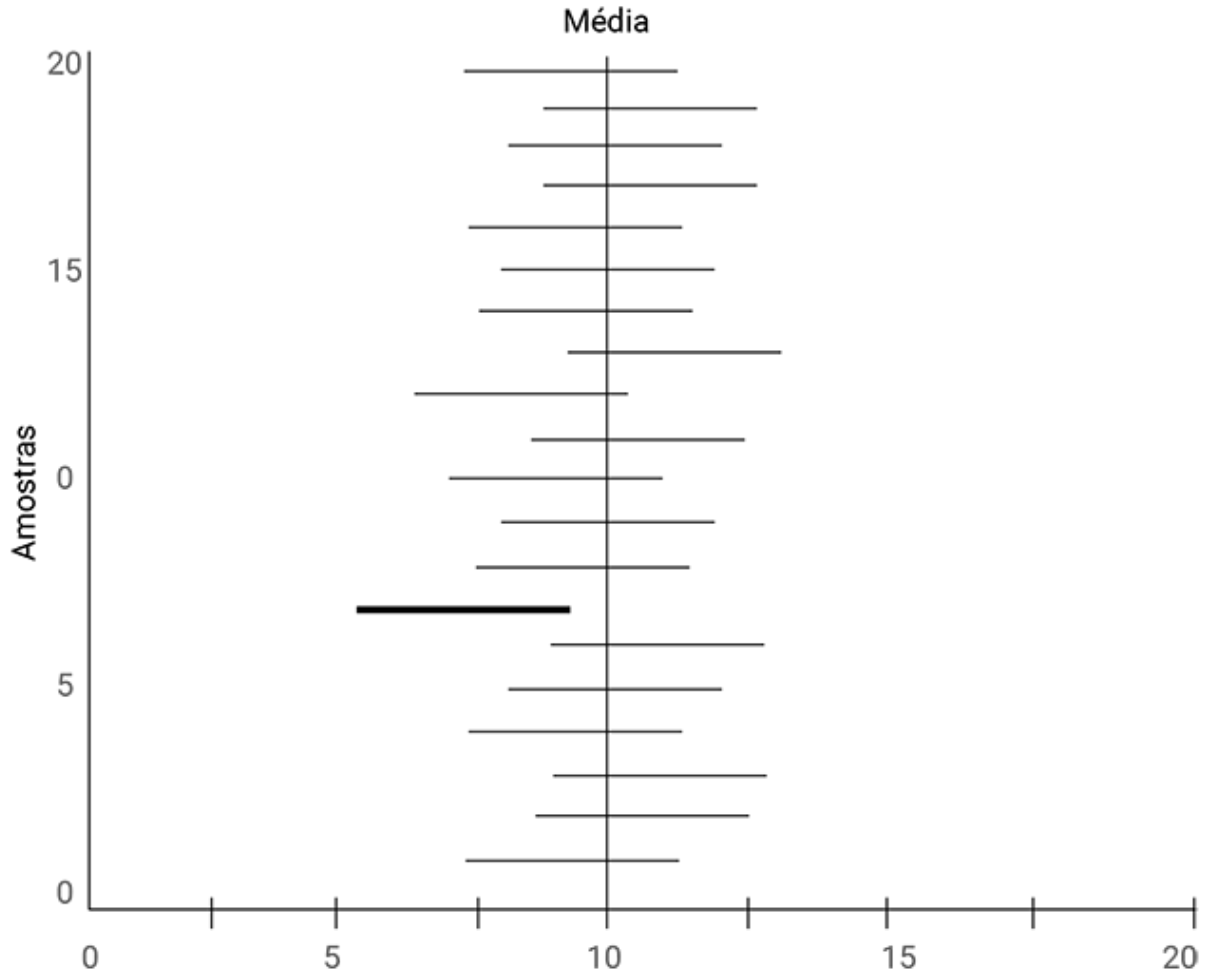
O mesmo gráfico da figura 2 de uma distribuição normal de salário anual com média de 80 mil reais. Pode se observar no gráfico que um desvio padrão é de 5 mil reais. Sendo assim, 68,2% das pessoas

dessa amostra ganham entre 75 mil e 85 mil reais. 95,4% das pessoas ganham entre 70 e 90 mil reais e 99,7% das pessoas ganham entre 65 e 95 mil reais.

Agora, vamos entender o **intervalo de confiança**. Observe que todas as medidas anteriores (média, moda, desvio padrão) se referem à amostra. Já o intervalo de confiança é um intervalo dentro do qual muito provavelmente estará aquele valor em questão caso tivéssemos testado toda a população e não apenas uma amostra dela.

Cada intervalo de confiança está associado com um nível de confiança, como 90%, 95% e 99%. Por exemplo, considere que um estudo uma nova terapia antidiabética reduziu a hemoglobina glicada em  $-0,8$  com um intervalo de confiança de 95% (IC 95%) entre  $-0,3$  e  $-1,0$ . Isso significa, *grosso modo*, que há 95% de probabilidade de que essa redução esteja mesmo entre  $0,3$  e  $1,0$ , caso aplicássemos o antidiabético em toda a população do estudo. De forma mais precisa, se calculássemos o IC 95% para diferentes amostras de uma mesma população, o real valor do efeito na população estará em 95% dos intervalos (Figura 7).<sup>4</sup>

**Figura 7.**



Intervalos de confiança de 95% de 20 amostras selecionadas aleatoriamente. Se for calculada a IC 95% da média dessas 20 amostras, aproximadamente 19 dessas 20 terão um intervalo de confiança que inclui a média da população.

Considere, agora, outro antidiabético que também reduz  $-0,8$  de hemoglobina glicada, mas com um IC 95% entre  $-1,0$  e  $0,2$ . Nesse caso, significa que o novo medicamento pode não apenas não reduzir, como aumentar a hemoglobina glicada.<sup>5</sup> Normalmente, quanto maior a amostra de um estudo, menor o seu intervalo de confiança (o primeiro exemplo provavelmente continha mais pacientes que o segundo, pois o intervalo de confiança abrangia  $0,7$  pontos de hemoglobina glicada no primeiro, e  $1,2$  no segundo).

Aqui já aprendemos um pouquinho de como funcionam os famosos *forest plots* das meta-análises (Figura 8).

**Figura 8.**

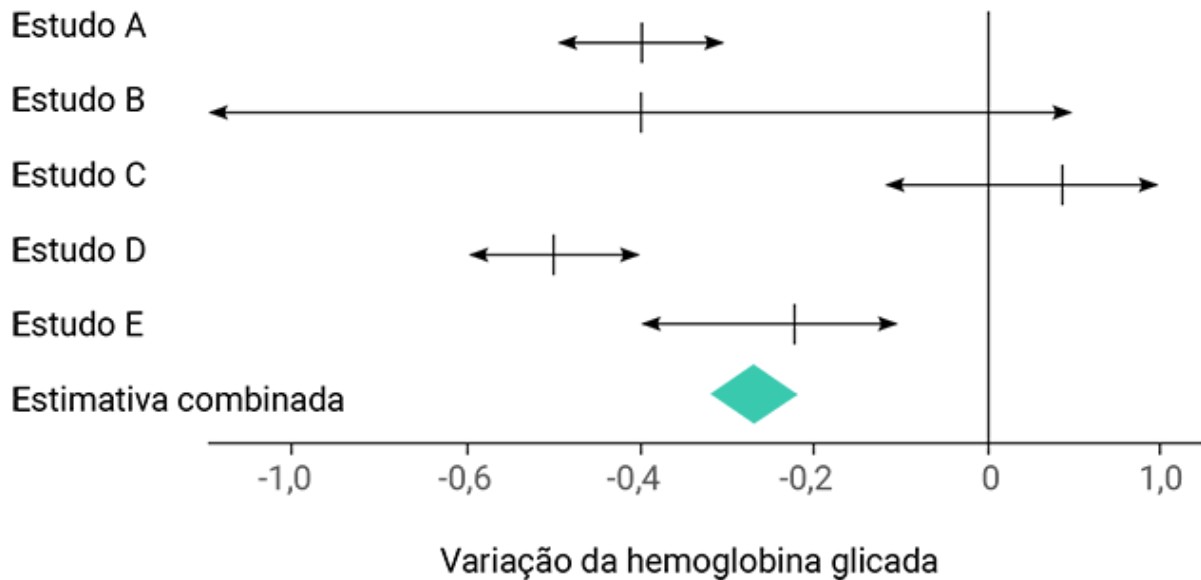


Gráfico do tipo “*Forest plot*” demonstrando intervalos de confiança da variação de hemoglobina glicada em cinco estudos. Quanto menor o estudo, maior seu intervalo de confiança, o que significa que o estudo B deve ser o de menor amostra. Perceba que o IC 95% dos estudos B e C cruza a linha do zero, o que significa que há chance de que os medicamentos testados nesses estudos não reduzam ou, pior, aumentem a hemoglobina glicada. O cálculo da estimativa combinada é muito utilizado em meta-análises.

## Comparando tendências e dispersões

Agora vamos falar de comparações de médias (ou medianas e modas) e desvios padrão (ou interquartis) obtidas de dois ou mais grupos testados ao mesmo tempo.

Imagine que duas pessoas jogam um dardo em um alvo. Só porque um deles conseguiu chegar mais perto do alvo central, não significa que ele é melhor. Eu sei que você é acostumado a pensar assim porque assiste esportes. Contudo, para chegar à final, o esportista já passou por várias provas antes. Além disso, muitas coisas podem acontecer no jogo final que acabem levando à derrota do melhor jogador. Para sabermos quem é o melhor, precisamos fazer mais ou menos como na NBA, que testa seus times em *playoffs*: uma série de 7 jogos em que o time que vencer pelo menos 4 é qualificado para a próxima fase. Na ciência, fazemos algo parecido com um *playoff* de 100 jogos para determinar, com algum grau de confiança, que o achado encontrado tem mesmo chance de ser real *naquela amostra*.

Uma pesquisa, então, inicia com a *hipótese nula*: assumimos que não há diferença real entre os grupos. Já a hipótese alternativa diz que a diferença



entre os grupos é real. As hipóteses nula e alternativa são comparadas através de testes de hipótese – equivalentes a um payoff de 100 jogos.

Para realizarmos um desses testes estatísticos, o primeiro passo é definir se os dados obtidos estão ou não distribuídos normalmente: os dados parecem mais com o gráfico da Figura 2 ou com o da Figura 3? Podemos realizar o teste de Kolmogorov-Smirnov (o nome acaba com v, hein?), para ver se os dados são distribuídos normalmente ou não. Podemos também transformar um dado não normal em normal usando o valor de logaritmo em vez do valor em si (mas fica tranquilo, não é você que vai fazer isso). O próximo passo é verificar a *significância estatística* da diferença das tendências ou dispersões que você obteve. Por que isso é importante e como funciona isso?

E é aí que entra o famoso *valor de p*. Vamos entender o que é isso: são realizados cálculos que determinam se há chance de que, naquele estudo ou teste, os resultados tenham acontecido por acaso e não porque há real diferença entre os grupos. O “p significativo” em Medicina foi convencionado como menor ou igual que 0,05. Se você demonstrar que há diferença do resultado entre dois grupos com p significativo, então você rejeitou a hipótese nula – o que significa que o estudo foi positivo (Figura 9).

**Figura 9.**

	Existe, de verdade, diferença	Não existe diferença, na verdade
Hipótese nula aceita = não se encontrou diferença	Falso-negativo = erro tipo 2	Achado verdadeiro
Hipótese nula rejeitada = se encontrou diferença	Achado verdadeiro	Falso-positivo = erro tipo 1

O conceito de aceitar e rejeitar hipótese nula é chato e contra intuitivo, mas vou fazer o meu melhor pra te fazer entender: hipótese nula é que não há diferença entre os grupos. Se o estudo rejeitou isso, é porque, no estudo, houve diferença entre os grupos. Se essa diferença realmente existir, ótimo, mas, se não existir, você obteve um resultado falso-positivo, que é sinônimo de erro tipo 1. Poderia ser mais simples? Poderia, mas não é.

Atenção: “significativo” não é sinônimo de relevante, mas algo que reduz a probabilidade de um falso-positivo (ou erro tipo 1). Muitas pessoas acreditam, equivocadamente, que um pequeno valor de  $p$  representa um grande efeito clínico. Uma revisão sistemática da Cochrane demonstrou que o uso de vitamina C reduz a duração dos sintomas de resfriado em 8%.<sup>6</sup> Como o resfriado dura algo em torno de 6 dias, essa redução significa que, tomando vitamina C, os sintomas tendem a durar 5 dias e meio. Mesmo não sendo tão relevante clinicamente, o estudo é considerado positivo, devido ao baixo valor de  $p$  reportado.

Antes de passarmos a diante, dê uma lida no Quadro 1, já demonstrado em outro capítulo do livro, mas que relembra os tipos básicos de erros que um pesquisador deve evitar: erros tipo 1 e tipo 2.

**Quadro 1. Para nunca mais esquecer: power e valor de  $p$ , erro tipo 1 e tipo 2**

Concordamos que a literatura científica podia ser mais palatável. Por isso, vamos resumir breve e didaticamente esses termos e deixar, aqui, uma analogia que pode ajudar você a entender de uma vez por todas essas questões.

Erro tipo 1 é quando o pesquisador encontra um resultado falso-positivo. Erro tipo 2 é quando o pesquisador encontra um resultado falso-negativo.

Imagine que 100 pessoas com tromboembolismo pulmonar realizam uma tomografia de tórax com contraste. Em 10% delas, a tomografia foi erroneamente normal, portanto, falso-negativas. Se isso fosse um estudo científico, se chamaria “erro tipo 2”. Se em 10% houve erro, significa que em 90% houve acerto. A capacidade de um teste ser verdadeiro positivo é chamada de “sensibilidade” em Medicina, mas em pesquisa se chama “*power*”. O *power* normalmente buscado em pesquisas médicas é de 0,8 ou 80% (ou seja, se aceita um máximo de 20% de falsos-negativos ou erros do tipo 2).

Agora imagine que 100 pessoas sem tromboembolismo pulmonar realizaram uma tomografia de tórax com contraste. Em 10% delas, a tomografia erroneamente detectou tromboembolismo pulmonar, um resultado falso-positivo. Se isso fosse um estudo científico, se chamaria “erro tipo 1” ou erro  $\alpha$ . Se em 10% houve erro, então em 90% houve acerto. A capacidade de um teste ser verdadeiro-negativo é chamada de “especificidade” em Medicina. Em pesquisas, é chamado de  $1 - \alpha$ .

Agora vamos usar os mesmos conceitos na análise de artigos científicos. Suponha que você deseja descobrir qual entre dois antidiabéticos tem maior eficácia no tratamento do diabetes. Para isso, você realiza um estudo e os seguintes cenários podem se apresentar:

- a. Você conclui que os antidiabéticos são igualmente eficazes e eles, de fato, são. Você aceitou a hipótese nula (a de que não há diferença) e, como ela é verdadeira na vida real, nenhum erro foi cometido.
- b. Você conclui que um antidiabético é melhor que o outro, mas na realidade eles são igualmente eficazes. Você, equivocadamente, cometeu um erro do tipo 1. Em outras palavras, um falso-positivo.
- c. Você conclui que os tratamentos são igualmente eficazes, mas, na vida real, um dos tratamentos é mesmo mais eficaz. Você cometeu um erro do tipo 2, que é o falso-negativo.
- d. Você conclui que um dos antidiabéticos é mais eficaz e, na vida real, isto está correto. Você rejeitou corretamente a hipótese nula (de que não havia diferença). Parabéns, você não cometeu nenhum erro!

Como se calcula o valor de p? De diversas maneiras. Depende da natureza do dado encontrado por você na pesquisa: normal ou não normal? Se você encontrou resultados “normais” (que se parecem com a Figura 2), tem de fazer *testes paramétricos*: teste t de Student, ANOVA, qui-quadrado ( $\chi^2$ ) ou teste exato de Fisher (usado para dados categóricos: SIM ou NÃO, por exemplo). Se você tiver encontrado resultados “não normais”, então deve usar Mann-Whitney. A Tabela 1 resume brevemente os testes estatísticos que devem ser usados em cada tipo de dado observado.

**Tabela 1.**

	Nominal	Ordinal	Quantitativo não normal	Quantitativo normal
Nominal (SIM ou NÃO, sexo do paciente etc.)	Qui-quadrado ou Fisher	Mann-Whitney	Mann-Whitney ou ajuste logarítmico	Teste de t Student
Ordinal (são ainda qualitativas, mas possuem várias categorias ordenadas: meses do ano, escolaridade...)	Qui-quadrado	Spearman	Spearman	Spearman
Quantitativo não normal	Regressão logística	*	<i>Forest plot</i> , Pearson ou Spearman	<i>Forest plot</i> , Pearson, Spearman e Regressão Linear
Quantitativo normal	Regressão logística		Regressão linear	Pearson e regressão linear

Resumo dos tipos de testes que devem ser usados para cada dado imputado (linhas) e para cada desfecho estudado (colunas). Não precisa decorar isso. Só deixei caso você tenha que fazer alguma pesquisa algum dia.

## Interpretando o valor de p

A definição precisa do valor de p pode parecer um pouco confusa por envolver probabilidade condicional. É importante entender que **o valor de p**

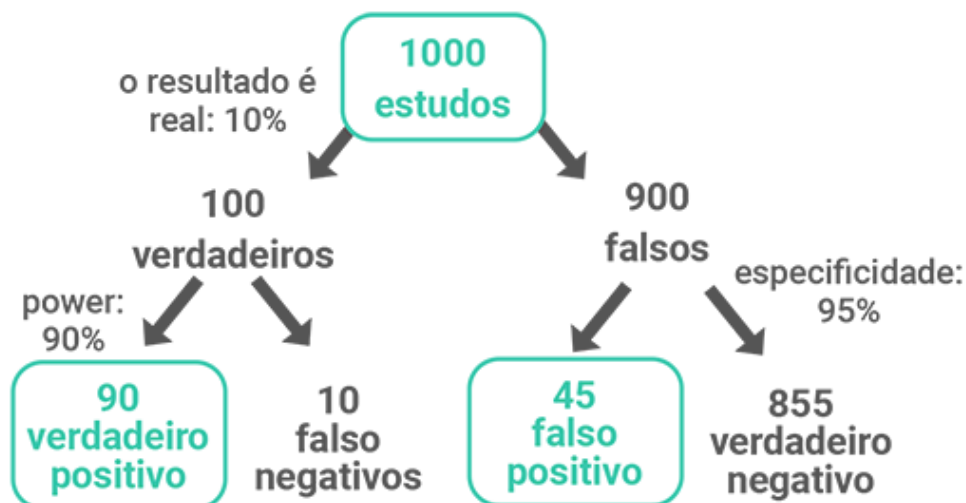
**não é a probabilidade de o estudo ter um resultado falso-positivo.**<sup>7</sup> O valor de  $p$  é a probabilidade de obtermos aquele resultado ou resultados mais extremos *assumindo que a hipótese nula é verdadeira* – ou seja, na vida real (cenário fora da pesquisa) não existe mesmo diferença entre os grupos.<sup>8</sup>

A interpretação dos valores de  $p$  deve ser feita da mesma forma que interpretamos um exame diagnóstico. Se os valores de  $p$  fossem exames de pacientes, considerando as seguintes informações, quantas vezes “diagnosticaríamos” corretamente nossos estudos?

A significância de 0,05 (ou 5%) equivale a um exame com especificidade de 95% – ou seja, 95% das pessoas saudáveis são corretamente identificadas, recebendo resultado negativo. Já a sensibilidade, em estudos ou testes de hipótese, é chamada de *power* (poder): a capacidade do estudo de achar algum efeito, se ele existir. No nosso exemplo, vamos assumir um poder de 90%. Se fosse um exame, significaria que 90% das pessoas doentes receberiam resultado positivo. Traduzindo para estudos: quando houver uma real diferença entre os grupos, 90% dos estudos detectariam essa diferença. Finalmente, precisamos ter uma ideia da probabilidade das hipóteses que estamos estudando (as hipóteses alternativas de cada estudo) serem verdadeiras. Isso é análogo à prevalência da doença na população. Apesar de não sabermos quais das nossas hipóteses são realmente verdadeiras – e é justamente por isso que estudos são elaborados – suponha que, para fins do exemplo, 10% são verdadeiras. Lendo 1000 estudos com essas características, qual a probabilidade de um desses estudos com  $p < 0,05$  ser um achado real?

Uma forma de responder a essa pergunta é construindo a árvore de probabilidades ilustrada na Figura 10 - veja ela agora, leia a legenda e só depois volte aqui. Vai! Quando a hipótese nula (de que não há diferença) é verdadeira, 45 estudos tiveram  $p$  significativo – ou seja, são falsos-positivos. No caso da hipótese alternativa verdadeira (de que há mesmo diferença), 90 estudos tiveram corretamente  $p$  significativo. Ao todo, temos  $45 + 90 = 135$  estudos com  $p$  significativo. A resposta da pergunta que fizemos é o número de estudos verdadeiramente significativos dividido pelo total de estudos significativos:  $90 / 135 = 66,6\%$ . Isso é chamado de *valor preditivo positivo*. Se você mudasse sua conduta a cada um desses estudos positivos, você estaria certo apenas 66,6% das vezes.

**Figura 10.**



Probabilidade de ser verdadeiro positivo =  $\frac{\text{Verdadeiro positivo}}{\text{Todos os positivos}}$

Probabilidade de ser verdadeiro positivo =  $\frac{90}{90 + 45}$

Probabilidade de ser verdadeiro positivo = 66,6%

Árvore de probabilidades que ilustra os resultados de 1.000 estudos, dos quais 10% têm efeito real (hipótese alternativa verdadeira), com nível de significância de 5% e poder 90%. O valor preditivo positivo neste exemplo é 66,6% e o valor preditivo negativo é 99%.

E se você se deparar com um resultado não significativo? Fazendo cálculo do valor *preditivo negativo* temos  $855 / (855 + 10) = 99\%$ . Isso significa que se você não mudar a sua conduta por um desses estudos negativos, você estaria correto 99% das vezes. Considerando estudos com tais características, devemos dar muito mais importância a um resultado negativo do que um resultado positivo.

O que acontece se esses valores fossem diferentes – se o poder fosse 80%, como frequentemente adotado em pesquisas médicas? O valor preditivo positivo, nesse caso, seria  $80 / (45 + 80) = 36\%$ . Imagina pensarmos que todos os estudos significativos da literatura estão corretos apenas 36% das vezes? Na realidade, é ainda pior que isso.

O poder é definido no momento do desenho do estudo e está completamente relacionado com o tamanho da amostra – quanto maior a amostra, mais poder tem o estudo. Outro detalhe importante no cálculo da amostra é que precisamos determinar a diferença entre os grupos – ou o tamanho do efeito – que queremos detectar. Quanto maior a diferença a ser detectada, maior será a amostra. Se quisermos detectar uma diferença real de

1% entre os grupos, a amostra será maior do que se esperarmos uma diferença de 20%. Isso é relevante porque, se realizarmos muitos estudos com baixo poder, o valor preditivo positivo cai justamente porque os resultados significativos têm ainda chances maiores de serem falsos-positivos.

Algumas considerações sobre o valor de  $p$  e a significância estatística:

- a. O cálculo do valor de  $p$  leva em consideração que a amostra foi obtida de forma aleatória. Se a amostra está enviesada de alguma maneira e não representa o mundo real, então ela não pode ser extrapolada para o mundo real.
- b. Estatisticamente significativa não é a mesma coisa que clinicamente relevante. Chamamos de relevância algo que vai, de fato, melhorar a qualidade de vida ou prolongar a vida do paciente, ou ajudar médicos a obterem esse desfecho positivo.
- c. Não deve ser analisado sozinho: aprendemos no Capítulo 5 que o valor de  $p$  deve ser usado como um dos componentes do Teorema de Bayes para chegar à probabilidade pós-teste de algo ser verdadeiro.

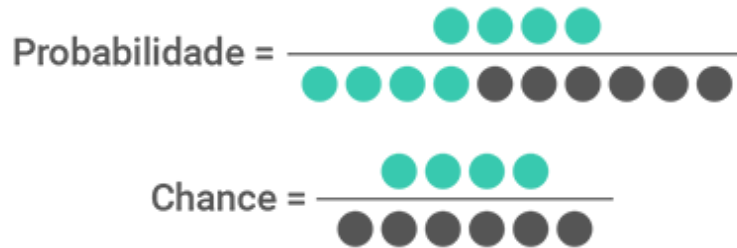
## Chance ou probabilidade? Uma introdução às medidas de associação

*Odds* e *probability* são conceitos diferentes, por mais que, se você pesquisar em um tradutor, essas palavras sejam sinônimas. Neste livro, vamos traduzir *odds* como chance e *probability* como probabilidade. Vamos entender a diferença entre elas:

**Probabilidade** é a fração de vezes que algo acontece em várias tentativas ( $X/\text{total}$ ): se você jogar um dado para cima, a probabilidade que ele caia com a face do número 1 virada para cima é  $1/6$ , portanto: 16% ou 0,16.

**Chance** é o número de vezes que esse evento pode ocorrer dividido pelo número de vezes que ele não ocorre. No mesmo exemplo do dado, a chance que ele caia com a face do número 1 virada para cima é  $1/5 = 0,2$  ou 20% (Figura 11).

**Figura 11.**



Diferença entre probabilidade e chance: probabilidade é o número de vezes em que um evento ocorre dividido pelo total de eventos. Chance é o número de vezes que um evento ocorre dividido pelo número de vezes que não ocorre.

Da chance surge um índice muito importante em bioestatística: a **razão de chances** (*odds ratio*), usada em estudos do tipo caso-controle (revisando o que falamos no Capítulo 3, é aquele tipo de estudo observacional que analisa o passado de indivíduos que já possuem uma doença a fim de encontrar fatores de risco que possam ter causado essa doença).<sup>9</sup> A Figura 12 mostra como se calcula uma razão de chances no exemplo de um estudo do tipo caso-controle. Se a razão de chances for  $> 1$ , então o fator existente no passado do paciente está mesmo aumentando a chance de ele ter a doença. Se  $< 1$ , então aquele fator existente no passado da pessoa teve algum feito em evitar a doença. Se o IC 95% não incluir o 1, o resultado é estatisticamente significativo (se incluir o 1, significa que, grosso modo, a probabilidade de o fator estudado não ter efeito nenhum é maior que 5%).

**Figura 12.**

	Doença	Controle
Exposto	A	B
Não exposto	C	D

Chance de um exposto ter a doença: A/B
Chance de um não exposto ter a doença: C/D
Razão de chances: (A/B)/(C/D)

Cálculo da razão de chances em um estudo do tipo caso-controle.

Da probabilidade surge mais um índice importante em bioestatística que é o **risco relativo** (RR) (*relative risk*). Risco é a probabilidade de que algo aconteça com alguém. O risco é “relativo” quando está sendo comparado com outra pessoa ou grupo de pessoas. Esse dado é usado em estudos do tipo

coorte (ou seja, estudos que analisam o futuro de pessoas expostas a um fator de risco ou a uma terapia).<sup>10</sup> Veja como se calcula o risco relativo na Figura 13. Quando se calcula o RR, se ele for  $> 1$  com IC 95% que não cruza o 1, significa que aquele fator ou terapia aumenta o risco de ter aquele desfecho. Caso seja  $< 1$  com IC 95% que não cruza o 1, significa que aquele fator ou terapia protege contra aquele desfecho. A **redução do risco relativo (RRR)** é  $1 - RR$ : ou seja,  $100\% - x\% =$  em quantos por cento foi reduzido o risco de ter aquele desfecho usando uma nova terapia. Já a **redução absoluta do risco (RAR)** é a subtração matemática do risco dos expostos ao fator ou à terapia pelo risco dos não expostos.

**Figura 13.**

	Exposto	Não exposto
Doente	A	B
Não doente	C	D

Probabilidade de um doente ter sido exposto:  $A/A + B$

Probabilidade de um não doente ter sido exposto:  $C/C + D$

Risco relativo:  $(A/A + B)/(C/C + D)$

Redução do risco relativo (RRR):  $1 - \text{risco relativo}$

Redução do risco absoluto (RAR):  $(A/A + B) - (C/C + D)$

Cálculo do risco relativo em um estudo do tipo coorte. Note que, na a figura 9, usamos o conceito de *chance* e, nesta figura, usamos o conceito de *risco* ou *probabilidade*.

Vamos aprender a fazer esses cálculos na vida real. Na Figura 14, observamos um estudo que objetivava estimar fatores de risco para melanoma cutâneo maligno na Argentina, um deles a quantidade de nevus nos braços.<sup>11</sup> Na Figura 15, observamos um hipotético estudo de coorte em que se avaliou se tabagistas teriam maior incidência de câncer de pulmão.

**Figura 14.**



	Melanoma	Controle
<b>≥ 10 nevos</b>	26	34
<b>&lt; 10 nevos</b>	67	205

Chance de um doente ter sido exposto:  $A/B = 26/34 = 0.76$

Chance de um não doente ter sido exposto:  $C/D = 64/205 = 0.31$

Razão de chances:  $(A/B)/(C/D) = 0.76/0.31 = 2.45$

Esse estudo era do tipo caso-controle e comparou pessoas com melanoma e sem melanoma para diversas características, entre elas, a quantidade de nevos no braço. O cálculo de razão de chances está demonstrado na figura. O fato de ser  $> 1$  (e ter IC95% que não cruza o 1) demonstrou que, nessa amostra, a quantidade de nevos é um fator de risco para melanoma.

**Figura 15.**

	Tabagista	Não tabagista
<b>Câncer de pulmão</b>	90	26
<b>Controle</b>	64	150

Probabilidade de um doente ter sido exposto:  $90/116$

Probabilidade de um não doente ter sido exposto:  $64/214$

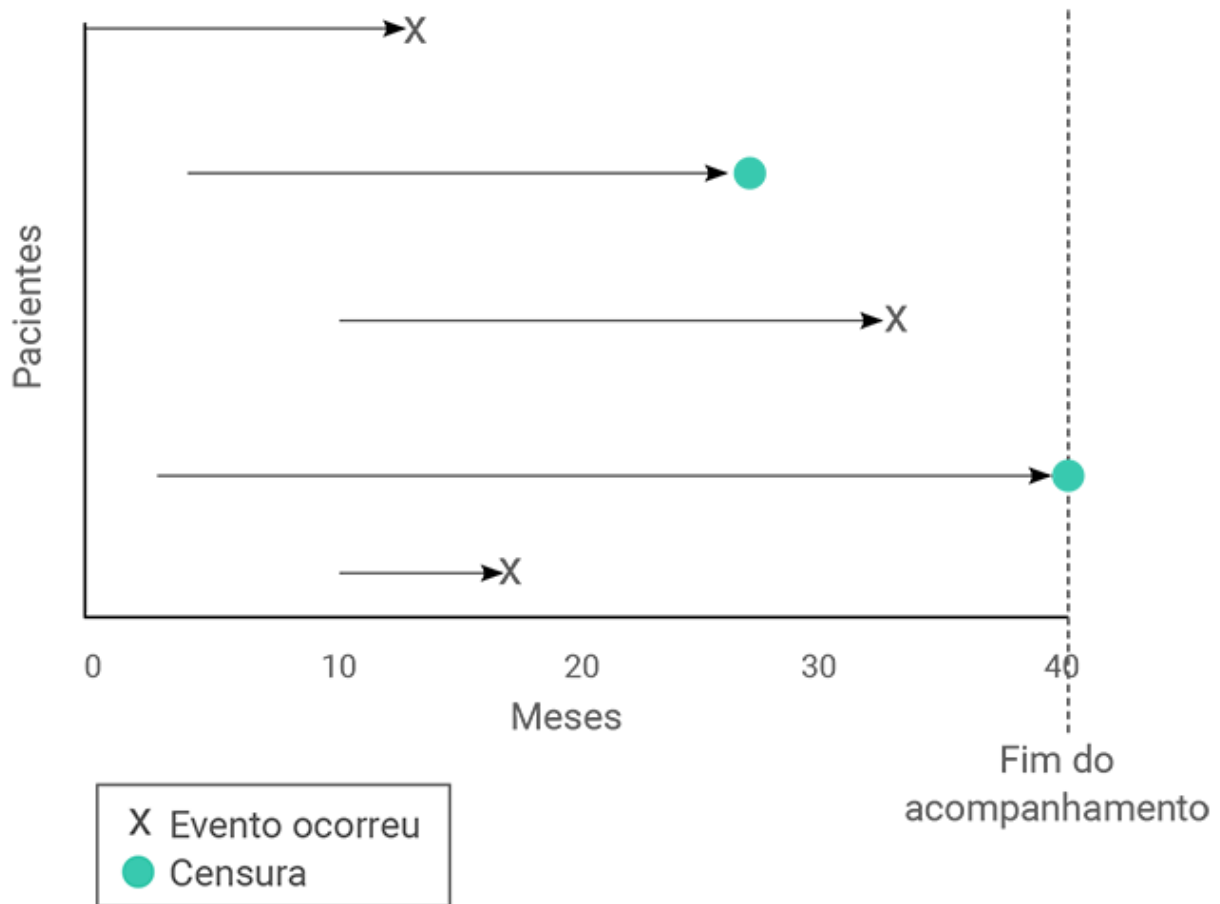
Risco relativo:  $(A/A+B)/(C/C+D) = 0,77/0,29 = 2,65$

Esse estudo hipotético do tipo coorte estudou se tabagistas e não tabagistas teriam incidências diferentes de câncer de pulmão. Por ser uma coorte, o cálculo realizado foi o de risco relativo, que foi de 2,65 com IC95% que não cruzava a linha do 1, portanto, estatisticamente significativo.

Em diferentes estudos, é comum que os pesquisadores meçam não apenas o risco do evento (ou desfecho) como também o *tempo até o evento acontecer*. Para lidar com dados desse tipo, é utilizada uma parte da estatística chamada “análise de sobrevida”<sup>12</sup> (só que o evento estudado não necessariamente deve ser “mortalidade”). O interessante da análise de sobrevida é que os pacientes que desistiram ou foram perdidos durante o acompanhamento (*follow-up*) não precisam ser eliminados da análise estatística. Esses pacientes são “censurados”, já que apenas sabemos que o evento não ocorreu até quando foram acompanhados. Na realidade, todos os pacientes que não tiveram o evento são censurados e os respectivos tempos

de acompanhamento, mesmo quando incompletos, contribuem na análise (Figura 16).

**Figura 16.**



Pacientes em estudo hipotético com diferentes tempos de acompanhamento. Alguns tiveram o evento e outros foram censurados. Utilizando análise de sobrevivência, os tempos de acompanhamento de todos os pacientes entram na análise.

Em estudos com a análise de sobrevivência, costuma-se usar uma medida chamada **hazard ratio**, que é o risco relativo do evento em função do tempo. Pode ser calculado através da razão das taxas de incidência (*incidence rate ratio*). A taxa de incidência é o número de eventos ocorridos em um estudo (ou grupo do estudo) dividido pela quantidade de pessoas-tempo, sendo pessoas-ano a medida de tempo mais utilizada. Por exemplo, em um estudo com 5 pessoas acompanhadas por 10 anos, o denominador será 50 pessoas-ano.<sup>13, 14</sup> A razão das taxas de incidências é calculada dividindo a taxa de incidência em um grupo (terapia) pela taxa de incidência em outro grupo

(controle). No entanto, normalmente o *hazard ratio* é calculado por meio do modelo de riscos proporcionais de Cox, que permite ajuste por outras variáveis.<sup>15</sup> Por exemplo, no estudo EMPA-REG<sup>16</sup> foi calculado o *hazard ratio* de um desfecho composto (quando se estuda a soma de diferentes eventos) do grupo que usou empagliflozina em comparação com o grupo placebo. Nesse estudo, o *hazard ratio* foi ajustado por idade, gênero, região geográfica, entre outras variáveis (Figura 17).

**Figura 17.**

Table 1. Primary and Secondary Cardiovascular Outcomes.					
Outcome	Placebo (N = 2333)		Empagliflozin (N = 4687)		Hazard Ratio (95% CI)
	no. (%)	rate/1000 patient-yr	no. (%)	rate/1000 patient-yr	
Death from cardiovascular causes, nonfatal myocardial infarction, or nonfatal stroke: primary outcome*	282 (12.1)	43.9	490 (10.5)	37.4	0.86 (0.74–0.99)

Hazard ratio (pelo método de Cox) = 0.86

Razão das taxas de incidência =  $\frac{\text{taxa de incidência no grupo}}{\text{taxa de incidência no grupo controle}}$

Razão das taxas de incidência =  $\frac{37.4}{43.9} = 0.85$

Risco relativo =  $\frac{10.5}{12.1} = 0.86$

Redução do risco relativo =  $1 - 0.86 = 0.14$  (o medicamento reduz em 14% os desfechos listados)

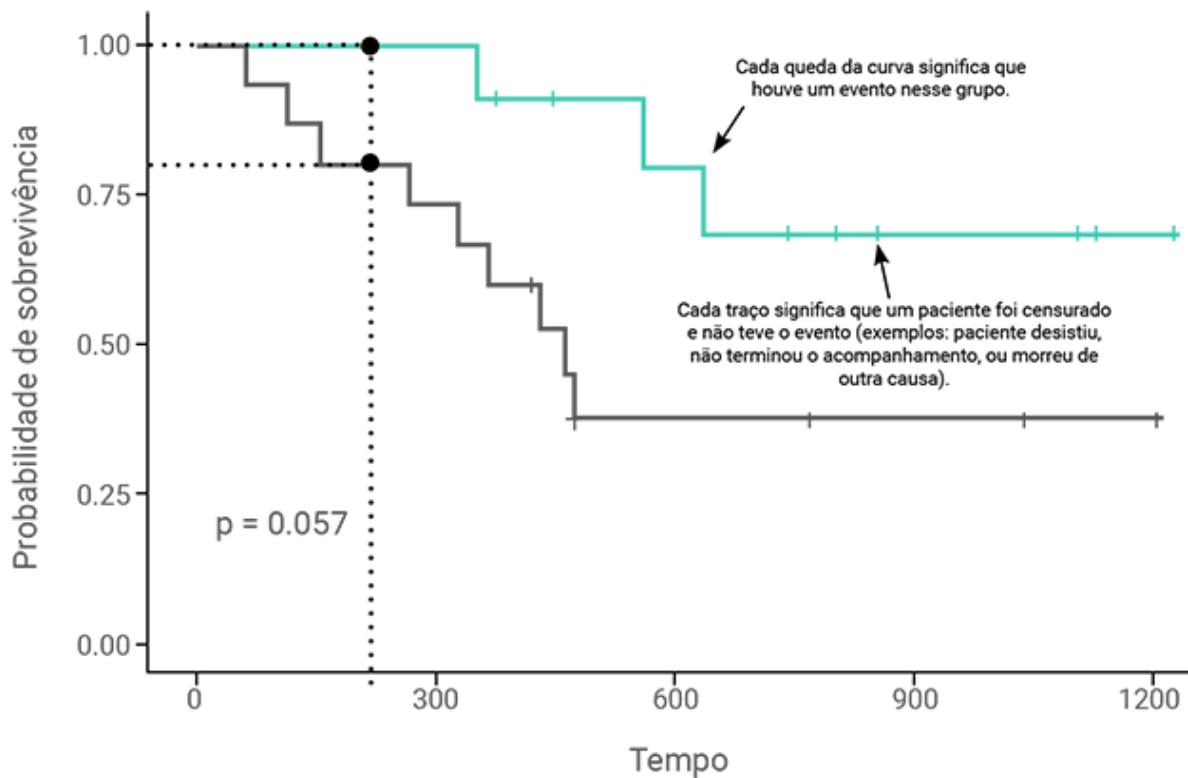
Redução absoluta do risco =  $12.1 - 10.5 = 1.6$

O cálculo do hazard ratio é muito comum em ensaios clínicos randomizados. O hazard ratio foi calculado pelo modelo de Cox e foi ajustado pela idade, gênero, região geográfica, entre outras variáveis. A razão de incidências é uma aproximação. O risco relativo se calcula igual ao que foi exemplificado no estudo de coorte: dividindo a probabilidade do desfecho dos expostos pela dos não expostos.

Também é muito utilizada a curva de sobrevida ou **curva de Kaplan-Meier**,<sup>15</sup> que relaciona a probabilidade da ocorrência do evento em relação ao tempo de acompanhamento. Nessa curva, o eixo horizontal (x) demonstra o tempo de acompanhamento dos pacientes e o eixo vertical (y) traz a probabilidade estimada do evento. Quando o evento é a morte do paciente, o

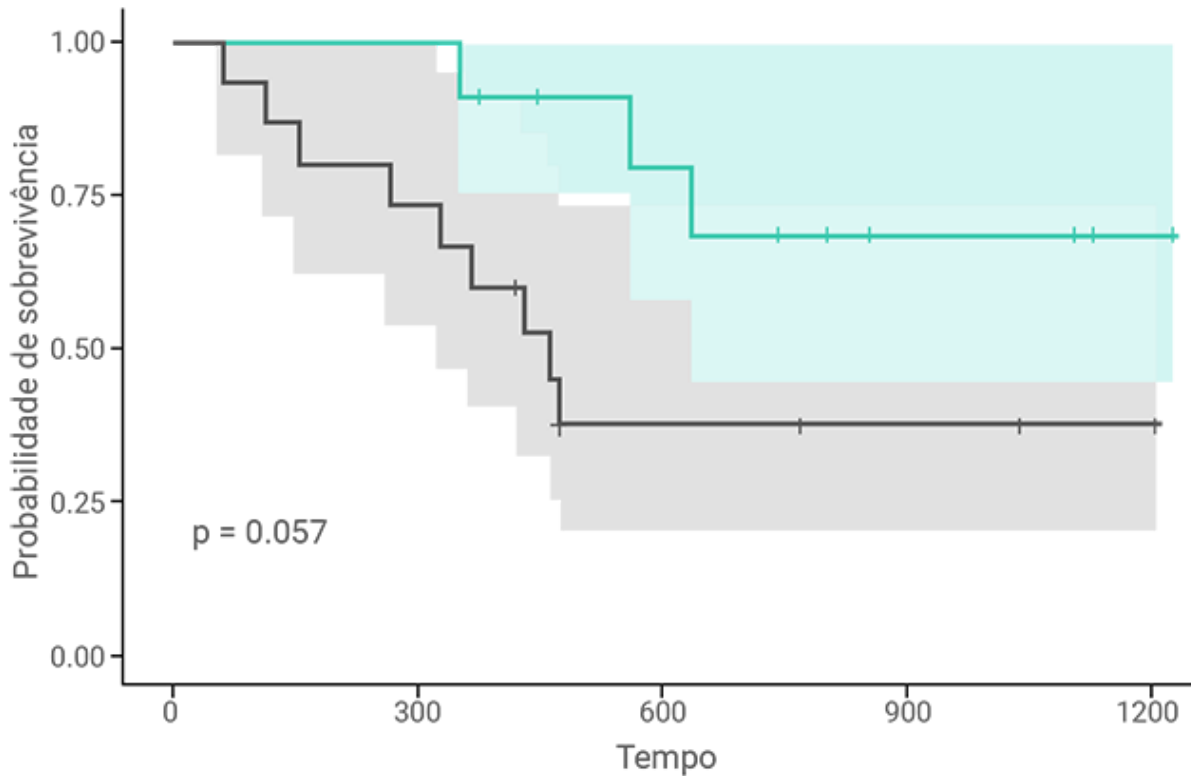
gráfico pode exibir a probabilidade da sobrevivência ao longo do tempo. Observe as Figuras 18 e 19 para compreender melhor.<sup>17, 18</sup>

**Figura 18.**



A curva de Kaplan-Meier consegue nos informar a sobrevida de acordo com o tempo de acompanhamento dos pacientes. Conseguimos perceber, também, que o benefício desse medicamento é precoce: as curvas de cada grupo se afastam logo nos primeiros dias, tanto que no dia 230, demonstrado na figura, a redução de risco relativo já é de 20% (100% – 20%).

**Figura 19.**



A curva de Kaplan-Meier exibindo o intervalo de confiança para as probabilidades de sobrevivência de cada grupo. Note como os intervalos se sobrepõem.

## Entendendo o NNT e o papel limitado dos médicos na história natural das doenças

O **NNT – número necessário para tratar** – é um dado importante na interpretação de pesquisas clínicas, pois significa o número de pessoas que precisam ser tratadas para evitar 1 evento ou desfecho. É calculado dividindo 100 pela redução absoluta do risco. No estudo EMPA-REG (Figura 14), o NNT da empaglifozina foi de  $100/1,6 = 62,5$ . Isso significa que é necessário prescrever a empaglifozina para 62,5 pacientes para que um 1 se beneficie (evitará o desfecho composto analisado no estudo em 1 paciente).<sup>16</sup>

Imagine que você quer comparar a eficácia de um medicamento, para o tratamento de pacientes com fibrilação atrial, com a varfarina, droga habitualmente utilizada. O objetivo do tratamento é evitar eventos isquêmicos e acidente vascular encefálico (AVE). Então, você entrega o medicamento aos pacientes e, após um acompanhamento de 590 dias, observa uma redução do risco relativo de 21% do desfecho composto “AVE

e eventos isquêmicos”. Excelente! Mas vamos observar a tabela do estudo (Figura 20).

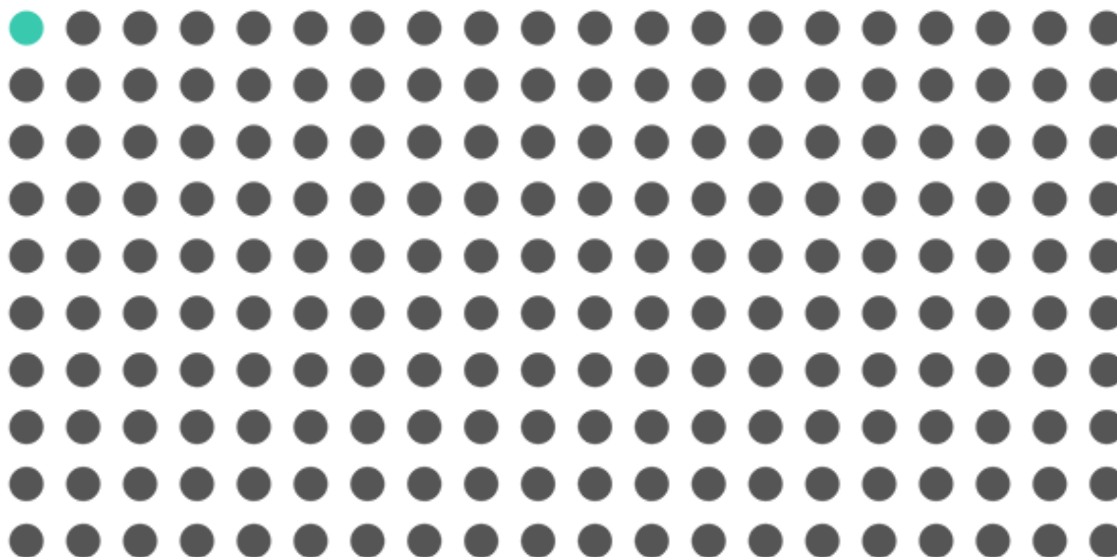
**Figura 20.**

	Novo medicamento	Varfarina
<b>AVE ou evento isquêmico</b>	1,7	2,2
Hazard ratio = $\frac{\text{risco no grupo novo medicamento}}{\text{risco no grupo Varfarina}}$		
Hazard ratio = $\frac{1,7}{2,2} = 0,79$		
Redução do risco relativo = $1 - 0,79 = 0,21$ (o medicamento reduz em 21% o desfecho composto)		
Redução absoluta do risco = $2,2 - 1,7 = 0,5$		
NNT = $100/\text{RAR} = 100/0,5 = 200$		

Interpretação dos dados de um estudo que testou uma nova terapia versus varfarina na redução de AVE ou eventos isquêmicos num acompanhamento de 590 dias. A taxa de incidência foi de 1,7 pacientes-ano no grupo “novo medicamento” e 2,2 pacientes-ano no grupo “varfarina”. A redução do risco relativo foi de 21%, mas o NNT é de 200. Isso que significa que, a cada 200 pacientes que substituir a varfarina pelo novo medicamento, 1 terá um evento reduzido. Esse novo medicamento é a rivaroxabana e os dados são do estudo ROCKET-AF.<sup>26</sup>

Você percebeu que o NNT é uma medida intuitiva sobre como um tratamento pode afetar seus pacientes? Ele traz informações relevantes sobre a quantidade de pacientes que você precisa tratar por um determinado acompanhamento, numa população como aquela, para evitar um desfecho por causa da nova conduta. O que acontece com os outros 199 pacientes? Observe na Figura 21.

**Figura 21.**



1 em cada 200 serão salvos pela nova terapia. Os outros 199:

Terão o evento mesmo assim

Terão o evento evitado por outra terapia

Não terão o evento porque não teriam mesmo (história natural da doença não é que 100% das pessoas tenham eventos isquêmicos)

O que significa um novo medicamento com NNT de 200? Que 1 em cada 200 pacientes terá um evento evitado pelo novo medicamento. Os outros 199 a) terão o evento mesmo assim, b) terão o evento evitado por outra terapia ou c) não terão o evento porque não teriam mesmo.

É muito importante que o médico entenda o real impacto que a história natural das doenças tem no acompanhamento dos seus pacientes. A Covid-19, por exemplo, não mata 100% dos seus doentes. Longe disso, a letalidade global estimada foi de 0,5 a 1% - quando se estimam também aqueles pacientes que nunca chegaram a realizar testes para confirmar a doença.<sup>19</sup> Se um pesquisador quer saber se um novo tratamento funciona, ele não pode simplesmente usar argumentos do tipo “eu usei em meus pacientes e os curei”, porque a chance *natural* de se curar dessa doença é de 99% a 99,5%.

Um medicamento clinicamente relevante é aquele que seu uso prolonga vida ou melhora a qualidade de vida, direta ou indiretamente. Um exemplo é o estudo RECOVERY (Figura 22), que avaliou o uso de dexametasona em pacientes com Covid-19.<sup>20</sup> O resultado é clinicamente relevante pois reportou considerável redução do risco de mortalidade (um desfecho clínico), a ponto de o NNT ser numericamente pequeno, o que é bom. Note que o RECOVERY foi feito com pacientes hospitalizados, por isso os

pacientes do estudo eram mais graves que a população geral (porque os pacientes que se hospitalizam o fazem exatamente por maior gravidade). Isso contribuiu para a alta letalidade reportada, mesmo nos pacientes que usaram dexametasona. Importante é perceber que, nessa população, a letalidade foi reduzida significativa e relevantemente.

**Figura 22.**

	Dexametasona	Placebo
<b>Mortalidade em 28 dias</b>	22,9	25,7
Risco relativo = $\frac{\text{proporção no grupo terapia}}{\text{proporção no grupo controle}}$		
Risco relativo = $\frac{22,9}{25,7} = 0,83$		
Redução do risco relativo = $1 - 0,83 = 0,17$ (o medicamento reduz em 17% os desfechos listados)		
Redução absoluta do risco = $25,7 - 22,9 = 2,8$		
NNT = $100/\text{RAR} = \frac{100}{2,8} = 35,7$		

Dados do estudo RECOVERY que testou dexametasona para pacientes com Covid-19 versus placebo. Observe que a dexametasona teve alta relevância clínica, chegando a um NNT arredondado de 36. Em pacientes em ventilação mecânica, o NNT foi de 8, o que significa que 1 em cada 8 pacientes desse subgrupo sobreviveram por conta da dexametasona (sem benefício para os outros 7 pacientes).

E você sabe quais são os NNTs das terapias da sua especialidade? Uma boa pesquisa no site [thennt.com](http://thennt.com) pode lhe ajudar bastante. A Tabela 2 resume alguns desses NNTs.

**Tabela 2.**

Procedimento	Desfecho estudado	Número necessário para tratar
Corticoides para faringite	Melhora da dor em 48 horas	3 <sup>21</sup>
Aspirina em infarto agudo	Mortalidade	42 <sup>2</sup>
Aspirina para prevenir infartos em quem nunca teve infarto	Mortalidade	Infinito <sup>2</sup>



Procedimento	Desfecho estudado	Número necessário para tratar
Estatinas em pacientes cardiopatas ou que já tiveram AVC	Mortalidade	83 <sup>24</sup>
Epinefrina em parada cardiorrespiratória extra-hospitalar	Mortalidade em 30 dias	112 <sup>25</sup>

Alguns NNTs calculados no site thennt.com após revisão sistemática de artigos de alta qualidade. Significa o número de pessoas que precisam receber o tratamento para se evitar 1 desfecho, portanto, quanto menor o número, melhor.

## Referências bibliográficas

1. Levitin DJ, Alves L. O guia contra mentiras: Como pensar criticamente na era da pós-verdade. Rio de Janeiro: Objetiva; 2019.
2. Whitley E, Ball J. Statistics review 1: presenting and summarising data. Crit Care. 2001/11/29. 2002 Feb;6(1):66-71.
3. Altman DG, Bland JM. Standard deviations and standard errors. BMJ. 2005 Oct;331(7521):903.
4. Morey RD, Hoekstra R, Rouder JN, Lee MD, Wagenmakers E-J. The fallacy of placing confidence in confidence intervals. Psychon Bull Rev. 2016 Feb;23(1):103-23.
5. Altman DG, Bland JM. How to obtain the P value from a confidence interval. BMJ. 2011;343:d2304.
6. Hemilä H, Chalker E. Vitamin C for preventing and treating the common cold. Cochrane Database Syst Rev. 2013;(1).
7. Goodman S. A Dirty Dozen: Twelve P-Value Misconceptions. Semin Hematol. 2008;45(3):135-40.
8. Colquhoun D. An investigation of the false discovery rate and the misinterpretation of p-values. R Soc Open Sci. 2014;1(3).
9. Bland JM, Altman DG. Statistics notes. The odds ratio. BMJ. 2000 May;320(7247):1468.
10. Andrade C. Understanding relative risk, odds ratio, and related terms: as simple as it can get. J Clin Psychiatry. 2015 Jul;76(7):e857-61.
11. Loria D, Matos E. Risk factors for cutaneous melanoma: a case-control study in Argentina. Int J Dermatol. 2001 Feb;40(2):108-14.
12. Altman DG, Bland JM. Time to event (survival) data. BMJ. 1998 Aug;317(7156):468-9.

13. Correia LC. Revisão Metodológica: Incidence Rate, Hazard Ratio. *Medicina Baseada em Evidências*. 2010.
14. Sedgwick P. Hazards and hazard ratios. *BMJ Br Med J*. 2012 Sep;345:e5980.
15. Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part I: basic concepts and first analyses. *Br J Cancer*. 2003 Jul;89(2):232-8.
16. Zinman B, Wanner C, Lachin JM, Fitchett D, Bluhmki E, Hantel S, Mattheus M, Devins T, Johansen OE, Woerle HJ, Broedl UC, Inzucchi SE. Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. *N Engl J Med*. 2015 Nov;373(22):2117-28.
17. Bland JM, Altman DG. Survival probabilities (the Kaplan-Meier method). *BMJ*. 1998 Dec;317(7172):1572.
18. Sedgwick P. How to read a Kaplan-Meier survival plot. *BMJ Br Med J*. 2014 Sep;349:g5608.
19. Meyerowitz-Katz G, Merone L. A systematic review and meta-analysis of published research data on COVID-19 infection fatality rates. *Int J Infect Dis IJID Off Publ Int Soc Infect Dis*. 2020 Dec;101:138-48.
20. Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, Linsell L, Staplin N, Brightling C, Ustianowski A, Elmahi E, Prudon B, Green C, Felton T, Chadwick D, Rege K, Fegan C, Chappell LC, Faust SN, Jaki T, Jeffery K, Montgomery A, Rowan K, Juszczak E, Baillie JK, Haynes R, Landray MJ. Dexamethasone in Hospitalized Patients with Covid-19 – Preliminary Report. *N Engl J Med*. 2020 Jul.
21. Hayward G, Thompson MJ, Perera R, Glasziou PP, Del Mar CB, Heneghan CJ. Corticosteroids as standalone or add-on treatment for sore throat. *Cochrane database Syst Rev*. 2012 Oct;10:CD008268.
22. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. *Lancet (London, England)*. 1988 Aug;2(8607):349-60.
23. Zheng SL, Roddick AJ. Association of Aspirin Use for Primary Prevention With Cardiovascular Events and Bleeding Events: A Systematic Review and Meta-analysis. *JAMA*. 2019 Jan;321(3):277-87.
24. Baigent C, Keech A, Kearney PM, Blackwell L, Buck G, Pollicino C, Kirby A, Sourjina T, Peto R, Collins R, Simes R. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. *Lancet (London, England)*. 2005 Oct;366(9493):1267-78.
25. Perkins GD, Ji C, Deakin CD, Quinn T, Nolan JP, Scomparin C, Regan S, Long J, Slowther A, Pocock H, Black JJM, Moore F, Fothergill RT, Rees N, O'Shea L, Docherty M, Gunson I, Han K, Charlton K, Finn J, Petrou S, Stallard N, Gates S, Lall R. A Randomized Trial of Epinephrine in Out-of-Hospital Cardiac Arrest. *N Engl J Med*. 2018 Jul;379(8):711-21.

26. Patel MR, Mahaffey KW, Garg J, Pan G, Singer DE, Hacke W, Breithardt G, Halperin JL, Hankey GJ, Piccini JP, Becker RC, Nessel CC, Paolini JF, Berkowitz SD, Fox KAA, Califf RM. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *N Engl J Med*. 2011 Sep;365(10):883-91.

## 7. COMO INTERPRETAR ARTIGOS SOBRE TERAPIAS?

### Introdução

Se você vem acompanhando o livro desde o início, já deve estar bem ciente de que artigos observacionais só geram hipóteses quando falamos de novas terapias. A primeira lição deste capítulo vem na sua própria introdução: quando você vir um estudo do tipo “caso-controle” ou do tipo “coorte”, saiba que esse artigo não tem *power* suficiente para confirmar uma hipótese. Óbvio: muitas vezes, artigos de natureza observacional são tudo o que temos (quando é antiético expor um grupo à ausência de um tratamento ou quando os desfechos estudados são raros demais a ponto de requerer que fosse feito o maior *trial* da história).

*Trial*, aliás, é outro nome para “ensaio clínico randomizado” (ECR). Neste capítulo, focaremos a análise dos estudos do tipo *trials*. E por quê? Porque eles sim é que podem confirmar hipóteses. O leitor deste livro perceberá, contudo, que não basta confirmar a hipótese (obter p significativo), tem que ser livre de vieses, tem que ter um bom *power*, esse resultado tem que ser clinicamente relevante, de boa magnitude e extrapolável para o mundo real. Poucas condutas na Medicina possuem todos esses pré-requisitos.

Também é importante que você, antes de ler este capítulo, perceba que os Capítulos 4, 5 e 6 serviram como uma introdução a este, quando vimos: pensamento bayesiano, quando listamos os possíveis vieses em pesquisa e quando introduzimos estatística básica. Fique à vontade para retornar aqueles capítulos caso pense que precisa relembrar algum conceito.

A fim de nunca mais fazer análises superficiais como: “mas o estudo obteve p significativo, então temos que prescrever tal medicamento”; vamos ao capítulo mais importante do livro?

### Por que prescrevemos medicamentos?

Nós, médicos, somos limitados

Essa é uma pergunta que eu tento fazer a mim mesmo várias vezes ao dia, por exemplo, quando, ao ser abordado por um paciente, sinto que ele está contando seus sintomas na expectativa de que eu lhe prescreva algo que lhe cure. Só que a “cura” é um desfecho muito raro em Medicina, especialmente quando se trata de medicamentos.

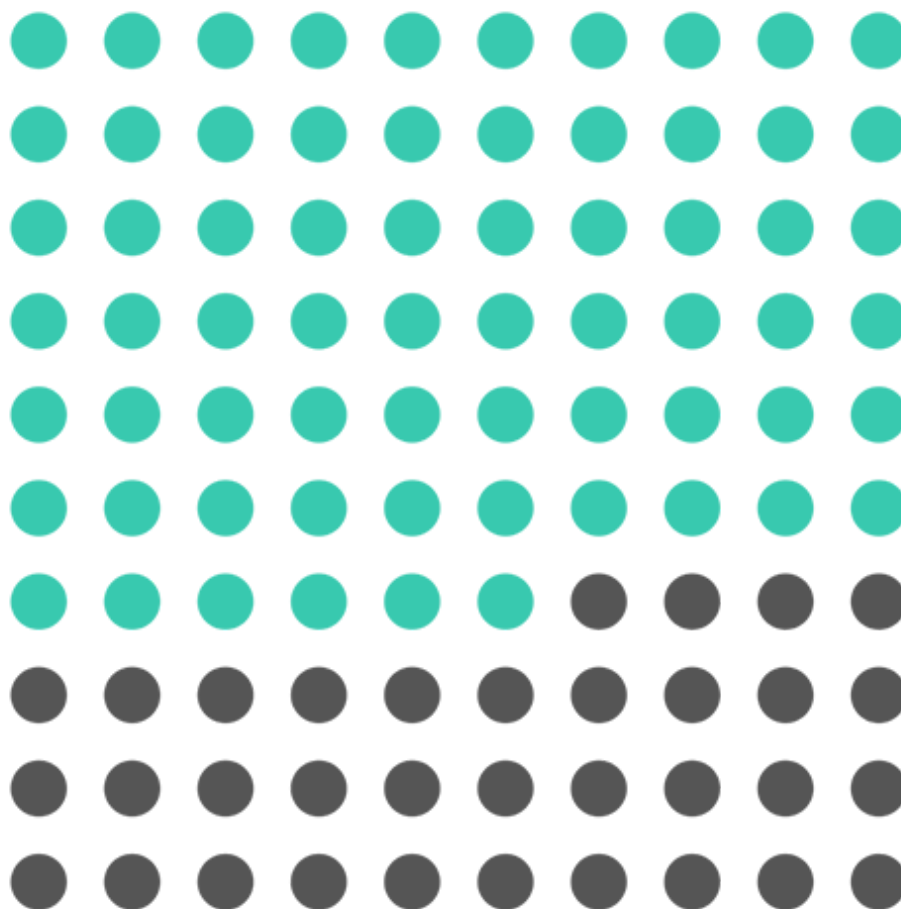
Você viu no Capítulo 6 o conceito de NNT (*number needed to treat*, ou número necessário para tratar) e, se entendeu bem o conceito, perdeu um pouco de fé nas receitas que escrevemos diariamente. Imagine que você está em seu consultório pronto para atender o último paciente do dia e, ao entrar na sala, cheio de coriza e fâcias de dor, o paciente lhe fala que tem estado com entupimento nasal há 6 dias e que nada que ele tomou ajudou seu problema. No afã de ajudar o paciente, você lhe prescreve um ciclo de antibióticos e pede retorno em 1 semana. Ao retornar, o paciente estava curado e você se sentiu o médico mais importante do mundo. É possível que a partir dali, por ter verificado “na pele” a ação do antibiótico, você se convença que esses “tais médicos que seguem a Medicina Baseada em Evidências estão errados”, e que “eu tenho mesmo é que prescrever antibióticos para todos agora”.

Pensamentos assim são muito comuns porque nossa mente está, o tempo inteiro, tentando criar associações. O caos não nos apetece. O problema é que há vários problemas nas conclusões que você tirou:

- a. A letalidade da sinusite aguda é praticamente nula e todos os pacientes vão se curar um dia, mesmo sem medicamentos.
- b. Seu poder de observação de um relato de caso ou mesmo de uma série de casos não é capaz de determinar a eficácia de um medicamento.

Imagine 100 pessoas com sinusite aguda. Dessas 100 pessoas, aproximadamente 50% terão obtido cura na primeira semana e um total de 66% estariam curadas na segunda semana, mesmo sem antibióticos.<sup>1</sup> A Figura 1 demonstra 66 pacientes que foram curados sem antibióticos. É o que chamamos de “história natural da doença”. É assim, dessa maneira relativamente benigna, que ela se apresenta. E, se é assim, como saber se seu paciente não seria um desses 66 que se curariam sozinhos?

**Figura 1.** 66 curados ao fim de duas semanas



A história natural da sinusite aguda é que, sem medicamento, 66 pacientes sejam curados ao final de duas semanas.

Nunca tinha parado pra pensar nisso? Bem-vindo, então, ao nosso **papel limitado como agente transformador do processo de saúde-doença** do indivíduo. Na verdade, nós perdemos feio para a magnífica máquina do corpo humano.

Na realidade, afirmar que o paciente se curou por conta da medicação que você prescreveu é um erro que eu gostaria que nenhum leitor deste livro jamais voltasse a cometer. No caso em específico, é como se você tivesse uma urna com 100 bolinhas, das quais 66 são da cor azul e as demais 34 da cor vermelha. Então você toma um comprimido, pega uma bolinha, ela é azul e você comemora porque houve interferência do comprimido. Percebeu o absurdo? A probabilidade de pegar uma bolinha azul já era de 66%.

Como comprovar, então, que o antibiótico realmente curou o paciente? Com pesquisas que consigam definir que a urna agora possui 70 ou 80 bola

azuis! O melhor tipo de pesquisa, quando se quer comprovar a eficácia de uma terapia, é o *trial*, como já falei.

## Os números são menos limitados

E aí você faz um *trial*, inclui nesse estudo algumas centenas de pacientes e, ao final, consegue demonstrar que a probabilidade de estar com sinusite ao final de duas semanas era de 34% no grupo que usou placebo e de 28% no grupo que usou antibióticos. Agora sim, você conseguiu demonstrar que houve uma redução do risco relativo de possuir sinusite ao final de duas semanas da ordem de 18% ( $RRR = 1 - 0,28/0,34$ ) com *power* adequado e *p* significativo – e a redução absoluta do risco (RAR) foi da ordem de 6% (34% - 28%). Usando o teorema de Bayes, você percebe que há uma boa probabilidade pré-teste de que esse resultado seja verdadeiro.

Agora somos capazes de calcular o NNT da terapia em questão. Para curar 1 paciente de sinusite em até 2 semanas, eu preciso dar antibióticos para 17 pacientes. ( $100/RAR = 100/6 = 17$ ).

Vou até repetir: sabe aquele remédio que você prescreveu e depois ficou se achando o salvador da pátria? Tenta mais 16 vezes, porque em uma delas, você salvou mesmo a pátria. NNT significa a quantidade de pessoas que eu preciso tratar para obter um resultado positivo. E se você acha que a interpretação acabou por aí, está enganado. Vou até te dar alguns segundos pra pensar na próxima pergunta que tem a me fazer.

Você precisa me perguntar sobre os eventos adversos da terapia. A magnitude dos adversos são medidos pelo NNH (*number needed to harm* ou número necessário para causar dano). Se o NNH for menor que o NNT, então temos um problema, porque isso significa que eu preciso tratar menos pessoas pra causar um problema do que para trazer um benefício. O leitor deve ficar atento ao fato de que: se a relevância clínica do problema não for lá essas coisas comparada ao benefício, então fique com o benefício. Por exemplo, o benefício de um medicamento hipotético é evitar uma morte em 1 a cada 20 pacientes e seu malefício é aumentar a quantidade de espirros diários em 1 a cada 6 pacientes. Apesar de ter um NNH menor que o NNT ( $6 < 20$ ), o benefício é muito, mas muito mais relevante clinicamente que o dano. Converse com seu paciente e explique: “vai espirrar mesmo, paciência”.

Agora sim, com todos esses dados em mente e com toda a humildade clínica que eles lhe trazem, entendemos melhor a célebre frase de William

Osler, pai da Medicina Interna, cujos pensamentos parecem nunca envelhecer: “uma das primeiras obrigações do médico é ensinar as massas a não usarem remédios”.

Só entre nós, o que se vê por aí é uma época de enorme desinformação, tanto por parte de médicos como dos leigos, do real papel dos medicamentos, e até mesmo da existência dessa entidade linda chamada “história natural das doenças”.

## Passo 1 – qual o tipo de pesquisa?

### O peso dos fatores confundidores

O tipo de estudo mais adequado para testar a eficácia de uma droga é o *trial*. Os estudos observacionais não têm poder estatístico suficiente para dar essa resposta e os próximos parágrafos vão explicar por quê. Tampouco os relatos de casos (“eu dei o remédio pra 20 pacientes e eles curaram”) servem pra esse intuito, como já foi explicado na seção anterior.

A falha dos estudos observacionais para o teste de novas terapias é que eles não têm metodologia suficiente para afastar os **fatores confundidores**, que são fatores que podem estar presentes na amostra que podem modificar o resultado de uma pesquisa.

Vamos a um exemplo da presença desses tais fatores confundidores: imagine que um pesquisador está realizando testes em uma possível nova terapia para o tratamento de câncer de pulmão (e, por acaso, essa terapia já servia para outro fim, digamos, para o tratamento da hipertensão). O pesquisador, então, observa pessoas doentes e saudáveis e calcula quantas pessoas tomaram e quantas pessoas nunca tomaram esse remédio no passado. O que eu descrevi foi um estudo do tipo “caso-controle”, em que o pesquisador parte do presente para o passado dos pesquisados a fim de encontrar alguma associação. Parece uma ideia boa, e é. Só que não tem poder suficiente para confirmar nada, exatamente porque ela não consegue afastar os fatores confundidores. Imagine que, ao final da pesquisa, o pesquisador encontrasse que, realmente, no grupo sem câncer de pulmão, mais pessoas eram usuários daquele anti-hipertensivo, mas, ao mesmo tempo, havia uma prevalência excessivamente baixa de fumantes nesse grupo. Opa. O fumo. “Não tinha pensado no fumo”.

Pois é, o tabagismo é, sabidamente, um fator de risco para câncer de pulmão. Se você colocou muitos pacientes tabagistas em um grupo, é



possível que isso, por si só, já justifique os seus achados.

## O peso dos fatores confundidores desconhecidos e a necessidade de alocar pacientes em grupos diferentes por sorteio

Tabagismo é um fator confundidor “conhecido”. Saiba que existem, ainda, os **fatores confundidores “desconhecidos”**: aqueles que nem foram descobertos ainda e, por isso, é impossível que o estatístico trate deles. Um exemplo de fator confundidor desconhecido seria alguma alteração genética que só se descobrirá daqui a 50 anos.

Os estatísticos conhecem diversas maneiras de evitar que um estudo do tipo observacional tenha fatores confundidores (estratificação, padronização, regressão logística...), mas, em pesquisas muito grandes, o pesquisador pode acabar “esquecendo” de mensurar um ou outro, e também lembre-se que, em estudos observacionais, é impossível controlar os fatores confundidores desconhecidos (Figura 2).<sup>2</sup>

**Figura 2.**

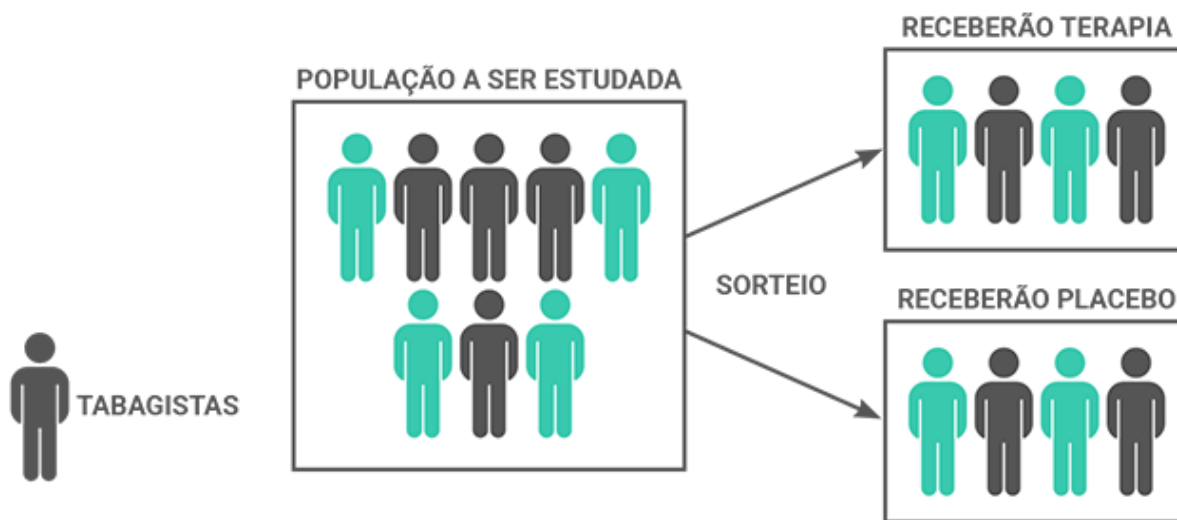


Os estudos observacionais são capazes de gerar ou fortalecer hipóteses porque, pela presença dos fatores confundidores conhecidos (possíveis de se controlar, mas por “possíveis” não entenda “fáceis”) e, nomeadamente, pela presença dos fatores confundidores desconhecidos, é impossível que se afaste a presença do acaso na amostra. Em uma análise bayesiana, esse seria um estudo com índice alto de vieses (Capítulo 5).

Agora o *trial*. Pela natureza experimental desse tipo de estudo, o pesquisador tem a oportunidade de randomizá-los, ou seja, distribuí-los entre

os grupos de maneira aleatória, desde o início da pesquisa. Por que isso é tão importante? Porque, assim, a probabilidade de que um tabagista entre em cada um dos grupos é de 50%, assim como a probabilidade dos hipertensos, e também como alguém que tenha uma alteração genética desconhecida. Ou seja, o sorteio, aumenta as chances de que esse estudo resolva, de fato, o problema dos fatores confundidores (Figura 3).

**Figura 3.**



A randomização (ou sorteio) em estudos experimentais deixa iguais as chances que alguém com alguma comorbidade ou fator confundidor (mesmo que desconhecido) caia no grupo terapia ou no grupo controle.

É por conta dessa randomização, e também de outros mecanismos capazes de evitar a ação de fatores confundidores e vieses, que os trials estão no topo da pirâmide fluida da Medicina Baseada em Evidências (MBE) (Capítulo 3). Um estudo como esse, quando evita vieses, e tem um resultado positivo e estatisticamente significativo, tem imenso peso no cálculo da probabilidade pós-teste, pelo método bayesiano ensinado no Capítulo 5.

## O trial testou superioridade ou de não inferioridade?

Um *trial* pode ser desenhado para avaliar a superioridade de um medicamento versus placebo ou outro, mas também para verificar a não inferioridade, que é diferente da superioridade.

Quando um *trial* de não inferioridade é feito, quer se provar que ele não é inferior à terapia usual, aceitando-se uma pequena margem em que o medicamento pode estar em “inferioridade” sem que seja considerado inferior. A Figura 4 ajuda a entender melhor isso.

**Figura 4.**

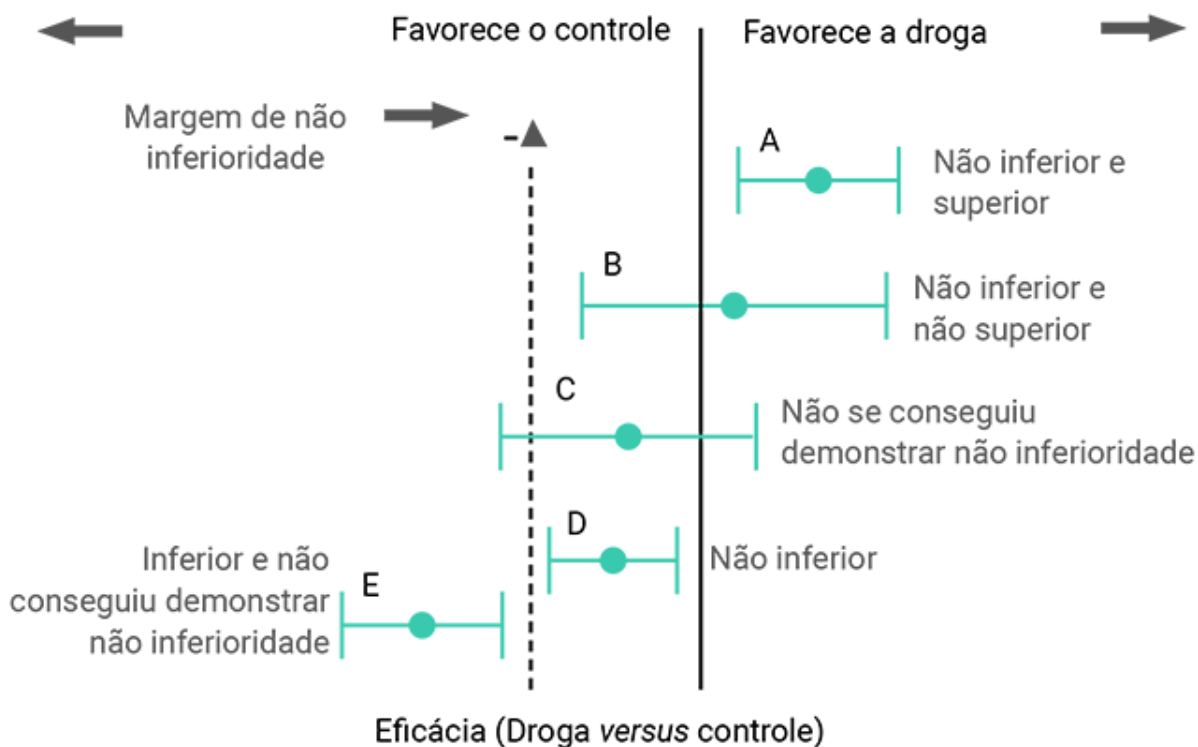


Gráfico do tipo *forest-plot* que representa os intervalos de confiança 95% de diferentes estudos – esses desenhos que mais parecem *TIE fighters* de *Star Wars* representam isso: se a asa ultrapassar a linha do 1, então não houve superioridade. O *TIE fighter* A, por exemplo, não cruza a linha do 1, então ela foi superior, já o B cruza, não atingindo superioridade. A novidade, nos estudos de não inferioridade, é que se desenha uma nova linha preestabelecida (a margem de não inferioridade) e, se o intervalo de confiança não a cruzar, então pode se afirmar que a terapia é não inferior: os *TIE fighters* B e D, por exemplo, são não inferiores.

Isso pode ser útil em casos que a nova terapia ofereça vantagens óbvias ao paciente em comparação à terapia vigente (por “óbvio”, entenda: algo que eu não preciso de um estudo pra me mostrar). Um exemplo muito claro são os estudos com os anticoagulantes diretos (Rivaroxabana,<sup>3</sup> Edoxabana,<sup>4</sup> Apixabana<sup>5</sup> e Dabigatrana<sup>6</sup>) na fibrilação atrial testados versus varfarina (a terapia já consolidada). No caso dos anticoagulantes diretos, a vantagem óbvia que eles apresentam sobre a varfarina é que, com eles, não é

necessário realizar periodicamente o teste do INR. Em outros casos, podem existir vantagens como: ser mais barato, mais conveniente, menos invasivo.<sup>7</sup>

Existem ainda, mas são menos comuns, os estudos de equivalência.

## Passo 2 – o estudo foi cego?

### Um balde de água fria em quem pensa que Medicina é uma ciência objetiva

Se você pensa que a análise de um raio X ou da pressão venosa jugular, ou mesmo de sintomas contados pelo paciente é objetiva, está redondamente enganado. Nós, médicos, encontramos o que esperamos achar. Quando eu digo que “esperamos” achar algo, eu me refiro a uma série de vieses cognitivos que nós, médicos, possuímos, que nos atrapalham na elaboração de hipóteses diagnósticas, a citar apenas um, o “viés de ancoragem”, quando temos uma ideia inicial e nos ancoramos nela, negligenciando novas informações que poderiam ser úteis àquele diagnóstico. Para evitar isso, sempre recomendo aos meus alunos que sistematizem o atendimento.

Pois é, existe até um índice, chamado de kappa, que mede a variabilidade interobservador de achados clínicos (ou seja, quando médicos discordam entre si). O achado de broncogramas aéreos na interpretação de raio X pediátrico, por exemplo, é de 0,13.<sup>8</sup> O índice kappa de médicos interpretando eletrocardiogramas para buscar supradesnivelamento do segmento ST em pacientes com suspeita de infarto agudo do miocárdio é de 0,33.<sup>9</sup> O índice kappa, em seu artigo original, pode ir de  $-1$  até  $+1$ , passando pelo 0. Valores  $\leq 0$  indicam nenhuma concordância entre médicos, valores entre 0,01 e 0,20 têm pouca concordância, entre 0,21 e 0,40 alguma concordância, 0,41 a 0,60 moderada concordância, 0,61 a 0,80 concordância substancial e 0,81 a 1 concordância quase perfeita.<sup>10</sup>

“Cegar os médicos” significa deixá-los inconscientes de quais pacientes são de um grupo ou de outro, assim, também a subjetividade vai incidir igual entre os grupos. Entre nós, também é importante para evitar que médicos deem alguma vantagem aos dados da nova terapia testada, algo que poderia acontecer caso eles acreditem muito na terapia ou caso estejam recebendo algum financiamento para isso.

Mas não só os médicos precisam ser “cegados”, os pacientes também. Isso porque os pacientes são ainda mais subjetivos que os médicos na análise

das suas próprias queixas. Como exemplo, uma pessoa pode sentir um alívio em sua dor apenas por ter recebido uma boa notícia, ou por receber um abraço da mãe. Isso tem até nome, é o efeito Hawthorne, e é confundido com “efeito placebo” (como veremos mais à frente). A maneira de cegar o paciente é fazendo com que tanto a terapia quanto o placebo se pareçam o mesmo. Assim, o paciente não saberá em qual grupo foi destacado.

## Desfechos duros ou moles? Ou quanto um participante ou um pesquisador podem nos enganar?

Então, por enquanto, é isso: **gostamos de estudos duplo-cegos** (aqueles em que médicos e pacientes estão inconscientes de quem está no grupo terapia e quem está no grupo controle) porque isso reduz a subjetividade na análise dos desfechos. Mas todo desfecho é subjetivo? Perceba, pelo efeito Hawthorne (de se sentir mais bem cuidado), alguém pode até ter menos dor, ou achar que aquele espirro que ainda continua nem é tão importante assim e já relatar melhora ao pesquisador. Mas o desfecho mais “duro” que existe em pesquisas clínicas é o de mortalidade, em que o efeito Hawthorne não incide. A diferença entre “*hard*” e “*soft endpoints*”, ou, em português, **desfechos duros e desfechos moles** é que no desfecho duro há mínima ou nenhuma subjetividade e variabilidade interobservador na interpretação daquilo. Na interpretação do autor que vos fala, os únicos desfechos duros que existem são a morte e sinais clínicos inequívocos (como “sangramento” – ninguém discorda se um paciente sangrou). A Tabela 1 exemplifica desfechos duros e moles.

**Tabela 1.**

Desfechos duros	Desfechos moles
Morte por qualquer causa	Morte por uma causa específica
Sinais clínicos inequívocos, como sangramento	Incidência de algum diagnóstico como “infarto”
	Redução de hospitalizações
	Redução de complicações, como tempo de internação ou sintomas, tosse, dor

Desfechos duros	Desfechos moles
	Graduação de sinais clínicos, mesmo que inequívocos, como classificar “sangramento” em “sangramento maior” ou “sangramento menor”
Resultados de exames laboratoriais ou de imagem (sujeitos a erros – nenhum exame é 100% acurado).	
Sintomas referidos pelos pacientes (se “sangramento” for algo não flagrado, mas referido pelo paciente, já passa a ser um “sintoma”, portanto, desfecho mole)	

Lista de desfechos duros e desfechos moles em pesquisas clínicas. Mesmo que se estudem desfechos duros (não sujeitos à interpretação dos médicos – nenhum médico discorda se alguém está morto ou não), o ideal é que, sempre que factível, os estudos sejam duplo-cegos, porque lembrem-se que “números não mentem, mas pessoas sim”.

É importante, caro leitor, que você entenda que o “cegamento” é muito difícil, senão impossível quando se trata de especialidades cirúrgicas: seria, obviamente, antiético “simular” uma cirurgia em um paciente apenas para que ele não saiba se ele é do grupo que operou ou não, porque, de uma cirurgia, no mínimo, saem pontos e suturas.<sup>11</sup> O que dá pra fazer é comparar técnicas cirúrgicas diferentes, mas daí nasce um problema: testar técnicas cirúrgicas diferentes e comprovar benefício de uma ou de outra não comprova que essa técnica vencedora é melhor do que não fazer nada – essa resposta só poderia ser dada testando terapia x nada (placebo).

Isso, aliás, nos leva ao passo 3.

### Passo 3 – o estudo comparou o que com o quê?

O estudo pode comprar terapia *versus* ausência de terapia ou terapia *versus* outra terapia. Se o estudo for cego, a comparação terapia *versus* ausência de terapia precisa ser feita com o uso de um placebo, algo que se pareça com a terapia e que tem o intuito de ajudar com que o paciente e/ou o médico não percebam a qual grupo pertencem (Figura 5).

**Figura 5.**



A comparação em ensaios clínicos randomizados deve ser feita de um grupo *versus* o controle. Normalmente se testam terapias *versus* nada ou terapias *versus* outras terapias para o mesmo fim. Caso queira se cegar o grupo que recebe “nada”, é preciso que se use o placebo. É para isso que o placebo existe.

Deve-se fazer o grupo controle usar uma terapia quando essa já for comprovadamente eficaz e se queira testar uma substituta a ela.<sup>12</sup>

Ao contrário do que muitos pensam, o denominado **“efeito placebo” não é unanimidade** e sua existência e definição geram controvérsia, que está descrita no Quadro 1.

#### Quadro 1. Efeito placebo: um mal-entendido

A tradição do “efeito placebo” data de um artigo de 1955 de Henry Beecher denominado “The Powerful Placebo”<sup>38</sup> em que ele afirma que os placebos aliviaram os sintomas de 35% dos 1.082 pacientes de 15 estudos revisados por ele. Hoje, os placebos são considerados “efetivos” em quase todas as doenças. Mas, se são efetivos, por que não se transformam em terapias?

Na verdade, a concepção que nos é passada pela mídia e, também, pela esmagadora maioria da literatura médica é, possivelmente, fruto de um viés cognitivo que provém da confusão que pessoas fazem entre efeito placebo *versus* efeito Hawthorne.

Partindo dessa premissa de que placebos não são realmente eficazes, porque se fossem teriam que ser vendidos, temos uma maior liberdade para fazer uma análise crítica desse efeito e começaremos fazendo isso elaborando critérios para que o efeito placebo exista verdadeiramente em uma determinada doença:

- Placebo tem que ter sido dado e comparado a outra terapia.
- O efeito não apareceu em quem não recebeu placebo.
- O efeito tem que ser clinicamente relevante.

Impondo esses três critérios que são mais do que justos (você há de concordar), nenhum dos artigos citados por Bleecher em seu estudo seminal verdadeiramente demonstraram o tal “efeito placebo”. O que pode causar essa falsa sensação de melhora, então? A principal razão é a de que, normalmente, em um *trial* só há dois braços: o grupo terapia (que recebe medicamento) e o grupo controle (que recebe placebo). Como não existe um terceiro grupo em que nada é dado, não há como saber se o placebo de fato demonstrou algum efeito. Observe a figura do quadro para entender.

**Figura do quadro 1.**

TERAPIA	X	PLACEBO		Terapia	Placebo	Interpretação
Desfecho estudado: mortalidade			Trial A	12,7	10,8	Terapia danosa
			Trial B	10,8	10,7	Terapia ineficaz
			Trial C	10,7	12,7	Terapia eficaz

São demonstrados resultados de três hipotéticos estudos comparando terapia *versus* placebo. Observe que em todas as situações, a interpretação é sobre a terapia: ela é eficaz no *trial C* porque reduziu mortalidade em comparação ao placebo, é ineficaz no *trial B* e é danosa no *trial A* porque aumentou a mortalidade em comparação ao placebo. Se alguém interpretasse o *trial A* como “placebo eficaz”, então aquele comprimido de farinha teria que ser, agora, chamado de “medicamento” e ser testado *versus* um composto diferente de placebo. Mas isso não existe, não é o placebo que estudamos em um *trial*, mas sim sua terapia.

Há ainda outras razões pelas quais o efeito placebo parece existir (mas, de acordo com essa linha de pensamento descrita aqui, não existe): melhora espontânea, que é o seguimento da tão falada “história natural da doença” e especialmente o efeito Hawthorne, que ocorre quando pessoas sabem que estão sendo observadas (ou, nesse caso, “melhor cuidadas”).<sup>39</sup>

## Passo 4 – a hipótese é plausível? Tem boa probabilidade pré-teste?

Já discutimos isso em profundidade no Capítulo 5. O raciocínio bayesiano, em comparação ao frequentista, solicita que nós avaliemos o resultado de um estudo sob a luz dos dados que já conhecemos do mundo real (plausibilidade, outros critérios de Bradford Hill, outros estudos com melhor ou pior metodologia que o que você está lendo agora ou até mesmo em níveis mais baixos da pirâmide fluida da MBE...). Se o frequentista lê um artigo e observa um resultado positivo com *p* significativo, ele assume imediatamente aquele novo dado como verdadeiro (caso seja um frequentista mais experiente, ele avaliará também os vieses para saber se acredita ou não naquela pesquisa). O bayesiano não: ele vai aplicar o teorema de Bayes com



os resultados da pesquisa e os dados prévios para entender como essa nova pesquisa impactará no conhecimento dele.



## Passo 5 – o estudo teve sensibilidade adequada?



Para que o estudo tenha sensibilidade (capacidade de ser positivo se a hipótese realmente for verdadeira), ele precisa ter um tamanho suficiente de amostra, ou seja, de participantes estudados. E esse tamanho é calculado a partir do que se consideraria que seria diferença clinicamente significativa com a nova droga. Por exemplo, eu posso testar um anti-hipertensivo com objetivo de demonstrar que ele, em comparação ao placebo (ou seja, nada), reduz 20 mmHg de pressão arterial sistólica em um exame de medida ambulatorial de pressão arterial (MAPA) com um intervalo de confiança predeterminado por mim. Posso também testar se um novo medicamento reduz em 16% a letalidade por uma determinada doença.

Além disso, para o cálculo do tamanho amostral, também precisamos definir, por antecipação, qual a taxa máxima de erro do tipo “falso-positivo” que eu posso ter na minha pesquisa. Vamos falar disso daqui a pouco, mas por enquanto, saiba que essa taxa é, normalmente, de 5%.

Com esses números, há diversas calculadoras que podem ser usadas para se obter tamanho de amostra, por exemplo, o <https://clincalc.com/stats/samplesize.aspx> (Figura 6).

**Figura 6.**


 vs.   
 Two independent study groups


 vs.   
 One study group vs. population

Two study groups will each receive different treatments.

---

### Primary Endpoint

  
 Dichotomous (yes/no)

  
 Continuous (means)

The primary endpoint is **binomial** - only two possible outcomes.  
*Eg, mortality (dead/not dead), pregnant (pregnant/not)*

---

### Statistical Parameters

#### Anticipated Incidence

Group 1 ?  %

Group 2 ?  %

Incidence ⌵

#### Type I/II Error Rate

Alpha ?

Power ?

Reset

Calculate

*Print screen* do site ClinCalc e da sua seção “*sample size*” que se traduz como “tamanho da amostra”. Observe que para demonstrar uma redução relativa de 16% (de 12% para 10%) de algum desfecho dicotômico (do tipo “sim” ou “não”), usando um alfa de 5% (taxa máxima de erro do tipo “falso positivo”) e uma sensibilidade (*power*) de 80%, clica-se em “*calculate*” para obter a resposta, que, nesse caso, foi de 7.682 pacientes. Ou seja, para se ter uma sensibilidade de 80% e uma especificidade de 95% de demonstrar uma diferença de 16% entre os grupos, eu preciso estudar 7.682 pessoas.

**A sensibilidade** (ou, na linguagem de artigos, o “*power*”) mínima buscada em artigos científicos médicos **varia de 80% a 90%**. Ao ler um artigo, observe se os autores escreveram (na seção “metodologia”) qual o *power* que eles buscaram para aquela pesquisa e se o objetivo deles era

cl clinicamente relevante ou apenas um desfecho “substituto” (daqui a pouco falaremos disso).

## Passo 6 – o que os desfechos do estudo têm a nos dizer?

### Desfecho clinicamente relevante ou substituto?

O ideal em uma pesquisa clínica é que se busque um desfecho clinicamente relevante, isto é, que esta nova terapia aumente a qualidade de vida ou a quantidade de dias vividos. Por mais que você tenha aprendido na graduação que um determinado medicamento é bom porque “melhora os valores de qualquer exame que seja (por exemplo: TSH, hemoglobina glicada, testosterona, valor de pressão arterial etc.)”, a partir de agora você precisa treinar seu cérebro pra fazer a seguinte pergunta: “Sim, mas qual o benefício *clínico* deste paciente?”.

Exatamente porque valores de exames são “**desfechos substitutos**” ou, em inglês, “*surrogate outcomes*”.<sup>13</sup> Um desfecho substituto é algum marcador (laboratorial ou clínico) que substitua o desfecho clínico porque reflete a doença em si ou algum fator de risco da doença em si. Observe a Figura 7 e as Tabelas 2 e 3 para ter uma ideia melhor do que são esses desfechos substitutos.

Figura 7.



Muitas variáveis incidem no curso desde o desfecho substituto até o desfecho clínico.

Tabela 2.

Condição pesquisada	Desfecho substituto	Desfecho clínico significativo a ser evitado
---------------------	---------------------	--

Condição pesquisada	Desfecho substituto	Desfecho clínico significativo a ser evitado
Hipertensão	Valor da pressão arterial	Infarto, AVC ou morte
Hipercolesterolemia	Valor do painel lipídico	Infarto, AVC ou morte
Câncer	Existência e tamanho do tumor e marcadores tumorais	Morte
Diabetes	Glicemia e hemoglobina glicada	Infarto, AVC, morte ou outros desfechos cardiovasculares
Doença de Alzheimer	Exames de imagem	Disfunção funcional e morte
Vacinas	Resposta sorológica	Proteção à doença
Doenças renais	Valor da creatinina	Evolução para hemodiálise

Lista de desfechos substitutos em diversas áreas. Observe que apesar de que sabidamente eles são um reflexo ou um prenúncio do desfecho clínico, não significam necessariamente que o paciente vai evoluir daquilo para o desfecho clínico.

**Tabela 3.**

Droga	Indicação	Desfecho substituto que era encontrado com o uso dessas drogas	Desfecho clínico real encontrado quando passaram por <i>trials</i>
Clofibrato	Pacientes com hipercolesterolemia	Redução dos níveis de colesterol	Aumento de mortalidade <sup>35</sup>
Flecainida	Extrassístoles ventriculares em pacientes pós-infarto agudo do miocárdio	Redução das extrassístoles no Holter	Aumento de mortalidade <sup>36</sup>
Rosiglitazona	Pacientes com diabetes tipo 2	Redução dos níveis de hemoglobina glicada	Aumentou a incidência de infarto agudo do miocárdio <sup>37</sup>

Drogas que eram usadas pelo seu mecanismo e pelos desfechos substitutos que encontravam, mas que, quando testadas em ensaios clínicos que buscavam desfechos clínicos, acabaram se provando malélicas.

Os desfechos substitutos são frequentemente usados em pesquisas clínicas porque podem ser medidos ou encontrados em fases mais precoces

da doença (mas isso leva ao problema do *overdiagnosis*, porque, como vimos na Figura 7 e Tabela 2, não necessariamente o desfecho substituto levará ao desfecho clínico final e verdadeiro); são mais convenientes e reduzem o tamanho das amostras necessárias para obter sensibilidade, especificidade e significância estatística, reduzindo também o tamanho da amostra; e podem servir para pavimentar o caminho de uma terapia até a sua aprovação como tratamento adequado para aquela condição.

Por exemplo, o grupo dos inibidores da proteína SGLT2 são antidiabéticos que têm, consistentemente, demonstrado benefício do seu uso também em doentes com insuficiência cardíaca e doença renal crônica. O estudo DAPA-CKD demonstrou que a dapaglifozina (um dos medicamentos dessa classe de drogas) reduziu o declínio da taxa de filtração glomerular (medida por meio dos valores da creatinina) dos pacientes em relação ao placebo.<sup>14</sup> O leitor precisa ficar atento que não está, ainda (não por esse estudo), comprovado que dapaglifozina reduza o desfecho “evolução para hemodiálise”, esse sim, clinicamente relevante, mas abriu espaço para um estudo mais robusto com esse desfecho.

## A invasão dos desfechos compostos

Outra sacada que o leitor precisa ficar atento é que, às vezes, para reduzir o tamanho da amostra necessária (com isso o custo e a dificuldade logística), os pesquisadores podem criar um **desfecho composto**. Muito comum na cardiologia, desfecho composto é a soma de diversos desfechos duros ou moles, e às vezes até substitutos, por exemplo: os MACE (“*major adverse cardiac events*”), que incluem insuficiência cardíaca, reinfarto não fatal, dor anginosa recorrente, re-hospitalização por causas cardíacas, necessidade de repetição de intervenção coronária percutânea, realização de cirurgia de revascularização miocárdica e morte por causas cardíacas.<sup>15</sup>

Um exemplo disso é o estudo COMPLETE, não cego, que randomizou pacientes tratados nas últimas 72 horas de infarto com supradesnivelamento do segmento ST para receber stents nas outras coronárias não culpadas pelo infarto que, por acaso, tivessem lesões ocupando > 70% da sua luz *versus* nada. O desfecho primário foi um composto de morte por causas cardiovasculares (desfecho mole) ou novo infarto do miocárdio (desfecho ainda mais mole, pois pode várias muito conforme a taxa de falsos-positivos da troponina), obtendo diferença significativa (7,8% x 10,5%, p: 0,004, IC95%: 0,6-0,91) a favor da estratégia de implantar stents em todas as

coronárias doentes, não só na que era culpada pelo infarto. Os problemas: (a) sendo um estudo não cego, deveriam se evitar os desfechos moles; (b) do desfecho composto, apenas “novo infarto do miocárdio” foi estatisticamente significativo, um desfecho muito mole.<sup>16</sup>

Os desfechos compostos podem ser muito problemáticos especialmente pela falácia do “*cherry picking*”, que, neste cenário, ocorre quando um pesquisador, diante de uma pesquisa que foi negativa, escolhe, entre todos os seus dados, aquele que teve um resultado positivo e o inclui na análise *post-hoc*, modificando, de última hora, o desfecho primário estudado. O leitor deve lembrar que isso não é correto porque, para calcular o tamanho amostral e a sensibilidade da amostra, o pesquisador tinha que predefinir quais desfechos ele esperava encontrar e quantos desfechos esperava encontrar (normalmente, isso se faz baseando-se na literatura prévia).

## Desfechos primários ou secundários?

O último conceito que o leitor precisa aprender sobre os desfechos é que existem desfechos primários e secundários. Desfecho primário é aquele que foi a premissa do pesquisador – todo o desenho do estudo e a quantidade de participantes na pesquisa foi baseada nele. Desfechos secundários são resultados adicionais, também pré-especificados, que podem adicionar conhecimento ou ajudar a interpretar os dados do desfecho primário.<sup>17</sup> Se o desfecho primário for negativo e o desfecho secundário positivo, deve se considerar aquele achado como “gerador de hipóteses” e se conduzir um novo *trial* com aquele desfecho secundário sendo o novo desfecho primário. Por exemplo, o estudo ASCOT, negativo em seu desfecho primário “redução de mortalidade”, mas positivo no desfecho secundário “redução de AVC, eventos cardiovasculares totais, morte por qualquer causa e diabetes de diagnóstico recente” quando comparava amlodipino *versus* atenolol para hipertensão arterial, dando vantagem ao amlodipino.<sup>18</sup> Sendo bem criterioso, um novo *trial* deveria ter sido realizado para testar essa hipótese, visto que, possivelmente, testar “redução de mortalidade” foi muito ambicioso da parte dos pesquisadores – poucos medicamentos na Medicina chegam a esse patamar.

Ao avaliar múltiplos desfechos, que é o caso dos desfechos compostos e também dos desfechos secundários, o leitor pode usar a fórmula de Bonferroni, que consiste em dividir o p perseguido (em Medicina, por

convenção, tipicamente  $< 0,05$ ) pela quantidade de variáveis pesquisadas, sendo esse resultado o novo valor de  $p$ .

## Passo 7 – qual a especificidade do resultado encontrado?

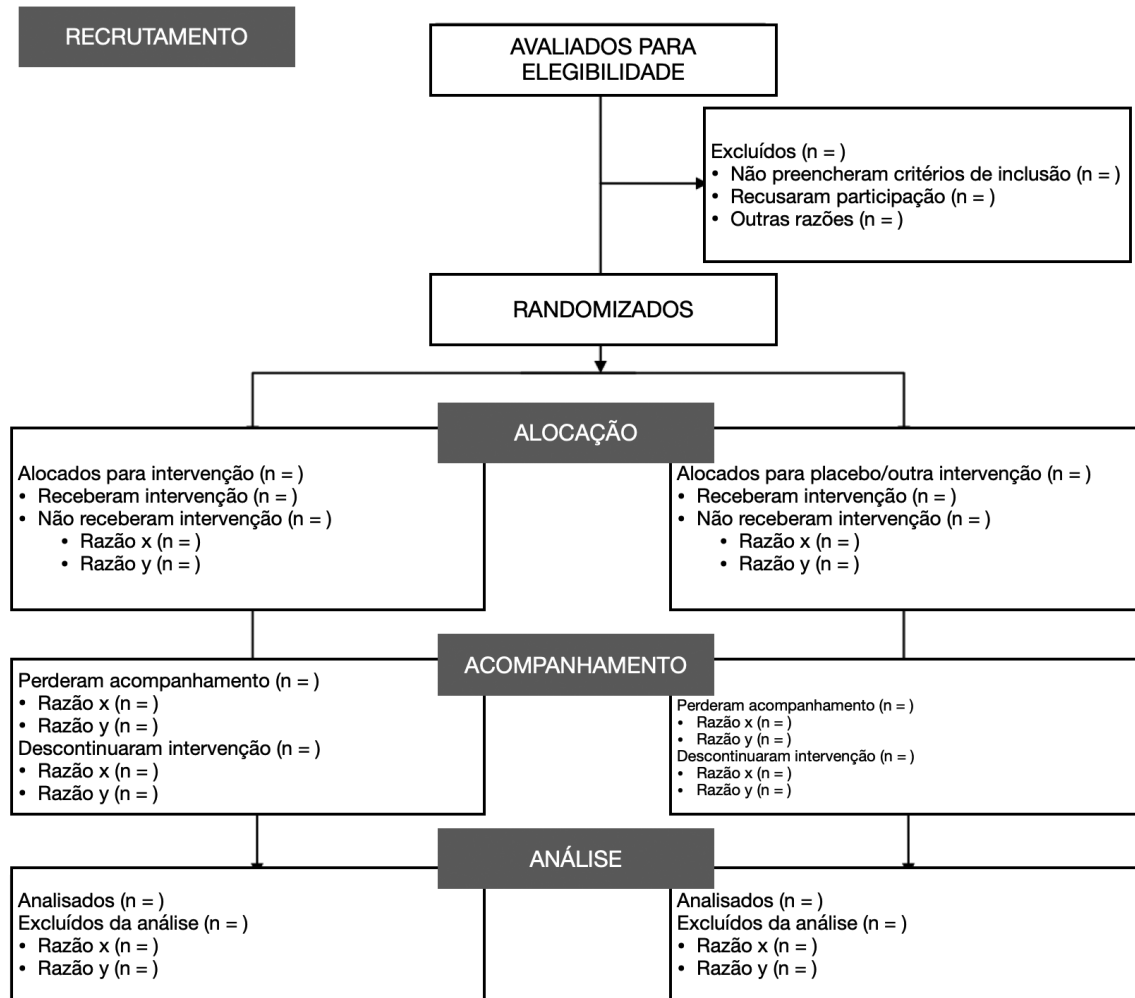
Em outras palavras, “eu coloco a minha mão no fogo por esse resultado?”. Alerta para os leitores: na forma como vem sendo produzida a pesquisa médica atualmente, quase nunca pode se pôr a mão no fogo.

Após a análise da metodologia, vem a seção “Resultados”, que é, tipicamente, regida pelas regras do consenso CONSORT, atualizado pela última vez em 2010.<sup>19</sup> Basicamente, você encontrará as informações nessa ordem:

1. Fluxograma ilustrando o progresso da pesquisa e seus dois grupos – quantos pacientes foram incluídos e excluídos, como foi sua alocação, quantos perderam *follow-up* e como foi a sua análise (Figura 8).
2. Recrutamento: em que mês e ano se iniciou o recrutamento de pacientes? Quando terminou? Quando terminou o acompanhamento? O estudo foi finalizado antes de que toda a amostra prevista fosse recrutada? Por quê?
3. Tabela de *baseline*. Esta, provavelmente, será a Tabela 1 do artigo que você está lendo. Ela ilustra os dados demográficos e clínicos dos participantes prévios à intervenção. Quantos eram hipertensos? Quantos diabéticos? É importante para que o leitor veja se a amostra está homogênea, isto é, se, estatisticamente, os pacientes dos dois grupos se parecem, evitando o impacto dos fatores confundidores. Às vezes, e de maneira inadequada, os *trialistas* colocam o valor de  $p$  para cada comparação, sendo que, aqui, se espera um valor de  $p > 0,05$  (pois não se deseja diferença significativa entre os grupos). Mas cuidado: (a) o uso dessa ferramenta é ilógico, visto que a randomização é, por definição, um processo realizado ao acaso; (b) se um fator confundidor tiver efeito muito potente no desfecho estudado, mesmo diferenças estatisticamente não significativas podem ser relevantes e enviesar

o resultado da pesquisa.<sup>20</sup> Observe uma tabela desse tipo na Figura 9.

**Figura 8.**



Fluxograma de fases do estudo conforme sugerido pelo consenso CONSORT de 2010. Este gráfico provavelmente será a Figura 1 do *trial* que você está lendo. É importante, aqui, ficar atento a possíveis vieses: se foram excluídos muitos pacientes, será que os pesquisadores estavam selecionando pacientes ideais (apenas os que eles já achavam que a terapia testada teria efeito)? Se muitos pacientes perderam o acompanhamento, então o estudo perdeu seu *power*, tendo menor sensibilidade.

**Figura 9.**



	NGT group (n=697)	GDM group (n=177)	P values
Age (years)	32.73±5.10	33.44±5.28	0.111
Race/Ethnicity			
Caucasian	470 (67.4)	117 (66.1)	0.032
Hispanic	207 (29.7)	53 (29.9)	
Others	20 (2.9)	7 (4.1)	
Family history			
Type 2 diabetes	138 (19.8)	42 (23.7)	0.450
MetS (>2 components)	145 (20.8)	38 (21.5)	
History			
GDM	15 (2.2)	10 (5.7)	0.043
Miscarriage	225 (32.3)	53 (29.9)	
Educational status			
Elementary education	53 (7.6)	25 (14.2)	0.030
Secondary school	167 (24.0)	45 (25.4)	
University degree	471 (67.6)	106 (59.9)	
UNK	6 (0.9)	2 (1.2)	
Employment	539 (77.3)	135 (76.3)	0.844
Number of pregnancies			
Primiparous	301 (43.2)	77 (43.5)	0.857
Second pregnancy	225 (32.3)	56 (31.6)	
>2 pregnancies	171 (24.5)	44 (25.9)	
Smoker			
Never	384 (55.1)	93 (52.5)	0.786
Current	55 (7.9)	17 (9.6)	
Gestational age (weeks) at entry	12.1±0.5	12.1±0.7	0.261
Body weight (kg)			
Prepregnancy	59.7±10.2	65.1±12.0	0.0001
At entry	61.8±10.5	67.2±12.3	0.0001
Weight gain at entry	2.0±2.9	2.2±2.9	0.508
BMI (kg/m <sup>2</sup> )			
Prepregnancy	22.6±3.5	24.9±4.2	0.0001
At baseline	23.4±3.6	25.7±4.4	0.0001
Systolic BP (mm Hg)	106±11	111±10	0.001
Diastolic BP (mm Hg)	65±10	67±8	0.063
FBG (mg/dL)	81±6	83±6	0.0001
HbA <sub>1c</sub> % (mmol/mol)	5.2±0.3 (33±0.9)	5.0±0.3 (31±0.9)	0.165
Cholesterol (mg/dL)	173±29	177±31	0.288
Triglycerides (mg/dL)	80±41	89±37	0.034
MEDAS score	4.90±1.73	4.89±1.56	0.931
Nutrition Score	0.37±3.2	0.27±3.16	0.699
Physical Activity Score ≥0	80 (12.4)	19 (10.8)	0.694

Data are mean±SD or number (%).

Physical Activity Score ≥0: (1) walking daily (>5 days / week); score 0: at least 30 min; score +1: if >60 min; score -1: if <30 min. (2) Climbing stairs (floors / day, >5 days a week): score 0, between 4 and 16; score +1, >16; score -1: <4.

BMI, body mass index;BP, blood pressure;FBG, fasting blood glucose;GDM, gestational diabetes mellitus;HbA<sub>1c</sub>, hemoglobin A<sub>1c</sub>;MEDAS, Mediterranean Diet Adherence Screener;MetS, metabolic syndrome;NGT, normal glucose tolerance;UNK, unknown.

Exemplo de uma tabela de *baseline*, a famosa “tabela 1”. Observe que aqui estão descritos dados demográficos e características clínicas que podem atuar como fatores confundidores em uma pesquisa.

Entenda que cada pesquisa pode incluir dados diferentes na Tabela 1 porque as várias doenças possuem fatores de risco e confundidores diferentes. Perceba também que os valores de p estão dispostos na última coluna dessa tabela, uma das maneiras encontradas para tentar definir significância estatística nessa diferença (que, nesse caso, o ideal é que não exista) e passando a ideia de que o sorteio realizado para alocação dos pacientes no estudo não foi suficiente para homogeneizar o IMC pré-gravidez (em inglês, *BMI prepregnancy*) das participantes desse estudo, havendo um valor de p significativo nessa comparação. Se esse fator confundidor for muito fortemente relevante para o desfecho estudado, essa diferença pode enviesar o resultado a ponto de ele ficar desacreditado. Há dois problemas nessa interpretação (a de adicionar valor de p à “Tabela 1”): (a) a randomização é, por definição, um processo realizado ao acaso (sorteio) e, por isso, o valor de p é inútil; (b) mesmo as diferenças não significativas no *baseline* podem gerar vieses importantes nos desfechos estudados.

4. Número de pessoas analisadas e a informação do tipo de análise feita: *intention-to-treat* ou *per-protocol*? Vamos explicar essa diferença adiante.
5. Resultados para cada desfecho primário, composto e secundário em cada grupo seguida da sua significância estatística (valor de p) e do seu intervalo de confiança (IC 95%). Também será analisado adiante.
6. Análises adjuntas: resultados das subanálises que podem **gerar hipóteses** para novos estudos, caso o desfecho primário tenha sido conseguido. Também será analisado adiante.
7. Resultados sobre os danos existentes em cada grupo.

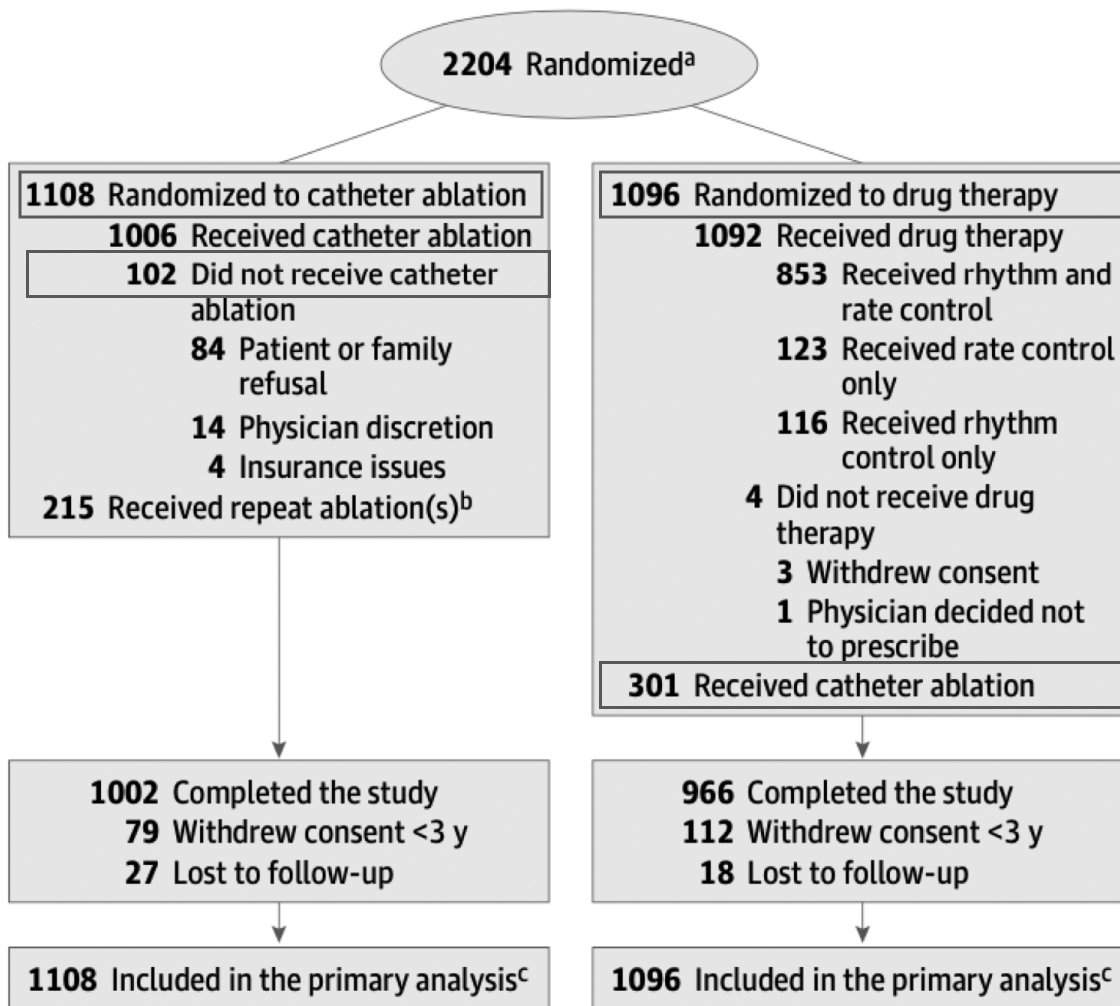
## Análise pela intenção de tratar ou pelo protocolo?

Imagine que um pesquisador quer comprovar o benefício de uma terapia intervencionista *versus* o tratamento farmacológico já existente para uma doença. Ele, com todo o esmero, recruta pacientes e os aloca nos grupos terapia *versus* controle. Só que, depois do acompanhamento, na hora de fazer a análise final do estudo, ele percebe que houve > 25% de migração dos pacientes de um grupo para o outro. Isto é: pacientes que começaram a pesquisa no braço “intervenção” acabaram não fazendo a intervenção, e pacientes que começaram a pesquisa no braço “terapia farmacológica” acabaram fazendo a intervenção. E agora, José?

Bem, isso existe e não é raro: o estudo CABANA teve uma taxa de **crossover** (nome que se dá quando o paciente é alocado para um braço, mas acaba recendo outro tratamento) de 27,5% do braço terapia farmacológica

para o braço intervenção (no caso, ablação de fibrilação atrial) e 9,2% do braço intervenção para a terapia farmacológica.<sup>21</sup> Observe a Figura 1 do próprio CABANA *trial* na Figura 10 do nosso capítulo.

**Figura 10.**



A Figura 1 (a do fluxograma de fases do estudo) do CABANA *trial* demonstra imenso peso do *crossover* nessa pesquisa: 9,2% dos pacientes que foram alocados para receber ablação acabaram não recebendo ablação e 27,5% dos pacientes alocados para receber tratamento farmacológico acabaram recebendo ablação. Isso limita fortemente o *power* e a especificidade do estudo.<sup>21</sup>

Agora o pesquisador tem duas opções: fazer a análise do tipo “intenção de tratar (ITT)” ou do tipo “pelo protocolo (PP)”.<sup>22</sup>

Intenção de tratar é o modo mais conservador e analisa todos os pacientes no braço em que eles foram incluídos inicialmente, mesmo que acabaram

fazendo o *crossover*. É conservador porque preserva o estudo de resultados falsos-positivos, isto é, encontrar, na pesquisa, um resultado positivo que não condiz com a vida real. Ou seja, é interessante quando se quer manter uma alta especificidade no estudo.

Já a análise pelo protocolo acaba abrindo mais espaço para o falso-positivo ao reduzir sua especificidade e aumenta a sua sensibilidade (ou *power*), analisando cada pessoa de acordo com a terapia que recebeu. A Figura 11 exemplifica essas análises. Esteja atento a alguns detalhes: a homogeneização que havia sido conseguida no sorteio da randomização pode se perder nesse *crossover*.

**Figura 11.**

TERAPIA	90 receberam terapia	60%, isto é, 54 tiveram benefício
	10 receberam placebo	40%, isto é, 4 tiveram benefício
PLACEBO	90 receberam placebo	40%, isto é, 46 tiveram benefício
	10 receberam terapia	60%, isto é, 6 tiveram benefício
<b>Intention-to-treat:</b>		<b>Per protocol:</b>
Terapia: 54 + 4 tiveram benefício (58%)		Terapia: 54 + 6 tiveram benefício (60%)
Placebo: 46 + 6 tiveram benefício (52%)		Placebo: 46 + 4 tiveram benefício (50%)
Hazard ratio: 6%		Hazard ratio: 10%

Diferença da análise *intention-to-treat* e *per-protocol* em uma pesquisa hipotética que teve 10% de *crossover*. Observe a mudança no *hazard ratio* de 6% para 10%. Isso pode ser o suficiente para se obter ou não significância estatística e convencer ou não a comunidade médica a respeito do uso de uma nova terapia.

## Os resultados são significativos e relevantes?

Nesta fase, o importante é analisar como se comportaram os desfechos entre os grupos. Na seção apropriada, você já aprendeu o que são desfechos relevantes e irrelevantes e no Capítulo 6, já aprendeu que o valor de *p* é a especificidade daquele estudo em específico.

Uma das grandes falhas na análise de um artigo científico é a superconfiança no valor de *p*. Como vimos, o valor de *p* é a probabilidade de que aquele dado encontrado seja tão extremo ou ainda mais extremo na vida real, ou seja, a probabilidade de ser verdadeiro positivo, ou, em outras

palavras, a especificidade. O valor de  $p$  é calculado em função da (a) magnitude do efeito, (b) do tamanho da amostra e (c) da variabilidade dos achados. Ou seja, não é porque um valor de  $p$  foi muito baixo (digamos  $< 0,001$ ), que isso signifique que a magnitude do efeito é alta (na verdade pode ter sido bem marginal, da ordem de 0,1%, caso a amostra seja grande o suficiente) e muito menos sobre a sua relevância (valor de  $p < 0,001$  não transforma um desfecho substituto em um desfecho clínico relevante). Também não é porque o valor de  $p$  foi  $> 0,05$  que a nova terapia se provou ineficaz: pode ter havido uma amostra não suficientemente grande para demonstrar a variabilidade real da droga.

## Análises de subgrupos: uma belíssima oportunidade para cometer “cherry picking” e encontrar algo significativo

A motivação para se realizar uma análise de subgrupos é a de encontrar quais pacientes se beneficiaram mais de uma terapia, baseada em seus fatores demográficos ou de risco. Contudo, essa análise é pobre e sofre de um ganho na possibilidade de falsos-positivos e falsos-negativos, isso porque, não tem *power* suficiente (lembre-se que o tamanho da amostra foi calculado para responder a uma questão maior que a análise de subgrupos que está sendo feita) e porque se baseia na realização de múltiplas comparações.<sup>23</sup>

Uma enorme fonte de vieses pode florescer da análise de subgrupos quando se comete o “*cherry picking*” ou “escolha de cerejas”, que é, ao final da pesquisa, procurar subgrupos em que o medicamento foi efetivo e dar ênfase a essa análise de subgrupo. Isso é especialmente ruim quando o estudo não obteve significância no seu desfecho primário. Escolher, em meio a tantos números, aquele que foi positivo, é dar ao acaso uma grande oportunidade, por isso, o ideal é que os pesquisadores anunciem antes da pesquisa quais serão os subgrupos analisados.

Outra regra importante da análise de subgrupos é que elas são apenas geradoras de hipóteses, especialmente se o estudo teve o desfecho primário significativamente modificado pela terapia. Um exemplo de um *trial* em que isso aconteceu foi o CHARISMA, que testou 15 mil pacientes que usaram aspirina + clopidogrel ou aspirina + placebo em um estudo duplo-cego e de análise *intention-to-treat* de doentes vasculares ou com múltiplos fatores de

risco com o desfecho primário composto tipo MACE negativo (6,8 x 7,3%, CI: 0,83 – 1,05, p: 0,22). A análise de subgrupo nesse estudo, contudo, demonstrou ter havido uma diferença significativa (p: 0,046) beneficiando o grupo que recebeu clopidogrel (e não placebo) quando os pacientes do subgrupo “aterotrombose clínica evidente”. Isso está inapropriadamente descrito na conclusão do artigo e também sem seu *abstract*.<sup>24</sup> O editorial do mesmo artigo na revista *New England* foi muito inteligente ao afirmar: “o carisma de extrair resultados de subgrupos favoráveis deve ser evitado”.<sup>25</sup>

## O desfecho primário falhou – e agora?

Quando o valor de p significativo (usualmente  $< 0,05$ ) não é atingido, significa que não se conseguiu rejeitar a hipótese nula – que era a de que a terapia e o placebo ou nova terapia são iguais. Muitas pessoas confundem isso com “conseguiu se comprovar que não há diferença entre os grupos”. Por isso, e também porque a análise bayesiana (proposta neste livro) é muito mais complexa que a avaliação do valor isolado de p, o leitor precisa entender muito bem o que significa quando um *trial* é negativo.

A primeira pergunta que deve ser feita, antes de sair por aí anunciando que uma terapia é ineficaz, é: o *trial* teve *power* suficiente, ou seja, teve amostra suficientemente grande para demonstrar diferença significativa no resultado? Lembre-se que *power*, ou sensibilidade, é a capacidade de evitar resultados falsos-negativos (ou, como se diz em pesquisas, um erro do tipo 2). Por exemplo, o CIBIS testou se bisoprolol, comparado a placebo, reduzia mortalidade por todas as causas em pacientes com insuficiência cardíaca e teve um resultado negativo (HR: 0,8; p: 0,22; IC 95%: 0,56 – 1,12 – perceba que o p é maior que 0,05 e o IC 95% cruza a linha do 1).<sup>26</sup> Contudo, esse estudo foi considerado *underpowered* (que significa baixa sensibilidade) e, por sorte, os patrocinadores realizaram o CIBIS II, esse sim com *power* adequado, demonstrando uma redução da mortalidade da ordem de 34% (p  $< 0,0001$ , IC95% 0,54 – 0,81).<sup>27</sup>

Outra pergunta importante é: os pesquisadores escolheram bem os desfechos primários? No já citado CABANA *trial*, que estudou ablação de fibrilação atrial *versus* medicamentos em um acompanhamento de 4 anos, os desfechos escolhidos não são intuitivos nem representam as expectativas do tratamento: morte, AVC debilitante, sangramento maior ou parada cardíaca.<sup>21</sup> Não se espera redução desses desfechos com a terapia ablativa – não em tão pouco tempo (apenas 4 anos). É por isso que, às vezes, os desfechos

compostos e até mesmo os desfechos substitutos podem servir – quando o desfecho clinicamente relevante for muito raro ou demorar muito tempo para acontecer.

Mais um fator que deve ser levado em consideração: a população estudada é mesmo aquela em que se espera que haja um benefício? Por exemplo, o SIGNIFY<sup>28</sup> e o BEAUTIFUL<sup>29</sup> *trials* falharam em demonstrar benefício da Ivabradina em pacientes com doença coronária crônica, mas, quando testada em pacientes com insuficiência cardíaca, O SHIFT *trial* conseguiu demonstrar que a Ivabradina reduz significativa nos índices do desfecho composto “morte, morte por causas cardiovasculares ou hospitalização por insuficiência cardíaca” (analise isso sob a visão de quem já conhece os desfechos compostos).<sup>30</sup>

A Tabela 4 resume algumas perguntas que um leitor de artigos precisa se fazer quando um artigo não atinge significância estatística.<sup>31</sup>

#### **Tabela 4.**

O estudo foi *underpowered*?

O desfecho primário foi bem escolhido?

A população estudada foi aquela em que se esperava mesmo haver benefício?

O regime de tratamento foi apropriado?

Houve deficiências na condução do *trial*?

Perguntas a se fazer quando um estudo falha em atingir significância estatística no desfecho primário.<sup>31</sup>

## Os vieses presentes no estudo são capazes de reduzir sua sensibilidade e especificidade?

Você aprendeu no Capítulo 4 os vieses que podem incidir na produção e na análise de estudos clínicos que podem levar à redução de sua sensibilidade e especificidade. Viu, também, no Capítulo 5, que a redução desses valores impacta fortemente na probabilidade pós-teste desses estudos, de acordo com a análise bayesiana. Esses capítulos são muito recomendados para o entendimento completo da análise de um artigo científico.

## Passo 8 – o estudo tem validação externa?

Validação interna é tudo que fizemos do passo 1 ao passo 7. Validação externa é quanto os resultados de um estudo podem ser generalizados para a população.<sup>32</sup> Um tipo de pergunta que pode ser feita para testar a validação externa é: a população do estudo se parece com a população normalmente acometida por essa doença? Algumas vezes, os *trials* podem ter selecionado pacientes jovens demais ou velhos demais, para conseguir obter resultado significativo, mas acabou se distanciando do mundo real.

Também faz parte da validação externa o cálculo da probabilidade pós-teste, que aprendemos a fazer no Capítulo 5.

## Observando cada passo em um artigo real

Vamos analisar cada um dos passos descritos neste capítulo em um artigo real, o SPRINT *trial*.<sup>33</sup> A Tabela 5 resume os passos 1 a 8. As Figuras 12 a 19 demonstram onde encontraríamos os dados requeridos pelos passos 1 a 8 em um artigo real.

**Tabela 5.**

Passos	Perguntas	Respostas possíveis	Vantagens e desvantagens	Onde encontrar essas respostas?
1. Qual o tipo de pesquisa?	1.1 Observacional ou <i>trial</i> ?	1.1.1 Observacional: gerador de hipóteses. 1.1.2 <i>Trials</i> : podem confirmar hipóteses.	1.1.1 Observacional: mais barato e mais rápido, mas não consegue confirmar hipóteses. 1.1.2 <i>Trials</i> : podem confirmar hipóteses, mas nem sempre são factíveis (pode ser antiético testar alguma terapia ou deixar de dar uma terapia a um grupo de pacientes).	Título e metodologia
	1.2 Superioridade ou não inferioridade?	1.2.1 Superioridade: aceita intervalos de confiança que não cruzam a linha do 1. 1.2.2 Não inferioridade: aceita intervalos de confiança que não cruzam a margem de não inferioridade.		



Passos	Perguntas	Respostas possíveis	Vantagens e desvantagens	Onde encontrar essas respostas?
2. O estudo foi cego?	2.1 O estudo foi <i>open label</i> , único-cego ou duplo-cego?	<p>2.1.1 <i>Open label</i>: não foram cegados o pesquisador nem os pesquisados.</p> <p>2.1.2 Único-cego: foram cegados apenas o pesquisador ou o pesquisado.</p> <p>2.1.3 Duplo-cego: o pesquisador e o pesquisado estão “cegos”.</p>	<p>2.1.1 <i>Open label</i>: Fica mais barata, mas muito sujeita à subjetividade dos desfechos, especialmente se forem “moles”.</p> <p>2.1.2 Único-cego: a pesquisa fica mais barata, mas muito sujeita à subjetividade dos desfechos, especialmente se forem “moles”.</p> <p>2.1.3 Duplo-cego: a pesquisa fica mais cara, mas é o padrão-ouro em termos de pesquisa médica</p>	Título e metodologia
3. O estudo comparou o que com o quê?	3.1 O estudo comparou terapia <i>versus</i> placebo ou terapia <i>versus</i> outra terapia? Ou mesmo terapia <i>versus</i> nada?	<p>3.1.1 Terapia <i>versus</i> placebo: é o mais usual – evita o “efeito placebo” e o efeito Hawthorne.</p> <p>3.1.2 Terapia <i>versus</i> outra terapia: pode acontecer se já existir uma terapia comprovadamente eficaz que seja antiético não dar a metade dos doentes.</p> <p>3.1.3 Terapia <i>versus</i> nada: Pode ser dado em estudos em que seja impossível cegar os pacientes (por exemplo: pesquisas com técnicas cirúrgicas) e que não existe ainda uma terapia padrão para aquela doença.</p>		Título e metodologia
4. A hipótese é plausível? Tem boa probabilidade pré-teste?	4.1 Tem boa probabilidade pré-teste?	<p>4.1.1 Sim: a hipótese existe e é forte porque tem plausibilidade, coerência e outros fatores de Bradford Hill.</p> <p>4.1.2 Não: a hipótese é ruim, já passou por vários estudos que foram negativos, é incoerente, não faz sentido do ponto de vista biológico...</p>	<p>4.1.1 Estudos plausíveis já partem de uma boa probabilidade pré-teste, que vai influenciar na sua probabilidade pós-teste.</p> <p>4.1.2 Estudos implausíveis partem de uma baixa probabilidade pré-teste e, mesmo que sejam positivos, serão provavelmente falsos-positivos.</p>	Introdução e referências bibliográficas.

Passos	Perguntas	Respostas possíveis	Vantagens e desvantagens	Onde encontrar essas respostas?
--------	-----------	---------------------	--------------------------	---------------------------------

5. O estudo teve sensibilidade adequada?	5.1 O tamanho amostral requerido pela pesquisa para se obter resultado significativo foi conseguido?	<p>5.1.1 Sim: os pesquisadores falaram que queriam x pesquisados e conseguiram finalizar a pesquisa analisando esse número de pessoas.</p> <p>5.1.2 Não: os pesquisadores não conseguiram selecionar pacientes porque eram raros demais, ou porque não queriam participar de pesquisas, ou porque o patrocinador cortou os investimentos.</p>		Metodologia e Resultados
--	--	---	--	--------------------------

6. O que os desfechos do estudo têm a nos dizer?	6.1 Desfecho clinicamente relevante ou substituto?	<p>6.1.1 Clinicamente relevante: desfecho que significa melhora da qualidade de vida ou aumento da quantidade de dias vividos pelo paciente.</p> <p>6.1.2 Substituto: um marcador laboratorial ou clínico que seja precoce como prenúncio de que um desfecho relevante possa acontecer no futuro.</p>	Usualmente se preferem os desfechos clinicamente relevantes. Mas, em casos de doenças cujo desfecho clínico pode demorar muito para acontecer (o que encareceria a pesquisa), os desfechos substitutos podem ser a única opção.	Metodologia e resultados
	6.2 Desfecho único ou composto?	<p>6.2.1 Desfecho único: quando um só desfecho é procurado.</p> <p>6.2.2 Desfecho composto: quando uma soma de desfechos é procurada, usualmente na tentativa de reduzir o tamanho amostral necessário para se obter significância estatística.</p>	Normalmente se preferem os desfechos únicos, mas desfechos compostos podem ser úteis quando são clinicamente relevantes.	
	6.3 Desfecho primário ou secundário?	<p>6.3.1 Desfecho primário: aquele para o qual o tamanho amostral foi calculado para demonstrar significância.</p> <p>6.3.2 Desfecho secundário: resultados adicionais.</p>		

Passos	Perguntas	Respostas possíveis	Vantagens e desvantagens	Onde encontrar essas respostas?
	7.1 A análise foi por intenção de tratar ou pelo protocolo?	7.1.1 Intenção de tratar: quando se analisam os pacientes de acordo com o grupo em que foram alocados, mesmo que tenha havido <i>crossover</i> . 7.1.2 Pelo protocolo: quando se analisam os pacientes de acordo com o tratamento que receberam.	Usualmente se prefere a análise por intenção de tratar, por ser mais pragmática e evitar erros do tipo 1 (falsos-positivos).	
	7.2 Os grupos do estudo foram homogêneos?	7.2.1 Sim: os grupos possuem prevalências similares de vários fatores de risco. 7.2.2 Não: houve diferença significativa e/ou relevante de prevalência de fatores de risco ou demográficos na pesquisa.	Se os fatores confundidores não forem controlados pela randomização, a pesquisa perde muito em especificidade.	
7. Qual a especificidade do resultado encontrado?	7.3 Os resultados são significativos?	7.3.1 Sim: com p significativo e IC 95% que não cruza a linha do 1. 7.3.2 Não: não obteve p significativo e/ou o IC 95% cruza a linha do 1.	Significância estatística é diferente de relevância, porque só quer dizer se os dados encontrados naquela pesquisa podem ser falsos-positivos naquela amostra, ou seja, é a especificidade do estudo.	Metodologia, resultados e discussão
	7.4 Os resultados são relevantes?	7.4.1 Sim. 7.4.2 Não.	Procuramos resultados relevantes: que melhoram a qualidade de vida ou a quantidade de dias vividos pelos pacientes.	
	7.5 A análise de subgrupos ajudou ou atrapalhou?	7.5.1 Ajudou. 7.5.2 Atrapalhou.	A análise de subgrupos pode ajudar a precisar ainda mais os resultados de uma pesquisa, mas também podem atrapalhar, caso sejam interpretados isoladamente, especialmente se o desfecho primário da pesquisa for negativo.	
	7.6 Há vieses?	7.6.1 Sim. 7.6.2 Não.	Vieses reduzem a sensibilidade e a especificidade das pesquisas.	
8. O estudo tem validação externa?	8.1 O estudo será capaz de modificar a prática clínica?  8.2 A probabilidade pós-teste é alta?	8.1.2 Sim. 8.1.3 Não.  8.2.2 Sim. 8.2.3 Não.	-	Discussão do artigo e análise realizada pelo próprio leitor (usando este livro, por exemplo, como referência).

Resumo dos passos sugeridos para análise de um artigo científico sobre uma terapia.

## Figura 12.

Passo 1: tipo de pesquisa  
Resposta: trial  
randomizado controlado



### A Randomized Trial of Intensive versus Standard Blood-Pressure Control

The SPRINT Research Group\*

#### ABSTRACT

##### BACKGROUND

The most appropriate targets for systolic blood pressure to reduce cardiovascular morbidity and mortality among persons without diabetes remain uncertain.

##### METHODS

We randomly assigned 9361 persons with a systolic blood pressure of 130 mm Hg or higher and an increased cardiovascular risk, but without diabetes, to a systolic blood-pressure target of less than 120 mm Hg (intensive treatment) or a target of less than 140 mm Hg (standard treatment). The primary composite outcome was myocardial infarction, other acute coronary syndromes, stroke, heart failure, or death from cardiovascular causes.

##### RESULTS

At 1 year, the mean systolic blood pressure was 121.4 mm Hg in the intensive-treatment group and 136.2 mm Hg in the standard-treatment group. The intervention was stopped early after a median follow-up of 3.26 years owing to a significantly lower rate of the primary composite outcome in the intensive-treatment group than in the standard-treatment group (1.65% per year vs. 2.19% per year; hazard ratio with intensive-treatment, 0.75; 95% confidence interval [CI], 0.64 to 0.89;  $P < 0.001$ ). All-cause mortality was also significantly lower in the intensive-treatment group (hazard ratio, 0.73; 95% CI, 0.60 to 0.90;  $P = 0.003$ ). Rates of serious adverse events of hypotension, syncope, electrolyte abnormalities, and acute kidney injury or failure, but not of injurious falls, were higher in the intensive-treatment group than in the standard-treatment group.

##### CONCLUSIONS

Among patients at high risk for cardiovascular events but without diabetes, targeting a systolic blood pressure of less than 120 mm Hg, as compared with less than 140 mm Hg, resulted in lower rates of fatal and nonfatal major cardiovascular events and death from any cause, although significantly higher rates of some adverse events were observed in the intensive-treatment group. (Funded by the National Institutes of Health; ClinicalTrials.gov number, NCT01206062.)

The members of the writing committee (Jackson T. Wright, Jr., M.D., Ph.D., Jeff D. Williamson, M.D., M.H.S., Paul K. Whelton, M.D., Joni K. Snyder, R.N., B.S.N., M.A., Kaycee M. Sink, M.D., M.A.S., Michael V. Rocco, M.D., M.S.C.E., David M. Reboussin, Ph.D., Mahboob Rahman, M.D., Suzanne Oparil, M.D., Cora E. Lewis, M.D., M.S.P.H., Paul L. Kimmel, M.D., Karen C. Johnson, M.D., M.P.H., David C. Goff, Jr., M.D., Ph.D., Lawrence J. Fine, M.D., Dr.P.H., Jeffrey A. Cutler, M.D., M.P.H., William C.ushman, M.D., Alfred K. Cheung, M.D., and Walter T. Ambrosius, Ph.D.) assume responsibility for the overall content and integrity of the article. The affiliations of the members of the writing group are listed in the Appendix. Address reprint requests to Dr. Wright at the Division of Nephrology and Hypertension, University Hospitals Case Medical Center, Case Western Reserve University, 1100 Euclid Ave. Cleveland, OH 44106-6053, or at jackson.wright@case.edu.

\*A complete list of the members of the Systolic Blood Pressure Intervention Trial (SPRINT) Research Group is provided in the Supplementary Appendix, available at NEJM.org.

This article was published on November 9, 2015, at NEJM.org.

N Engl J Med 2015;373:2103-16.

DOI: 10.1056/NEJMoa1511959

Copyright © 2015 Massachusetts Medical Society.

Página 1 do estudo SPRINT. Aqui, pelo título, já descobrimos que se trata de um estudo do tipo “*trial*” controlado e randomizado. Também já se percebe pelo título que é uma comparação de um tratamento mais intensivo *versus* o tratamento padrão para controle da pressão arterial. O leitor deve ficar agora curioso: mas o que será controle intensivo? Não seja apressado e não leia o *abstract*.

## Figura 13.

Passo 4: Na introdução se apresenta a plausibilidade da hipótese: há estudos prévios que apontam para essa mesma direção? Há plausibilidade biológica?

**H**YPERTENSION IS HIGHLY PREVALENT in the adult population in the United States, especially among persons older than 60 years of age, and affects approximately 1 billion adults worldwide.<sup>1,2</sup> Among persons 50 years of age or older, isolated systolic hypertension is the most common form of hypertension,<sup>3,4</sup> and systolic blood pressure becomes more important than diastolic blood pressure as an independent risk predictor for coronary events, stroke, heart failure, and end-stage renal disease (ESRD).<sup>5,13</sup> The Global Burden of Disease Study identified elevated blood pressure as the leading risk factor, among 67 studied, for death and disability-adjusted life-years lost during 2010.<sup>14</sup>

Clinical trials have shown that treatment of hypertension reduces the risk of cardiovascular disease outcomes, including incident stroke (by 35 to 40%), myocardial infarction (by 15 to 25%), and heart failure (by up to 64%).<sup>5,15,16</sup> However, the target for systolic blood-pressure lowering is uncertain. Observational studies have shown a progressive increase in cardiovascular risk as systolic blood pressure rises above 115 mm Hg,<sup>10</sup> but the available evidence from randomized, controlled trials in the general population of patients with hypertension only documents the benefit of treatment to achieve a systolic blood-pressure target of less than 150 mm Hg, with limited data concerning lower blood-pressure targets.<sup>11,17,21</sup> In a trial involving patients with type 2 diabetes mellitus, the rate of major cardiovascular events was similar with a systolic blood-pressure target of less than 120 mm Hg and the commonly recommended target of less than 140 mm Hg, though the rate of stroke was lower with the target of less than 120 mm Hg.<sup>22</sup> A recent trial involving patients who had had a stroke compared treatment to lower systolic blood pressure to less than 130 mm Hg with treatment to lower it to less than 150 mm Hg and showed no significant benefit of the lower target with respect to the overall risk of another stroke but a significant benefit with respect to the risk of hemorrhagic stroke.<sup>23</sup>

The hypothesis that a lower systolic blood-pressure goal (e.g., <120 mm Hg) would reduce clinical events more than a standard goal was designated by a National Heart, Lung, and Blood Institute (NHLBI) expert panel in 2007 as the most important hypothesis to test regarding the prevention of hypertension-related complications

among patients without diabetes.<sup>24</sup> The current article describes the primary results of the Systolic Blood Pressure Intervention Trial (SPRINT), which compared the benefit of treatment of systolic blood pressure to a target of less than 120 mm Hg with treatment to a target of less than 140 mm Hg.

#### METHODS

##### STUDY DESIGN AND OVERSIGHT

SPRINT was a randomized, controlled, open-label trial that was conducted at 102 clinical sites (organized into 5 clinical center networks) in the United States, including Puerto Rico (see the Supplementary Appendix, available with the full text of this article at NEJM.org). A trial coordinating center served as a data and biostatistical core center and supervised the central laboratory, the electrocardiography reading center, the magnetic resonance imaging reading center, and the drug-distribution center. The rationale and protocol for the trial are publicly available,<sup>25,26</sup> and the protocol is available at NEJM.org.

SPRINT was sponsored by the NHLBI, with cosponsorship by the National Institute of Diabetes and Digestive and Kidney Diseases, the National Institute of Neurological Disorders and Stroke, and the National Institute on Aging. An independent data and safety monitoring board monitored unblinded trial results and safety events. The study was approved by the institutional review board at each participating study site. The steering committee designed the study, gathered the data (in collaboration with investigators at the clinics and other study units), made the decision to submit the manuscript for publication, and vouches for the fidelity of the study to the protocol. The writing committee wrote the manuscript and vouches for the completeness and accuracy of the data and analysis. The coordinating center was responsible for analyzing the data. Scientists at the National Institutes of Health participated in the design of the study and as a group had one vote on the steering committee of the trial.

##### STUDY POPULATION

Participants were required to meet all the following criteria: an age of at least 50 years, a systolic blood pressure of 130 to 180 mm Hg (see the Supplementary Appendix), and an increased risk

Passo 2: Foi cego?  
Resposta: não, foi *open-label* (o oposto de duplo-cego)

Na segunda página do SPRINT, podemos aprender sobre a sua probabilidade pré-teste, que é o passo 4 (há outros estudos apontando para a mesma direção ou a hipótese “caiu do céu”?) e também responder ao passo 2 quando, logo na primeira linha da seção “métodos”, se lê que o estudo é do tipo “*open-label*”.

## Figura 14.

of cardiovascular events. Increased cardiovascular risk was defined by one or more of the following: clinical or subclinical cardiovascular disease other than stroke; chronic kidney disease, excluding polycystic kidney disease, with an estimated glomerular filtration rate (eGFR) of 20 to less than 60 ml per minute per 1.73 m<sup>2</sup> of body-surface area, calculated with the use of the four-variable Modification of Diet in Renal Disease equation; a 10-year risk of cardiovascular disease of 15% or greater on the basis of the Framingham risk score; or an age of 75 years or older. Patients with diabetes mellitus or prior stroke were excluded. Detailed inclusion and exclusion criteria are listed in the Supplementary Appendix. All participants provided written informed consent.

#### RANDOMIZATION AND INTERVENTIONS

Eligible participants were assigned to a systolic blood-pressure target of either less than 140 mm Hg (the standard-treatment group) or less than 120 mm Hg (the intensive-treatment group). Randomization was stratified according to clinical site. Participants and study personnel were aware of the study-group assignments, but outcome adjudicators were not.

After the participants underwent randomization, their baseline antihypertensive regimens were adjusted on the basis of the study-group assignment. The treatment algorithms were similar to those used in the Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial.<sup>22</sup> These algorithms and our formulary are listed in Figures S1 and S2 and Table S1 in the Supplementary Appendix. All major classes of antihypertensive agents were included in the formulary and were provided at no cost to the participants. SPRINT investigators could also prescribe other antihypertensive medications (not provided by the study). The protocol encouraged, but did not mandate, the use of drug classes with the strongest evidence for reduction in cardiovascular outcomes, including thiazide-type diuretics (encouraged as the first-line agent), loop diuretics (for participants with advanced chronic kidney disease), and beta-adrenergic blockers (for those with coronary artery disease).<sup>5,27</sup> Chlorthalidone was encouraged as the primary thiazide-type diuretic, and amlodipine as the preferred calcium-channel blocker.<sup>28,29</sup> Azilsartan and azilsartan combined with chlorthalidone were donated by

Takeda Pharmaceuticals International and Arbor Pharmaceuticals; neither company had any other role in the study.

Participants were seen monthly for the first 3 months and every 3 months thereafter. Medications for participants in the intensive-treatment group were adjusted on a monthly basis to target a systolic blood pressure of less than 120 mm Hg. For participants in the standard-treatment group, medications were adjusted to target a systolic blood pressure of 135 to 139 mm Hg, and the dose was reduced if systolic blood pressure was less than 130 mm Hg on a single visit or less than 135 mm Hg on two consecutive visits. Dose adjustment was based on a mean of three blood-pressure measurements at an office visit while the patient was seated and after 5 minutes of quiet rest; the measurements were made with the use of an automated measurement system (Model 907, Omron Healthcare). Lifestyle modification was encouraged as part of the management strategy. Retention in the study and adherence to treatment were monitored prospectively and routinely throughout the trial.<sup>26</sup>

#### STUDY MEASUREMENTS

Demographic data were collected at baseline. Clinical and laboratory data were obtained at baseline and every 3 months thereafter. A structured interview was used in both groups every 3 months to obtain self-reported cardiovascular disease outcomes. Although the interviewers were aware of the study-group assignments, they used the same format for interviews in the two groups to minimize ascertainment bias. Medical records and electrocardiograms were obtained for documentation of events. Whenever clinical-site staff became aware of a death, a standard protocol was used to obtain information on the event.

Serious adverse events were defined as events that were fatal or life-threatening, that resulted in clinically significant or persistent disability, that required or prolonged a hospitalization, or that were judged by the investigator to represent a clinically significant hazard or harm to the participant that might require medical or surgical intervention to prevent one of the other events listed above.<sup>30,31</sup> A short list of monitored conditions were reported as adverse events if they were evaluated in an emergency department:

Passo 3: O estudo comparou um alvo de pressão arterial sistólica de 140 mmHg *versus* um alvo de pressão de 120 mmHg.

Aqui, ele descreve como conseguiu esses alvos e como fez essas medidas

Na página 3 do artigo, observamos que o estudo não comparou terapias *versus* placebo ou *versus* outras terapias, mas sim um alvo específico de pressão arterial sistólica (140 mmHg) *versus* outro (120 mmHg).

### Figura 15.

Passo 6: os desfechos procurados são um composto de infarto do miocárdio, síndrome coronária aguda não resultante em infarto, insuficiência cardíaca agudamente descompensada, AVC ou morte por causas cardiovasculares

hypotension, syncope, injurious falls, electrolyte abnormalities, and bradycardia. We also monitored occurrences of acute kidney injury or acute renal failure if they were noted on admission or occurred during a hospitalization and were reported in the hospital discharge summary as a primary or main secondary diagnosis. The *Medical Dictionary for Regulatory Activities* was used to classify the safety events. Coding was performed at the coordinating center, and up to three codes were assigned to each safety event. The relationship of serious adverse events to the intervention was assessed by the trial safety officer and reviewed monthly by the safety committee.

**STUDY OUTCOMES**

Definitions of study outcomes are outlined in the Supplementary Appendix. A committee whose members were unaware of the study-group assignments adjudicated the clinical outcomes specified in the protocol. The primary hypothesis was that treatment to reach a systolic blood-pressure target of less than 120 mm Hg, as compared with a target of less than 140 mm Hg, would result in a lower rate of the composite outcome of myocardial infarction, acute coronary syndrome not resulting in myocardial infarction, stroke, acute decompensated heart failure, or death from cardiovascular causes. Secondary outcomes included the individual components of the primary composite outcome, death from any cause, and the composite of the primary outcome or death from any cause.

We also assessed renal outcomes, using a different definition for patients with chronic kidney disease (eGFR <60 ml per minute per 1.73 m<sup>2</sup>) at baseline and those without it. The renal outcome in participants with chronic kidney disease at baseline was a composite of a decrease in the eGFR of 50% or more (confirmed by a subsequent laboratory test) or the development of ESRD requiring long-term dialysis or kidney transplantation. In participants without chronic kidney disease at baseline, the renal outcome was defined by a decrease in the eGFR of 30% or more to a value of less than 60 ml per minute per 1.73 m<sup>2</sup>. Incident albuminuria, defined for all study participants by a doubling of the ratio of urinary albumin (in milligrams) to creatinine (in grams) from less than 10 at baseline to greater than 10 during follow-up, was also a prespecified renal outcome.

Prespecified subgroups of interest for all outcomes were defined according to status with respect to cardiovascular disease at baseline (yes vs. no), status with respect to chronic kidney disease at baseline (yes vs. no), sex, race (black vs. non-black), age (<75 vs. ≥75 years), and baseline systolic blood pressure in three levels (≤132 mm Hg, >132 to <145 mm Hg, and ≥145 mm Hg). We also planned a comparison of the effects of systolic blood-pressure targets on incident dementia, changes in cognitive function, and cerebral small-vessel ischemic disease; these results are not presented here.

**STATISTICAL ANALYSIS**

We planned a 2-year recruitment period, with a maximum follow-up of 6 years, and anticipated a loss to follow-up of 2% per year. With an enrollment target of 9250 participants, we estimated that the trial would have 88.7% power to detect a 20% effect with respect to the primary outcome, assuming an event rate of 2.2% per year in the standard-treatment group.

Our primary analysis compared the time to the first occurrence of a primary outcome event between the two study groups with the use of the intention-to-treat approach for all randomly assigned participants; for this analysis, we used Cox proportional-hazards regression with two-sided tests at the 5% level of significance, with stratification according to clinic. Follow-up time was censored on the date of last event ascertainment. Interactions between treatment effect and prespecified subgroups were assessed with a likelihood-ratio test for the interaction with the use of Hommel-adjusted P values.<sup>32</sup> Interim analyses were performed for each meeting of the data and safety monitoring board, with group-sequential stopping boundaries defined with the use of the Lan-DeMets method with an O'Brien-Fleming-type spending function.<sup>33</sup> The Fine-Gray model for the competing risk of death was used as a sensitivity analysis.<sup>34</sup>

**RESULTS**

**STUDY PARTICIPANTS**

A total of 9361 participants were enrolled between November 2010 and March 2013 (Fig. 1). Descriptive baseline statistics are presented in Table 1. On August 20, 2015, the NHLBI director accepted a recommendation from the data and

Passo 5: A sensibilidade foi adequada? Se fossem analisados 9520 participantes, haveria um power de 88,7% assumindo uma redução esperada de 20% do desfecho composto

Passo 5: A sensibilidade foi adequada? Foram analisados 9361 participantes

A página 4 do SPRINT (ainda a seção “metodologia”) traz informações que respondem às perguntas dos passos 5 e 6.

**Figura 16.**

safety monitoring board of the trial to inform the investigators and participants of the cardiovascular-outcome results after analyses of the primary outcome exceeded the monitoring boundary at two consecutive time points (Fig. S3 in the Supplementary Appendix), thus initiating the process to end the blood-pressure intervention early. The median follow-up on August 20, 2015, was 3.26 years of the planned average of 5 years.

**BLOOD PRESSURE**

The two treatment strategies resulted in a rapid and sustained between-group difference in systolic blood pressure (Fig. 2). At 1 year, the mean systolic blood pressure was 121.4 mm Hg in the intensive-treatment group and 136.2 mm Hg in the standard-treatment group, for an average difference of 14.8 mm Hg. The mean diastolic blood pressure at 1 year was 68.7 mm Hg in the intensive-treatment group and 76.3 mm Hg in the standard-treatment group (Fig. S4 in the Supplementary Appendix). Throughout the 3.26 years of follow-up, the mean systolic blood pressure was 121.5 mm Hg in the intensive-treatment group and 134.6 mm Hg in the standard-treatment group, and the mean number of blood-pressure medications was 2.8 and 1.8, respectively. The relative distribution of antihypertensive medication classes used was similar in the two groups, though the use of each class was greater in the intensive-treatment group (Table S2 in the Supplementary Appendix).

**CLINICAL OUTCOMES**

A primary outcome event was confirmed in 562 participants — 243 (1.65% per year) in the intensive-treatment group and 319 (2.19% per year) in the standard-treatment group (hazard ratio with intensive treatment, 0.75; 95% confidence interval [CI], 0.64 to 0.89;  $P < 0.001$ ) (Table 2). Separation in the primary outcome between the groups was apparent at 1 year (Fig. 3A). The between-group differences were consistent across the components of the primary outcome and other prespecified secondary outcomes (Table 2).

A total of 365 deaths occurred — 155 in the intensive-treatment group and 210 in the standard-treatment group (hazard ratio, 0.73; 95% CI, 0.60 to 0.90;  $P = 0.003$ ). Separation in mortality between the groups became apparent at approximately 2 years (Fig. 3B). Causes of death are provided in Table S3 in the Supplementary Ap-

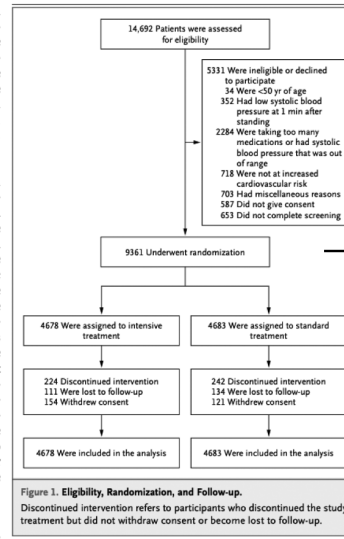


Figure 1. Eligibility, Randomization, and Follow-up. Discontinued intervention refers to participants who discontinued the study treatment but did not withdraw consent or become lost to follow-up.

Passo 7: a clássica “figura 1” com o fluxograma ilustrando o progresso da pesquisa e o seu recrutamento. Não foram reportados “crossovers”, mas houve perdas de pacientes

Passo 7: os resultados foram demonstrados em *hazard ratio*, valor de *p* e IC95%

pendix. The relative risk of death from cardiovascular causes was 43% lower with the intensive intervention than with the standard treatment ( $P = 0.005$ ) (Table 2).

The numbers needed to treat to prevent a primary outcome event, death from any cause, and death from cardiovascular causes during the median 3.26 years of the trial were 61, 90, and 172, respectively. The effects of the intervention on the rate of the primary outcome and on the rate of death from any cause were consistent across the prespecified subgroups (Fig. 4, and Fig. S5 in the Supplementary Appendix). There were no significant interactions between treatment and subgroup with respect to the primary outcome or death from any cause. When death

Passo 7: os NNTs estão dispostos nessa seção

A página 5 do SPRINT traz o fluxograma do progresso da pesquisa e também seus resultados. Nem todos os estudos traduzem seus resultados a NNTs, mas esse o fez.

**Figura 17.**



Passo 7: a famosa “tabela 1” traz um panorama das prevalências de diversos fatores demográficos ou de risco nos dois grupos da pesquisa. Leia com atenção e veja se consegue perceber alguma diferença significativa (numericamente larga) ou relevante

→

Characteristic	Intensive Treatment (N=4678)	Standard Treatment (N=4683)
Criterion for increased cardiovascular risk — no. (%)†		
Age ≥75 yr	1317 (28.2)	1319 (28.2)
Chronic kidney disease‡	1330 (28.4)	1316 (28.1)
Cardiovascular disease	940 (20.1)	937 (20.0)
Clinical	779 (16.7)	783 (16.7)
Subclinical	247 (5.3)	246 (5.3)
Framingham 10-yr cardiovascular disease risk score ≥15%	2870 (61.4)	2867 (61.2)
Female sex — no. (%)	1684 (36.0)	1648 (35.2)
Age — yr		
Overall	67.9±9.4	67.9±9.5
Among those ≥75 yr of age	79.8±3.9	79.9±4.1
Race or ethnic group — no. (%)§		
Non-Hispanic black	1379 (29.5)	1423 (30.4)
Hispanic	503 (10.8)	481 (10.3)
Non-Hispanic white	2698 (57.7)	2701 (57.7)
Other	98 (2.1)	78 (1.7)
Black race¶	1454 (31.1)	1493 (31.9)
Baseline blood pressure — mm Hg		
Systolic	139.7±15.8	139.7±15.4
Diastolic	78.2±11.9	78.0±12.0
Distribution of systolic blood pressure — no. (%)		
≤132 mm Hg	1583 (33.8)	1553 (33.2)
>132 mm Hg to <145 mm Hg	1489 (31.8)	1549 (33.1)
≥145 mm Hg	1606 (34.3)	1581 (33.8)
Serum creatinine — mg/dl	1.07±0.34	1.08±0.34
Estimated GFR — ml/min/1.73 m <sup>2</sup>		
Among all participants	71.8±20.7	71.7±20.5
Among those with estimated GFR ≥60 ml/min/1.73 m <sup>2</sup>	81.3±15.5	81.1±15.5
Among those with estimated GFR <60 ml/min/1.73 m <sup>2</sup>	47.8±9.5	47.9±9.5
Ratio of urinary albumin (mg) to creatinine (g)	44.1±178.7	41.1±152.9
Fasting total cholesterol — mg/dl	190.2±41.4	190.0±40.9
Fasting HDL cholesterol — mg/dl	52.9±14.3	52.8±14.6
Fasting total triglycerides — mg/dl	124.8±85.8	127.1±95.0
Fasting plasma glucose — mg/dl	98.8±13.7	98.8±13.4
Statin use — no./total no. (%)	1978/4645 (42.6)	2076/4640 (44.7)
Aspirin use — no./total no. (%)	2406/4661 (51.6)	2350/4666 (50.4)
Smoking status — no. (%)		
Never smoked	2050 (43.8)	2072 (44.2)
Former smoker	1977 (42.3)	1996 (42.6)
Current smoker	639 (13.7)	601 (12.8)
Missing data	12 (0.3)	14 (0.3)
Framingham 10-yr cardiovascular disease risk score — %	20.1±10.9	20.1±10.8

A página 6 do SPRINT traz a famosa “Tabela 1” que é a que traz os fatores demográficos e de risco em cada grupo do estudo. O ideal aqui é que não haja muitas discrepâncias, porque foi pra isso que os pacientes foram sorteados – para que os fatores incidissem igualmente.

## Figura 18.

Passo 7: a famosa "tabela 1" continua

Characteristic	Intensive Treatment (N = 4678)	Standard Treatment (N = 4683)
Body-mass index <sup>†</sup>	29.9±5.8	29.8±5.7
Antihypertensive agents — no./patient	1.8±1.0	1.8±1.0
Not using antihypertensive agents — no. (%)	432 (9.2)	450 (9.6)

\* Plus-minus values are means ±SD. There were no significant differences (P<0.05) between the two groups except for statin use (P=0.04). To convert the values for creatinine to micromoles per liter, multiply by 88.4. To convert the values for cholesterol to millimoles per liter, multiply by 0.02586. To convert the values for triglycerides to millimoles per liter, multiply by 0.01129. To convert the values for glucose to millimoles per liter, multiply by 0.05551. GFR denotes glomerular filtration rate, and HDL high-density lipoprotein.

<sup>†</sup> Increased cardiovascular risk was one of the inclusion criteria.

<sup>‡</sup> Chronic kidney disease was defined as an estimated glomerular filtration rate of less than 60 ml per minute per 1.73 m<sup>2</sup> of body-surface area.

<sup>§</sup> Race and ethnic group were self-reported.

<sup>¶</sup> Black race includes Hispanic black and black as part of a multiracial identification.

<sup>||</sup> The body-mass index is the weight in kilograms divided by the square of the height in meters.

was treated as a competing risk in a Fine-Gray model, the results with respect to the primary outcome were virtually unchanged (hazard ratio, 0.76; 95% CI, 0.64 to 0.89).

Among participants who had chronic kidney disease at baseline, no significant between-group difference in the composite outcome of a decrease in the eGFR of 50% or more or the development of ESRD was noted, though the number of events was small (Table 2). Among participants who did not have chronic kidney disease at baseline, the incidence of the outcome defined by a decrease in the eGFR of 30% or more to a value of less than 60 ml per minute per 1.73 m<sup>2</sup> was higher in the intensive-treatment group than in the standard-treatment group (1.21% per year vs. 0.35% per year; hazard ratio, 3.49; 95% CI, 2.44 to 5.10; P<0.001).

#### SERIOUS ADVERSE EVENTS

Serious adverse events occurred in 1793 participants in the intensive-treatment group (38.3%) and in 1736 participants in the standard-treatment group (37.1%) (hazard ratio with intensive treatment, 1.04; P=0.25) (Table 3, and Table S4 in the Supplementary Appendix). Serious adverse events of hypotension, syncope, electrolyte abnormalities, and acute kidney injury or acute renal failure, but not injurious falls or bradycardia, occurred more frequently in the intensive-treatment group than in the standard-treatment group. Orthostatic hypotension as assessed during a clinic visit was significantly less common in the intensive-treatment group. A total of 220

participants in the intensive-treatment group (4.7%) and 118 participants in the standard-treatment group (2.5%) had serious adverse events that were classified as possibly or definitely related to the intervention (hazard ratio, 1.88; P<0.001) (Table S5 in the Supplementary Appendix). The magnitude and pattern of differences in adverse events according to treatment assignment among participants 75 years of age or older were similar to those in the overall cohort (Table S6 in the Supplementary Appendix).

#### DISCUSSION

SPRINT showed that among adults with hypertension but without diabetes, lowering systolic blood pressure to a target goal of less than 120 mm Hg, as compared with the standard goal of less than 140 mm Hg, resulted in significantly lower rates of fatal and nonfatal cardiovascular events and death from any cause. Trial participants assigned to the lower systolic blood-pressure target (intensive-treatment group), as compared with those assigned to the higher target (standard-treatment group), had a 25% lower relative risk of the primary outcome; in addition, the intensive-treatment group had lower rates of several other important outcomes, including heart failure (38% lower relative risk), death from cardiovascular causes (43% lower relative risk), and death from any cause (27% lower relative risk). During the follow-up period of the trial (median, 3.26 years), the number needed to treat with a strategy of intensive blood-pressure control to

Passo 8: a leitura da "discussão" é importante porque é aqui que o autor tentará convencer você que o estudo dele é capaz de mudar a sua conduta, ou seja, tem validade externa. Ele citará também as limitações do seu estudo. Alie essas informações à sua interpretação pessoal para chegar a uma conclusão.

Passo 7: dados sobre os danos aos pacientes

A página 7 do SPRINT segue com a Tabela 1 e termina a seção "resultados" falando dos danos que podem ter sido causados aos pacientes. E aí se inicia a seção "discussão", em que o autor vai explicar melhor seus resultados e tentar convencer você de que a validação externa dos resultados é possível.

Alie essa leitura ao seu conhecimento de análise bayesiana para chegar a essa conclusão.

## Figura 19.

Na página 9, o autor apresenta os resultados primários e secundários em forma de tabela

Table 2. Primary and Secondary Outcomes and Renal Outcomes.\*

Outcome	Intensive Treatment		Standard Treatment		Hazard Ratio (95% CI)	P Value
	no. of patients (%)	% per year	no. of patients (%)	% per year		
<b>All participants</b>	<b>(N=4678)</b>		<b>(N=4683)</b>			
Primary outcome†	243 (5.2)	1.65	319 (6.8)	2.19	0.75 (0.64–0.89)	<0.001
Secondary outcomes						
Myocardial infarction	97 (2.1)	0.65	116 (2.5)	0.78	0.83 (0.64–1.09)	0.19
Acute coronary syndrome	40 (0.9)	0.27	40 (0.9)	0.27	1.00 (0.64–1.55)	0.99
Stroke	62 (1.3)	0.41	70 (1.5)	0.47	0.89 (0.63–1.25)	0.50
Heart failure	62 (1.3)	0.41	100 (2.1)	0.67	0.62 (0.45–0.84)	0.002
Death from cardiovascular causes	37 (0.8)	0.25	65 (1.4)	0.43	0.57 (0.38–0.85)	0.005
Death from any cause	155 (3.3)	1.03	210 (4.5)	1.40	0.73 (0.60–0.90)	0.003
Primary outcome or death	332 (7.1)	2.25	423 (9.0)	2.90	0.78 (0.67–0.90)	<0.001
<b>Participants with CKD at baseline</b>						
<b>(N=1330)</b>		<b>(N=1316)</b>				
Composite renal outcome‡	14 (1.1)	0.33	15 (1.1)	0.36	0.89 (0.42–1.87)	0.76
≥50% reduction in estimated GFR§	10 (0.8)	0.23	11 (0.8)	0.26	0.87 (0.36–2.07)	0.75
Long-term dialysis	6 (0.5)	0.14	10 (0.8)	0.24	0.57 (0.19–1.54)	0.27
Kidney transplantation	0		0			
Incident albuminuria¶	49/526 (9.3)	3.02	59/500 (11.8)	3.90	0.72 (0.48–1.07)	0.11
<b>Participants without CKD at baseline</b>						
<b>(N=3332)</b>		<b>(N=3345)</b>				
≥30% reduction in estimated GFR to <60 ml/min/1.73 m²§	127 (3.8)	1.21	37 (1.1)	0.35	3.49 (2.44–5.10)	<0.001
Incident albuminuria¶	110/1769 (6.2)	2.00	135/1831 (7.4)	2.41	0.81 (0.63–1.04)	0.10

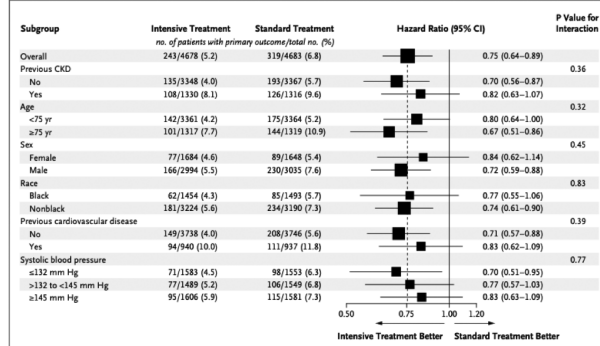


Figure 4. Forest Plot of Primary Outcome According to Subgroups. The dashed vertical line represents the hazard ratio for the overall study population. The box sizes are proportional to the precision of the estimates (with larger boxes indicating a greater degree of precision). The subgroup of no previous chronic kidney disease (CKD) includes some participants with unknown CKD status at baseline. Black race includes Hispanic black and black as part of a multiracial identification.

Outros dados importantes encontrados no SPRINT.

Na página 11, apresenta a análise de subgrupos. Ajuda a precisar ainda mais os resultados

## Analisando meta-análises

Relembrar é viver: meta-análises não são mais consideradas o topo da pirâmide da MBE

Meta-análises são um tipo de pesquisa em que se usam técnicas estatísticas para combinar dados de mais de uma pesquisa e analisá-los em conjunto. Isso tem que ser feito de maneira muito sistemática: o pesquisador deve, antes de começar a pesquisa, definir qual será o objetivo da sua pesquisa e quais artigos vai incluir em sua meta-análise (por exemplo: artigos do tipo randomizados e controlados que estejam indexados na base de dados PubMed no intervalo do ano 2010-2020). Se não for assim, há um grande risco de o pesquisador acabar cometendo “*cherry picking*” e selecionando apenas os artigos que lhe convêm – que tiveram resultados que apontam na mesma direção do ponto que ele quer provar com a pesquisa.

Não é fácil conduzir uma meta-análise. Imagine uma meta-análise pequena de 5 pesquisas que tentaram responder se a aspirina é melhor que o placebo em prevenção primária cardiovascular – ou seja, aspirina reduz a chance de ter infarto em quem nunca teve infarto? Ele abre as pesquisas e vai ler: a primeira delas teve mais pacientes orientais e usaram a dose de 81 mg de aspirina. A segunda pesquisa teve mais pacientes na Europa e usaram a dose de 124 mg, além de um consumo maior de vinho. A terceira pesquisa tinha apenas pacientes com mais de 70 anos, enquanto a quarta tinha uma média de idade de 37. A quinta, por sua vez, usou como desfecho primário a evolução para câncer colorretal, ficando com a evolução para infarto como desfecho secundário. Percebeu as dificuldades que esse pesquisador vai ter em homogeneizar esses dados?

De fato, a heterogeneidade é um dos principais problemas das meta-análises, tanto que existem elementos para quantificar isso, como a estatística Q e o teste  $I^2$ .<sup>34</sup> Quanto maior o  $I^2$ , por exemplo, maior será a heterogeneidade dos estudos escolhidos para aquela análise (são muito diferentes).

Como vimos no Capítulo 3, as meta-análises e as revisões sistemáticas deixaram, recentemente, de ser consideradas o topo da pirâmide das evidências, para servir mais como uma lupa, que pode ajudar ou atrapalhar a analisar os dados da pirâmide.

## Fique muito atento à qualidade dos estudos que entraram na análise e cheque se estudos importantes deixaram de entrar

Em meta-análises, obedecemos a um mantra, que é o de “*garbage in, garbage out*” – lixo entra, lixo sai. Significa que se o pesquisador escolher artigos de maneira aleatória (ou, pior, deliberadamente escolher artigos duvidosos, de má qualidade, sem *peer review* adequado, influenciados politicamente etc.), ele pode estar montando uma meta-análise cujo resultado será lixo.

Se o leitor tiver que fazer a interpretação de uma meta-análise, observe se algum *trial* clássico sobre o tema foi deixado de lado e, caso tenha sido, tente encontrar no texto a razão para isso ter acontecido (pode ser que o pesquisador até justifique ter retirado por razões metodológicas).

## O gráfico forest plot nas meta-análises

Os já explicados *forest plots* estão de volta. Cada linha horizontal no gráfico representa o intervalo de confiança da redução ou aumento da chance de algum desfecho específico em diferentes estudos. Intervalos com maior largura significam pesquisas com menor tamanho amostral e intervalos mais curtos significam, em geral, pesquisas com maior *power* – são também mais confiáveis.

Na última linha do forest plot está o diamante que vai sumarizar os resultados dos estudos combinados. Estudos maiores têm mais peso no desenho desse diamante que estudos menores.

Na linha do centro, a famosa “linha do 1”, temos o ponto em que o intervalo de confiança do *hazard ratio* cruza o 1.0, ou seja, há > 5% de chance de que a média do resultado real do estudo seja de ausência de benefício ou mesmo causar dano (vimos isso no Capítulo 6).

A Figura 20 ilustra um *forest plot* da meta-análise de cinco artigos. Observe que nada se diz sobre a qualidade metodológica ou do *peer review*, ou até mesmo a índole de cada um desses autores.

**Figura 20.**

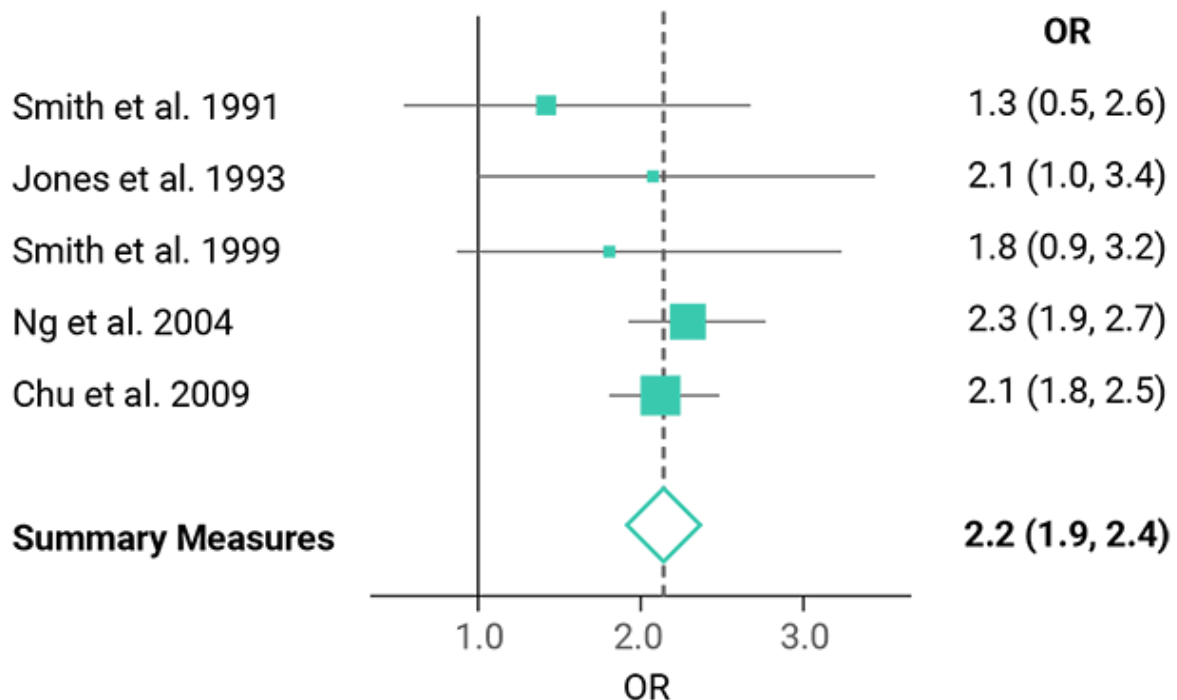


Gráfico do tipo *forest plot* em uma meta-análise para um determinado desfecho não citado na figura.

Observe que os estudos menores (da década de 1990) não conseguiam demonstrar diferença

significativa, pelo fato de que o intervalo de confiança cruzada a linha do 1 (então, havia mais que 5% de probabilidade de que a média do resultado definitivo fosse neutro ou até danoso). Mas as pesquisas de 2004 e 2009 encontraram um intervalo de confiança mais curto (provavelmente eram pesquisas com maior *power* e amostra). Combinando todas, se desenhou o diamante na última linha, que demonstra um benefício da ordem de 2,2 no *odds ratio*.

## Revisões sistemáticas

Também consideradas como “lupas”, as revisões sistemáticas são como um capítulo de livro sobre uma determinada doença ou tratamento, com a diferença que seu principal objetivo não é “ensinar”, mas “informar” sobre as evidências existentes sobre aquele assunto.

Para isso, o autor precisa explicitar de que maneira ele escolherá cada referência bibliográfica dele e também de que modo ele interpretará e sintetizará aqueles resultados em seu texto.

Para escolher suas referências bibliográficas, o autor pode definir critérios para que as pesquisas que ele encontre sejam minimamente sérias, realistas e bem conduzidas – acredite, isso é muito raro na pesquisa médica. Então, provavelmente, o autor buscará seus artigos nas bases PubMed e MEDLINE, por exemplo.

Para interpretar bem os estudos, o autor pode sistematizar a sua análise respondendo, para cada artigo selecionado, as seguintes questões:

1. A alocação dos pacientes foi randomizada?
2. A alocação foi cega?
3. Os grupos eram similares em termos de fatores de risco e prognósticos?
4. Os critérios de elegibilidade foram especificados?
5. Os pesquisadores, os médicos e os pacientes foram cegos?
6. As medidas de variabilidade e as estimativas de ponto foram apresentadas para o desfecho primário?
7. A análise foi do tipo *intention-to-treat*?

Quanto mais respostas negativas a essas perguntas, mais enviesado é o artigo e maior o risco de que ele polua o resultado final.

## Referências bibliográficas

1. Lemiengre MB, van Driel ML, Merenstein D, Liira H, Mäkelä M, De Sutter AI. Antibiotics for acute rhinosinusitis in adults. *Cochrane database Syst Rev*. 2018 Sep;9(9):CD006089.
2. Kahlert J, Gribsholt SB, Gammelager H, Dekkers OM, Luta G. Control of confounding in the analysis phase – an overview for clinicians. *Clin Epidemiol [Internet]*. 2017 Mar 31;9:195-204 [[acesso](#) em: 18 mar. 2021].
3. Patel MR, Mahaffey KW, Garg J, Pan G, Singer DE, Hacke W, Breithardt G, Halperin JL, Hankey GJ, Piccini JP, Becker RC, Nessel CC, Paolini JF, Berkowitz SD, Fox KAA, Califf RM. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *N Engl J Med*. 2011 Sep;365(10):883-91.
4. Giugliano RP, Ruff CT, Braunwald E, Murphy SA, Wiviott SD, Halperin JL, Waldo AL, Ezekowitz MD, Weitz JI, Špinar J, Ruzyllo W, Ruda M, Koretsune Y, Betcher J, Shi M, Grip LT, Patel SP, Patel I, Hanyok JJ, Mercuri M, Antman EM. Edoxaban versus Warfarin in Patients with Atrial Fibrillation. *N Engl J Med [Internet]*. 2013 Nov 19;369(22):2093–104 [[acesso](#) em: 18 mar. 2021].
5. Granger CB, Alexander JH, McMurray JJ V, Lopes RD, Hylek EM, Hanna M, Al-Khalidi HR, Ansell J, Atar D, Avezum A, Bahit MC, Diaz R, Easton JD, Ezekowitz JA, Flaker G, Garcia D, Geraldles M, Gersh BJ, Golitsyn S, Goto S, Hermosillo AG, Hohnloser SH, Horowitz J, Mohan P, Jansky P, Lewis BS, Lopez-Sendon JL, Pais P, Parkhomenko A, Verheugt FWA, Zhu J, Wallentin L. Apixaban versus warfarin in patients with atrial fibrillation. *N Engl J Med*. 2011 Sep;365(11):981-92.
6. Connolly SJ, Ezekowitz MD, Yusuf S, Eikelboom J, Oldgren J, Parekh A, Pogue J, Reilly PA, Themeles E, Varrone J, Wang S, Alings M, Xavier D, Zhu J, Diaz R, Lewis BS, Darius H, Diener H-C, Joyner CD, Wallentin L. Dabigatran versus warfarin in patients with atrial fibrillation. *N Engl J Med*. 2009 Sep;361(12):1139-51.
7. Leung JT, Barnes SL, Lo ST, Leung DY. Non-inferiority trials in cardiology: what clinicians need to know. *Heart [Internet]*. 2020;106(2):99-104 [[acesso](#) em: 18 mar. 2021].
8. Bada C, Carreazo NY, Chalco JP, Huicho L. Inter-observer agreement in interpreting chest X-rays on children with acute lower respiratory tract infections and concurrent wheezing. *Sao Paulo Med J*. 2007 May;125(3):150-4.
9. McCabe JM, Armstrong EJ, Ku I, Kulkarni A, Hoffmayer KS, Bhave PD, Waldo SW, Hsue P, Stein JC, Marcus GM, Kinlay S, Ganz P. Physician accuracy in interpreting potential ST-segment elevation myocardial infarction electrocardiograms. *J Am Heart Assoc*. 2013 Oct;2(5):e000268.
10. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20(1):37-46.
11. Karanicolas PJ, Farrokhyar F, Bhandari M. Practical tips for surgical research: blinding: who, what, when, why, how? *Can J Surg [Internet]*. 2010 Oct;53(5):345-8 [[acesso](#) em: 18 mar. 2021].

12. Gupta U, Verma M. Placebo in clinical trials. *Perspect Clin Res* [Internet]. 2013 Jan;4(1):49-52 [[acesso](#) em: 18 mar. 2021].
13. Aronson JK. Biomarkers and surrogate endpoints. *Br J Clin Pharmacol* [Internet]. 2005 May;59(5):491-4 [[acesso](#) em: 18 mar. 2021].
14. Heerspink HJL, Stefánsson B V, Correa-Rotter R, Chertow GM, Greene T, Hou F-F, Mann JFE, McMurray JJ V, Lindberg M, Rossing P, Sjöström CD, Toto RD, Langkilde A-M, Wheeler DC. Dapagliflozin in Patients with Chronic Kidney Disease. *N Engl J Med*. 2020 Oct;383(15):1436-46.
15. Poudel I, Tejpal C, Rashid H, Jahan N. Major Adverse Cardiovascular Events: An Inevitable Outcome of ST-elevation myocardial infarction? A Literature Review. *Cureus* [Internet]. 2019 Jul 30;11(7):e5280-e5280 [[acesso](#) em: 18 mar. 2021].
16. Mehta SR, Wood DA, Storey RF, Mehran R, Bainey KR, Nguyen H, Meeks B, Di Pasquale G, López-Sendón J, Faxon DP, Mauri L, Rao S V, Feldman L, Steg PG, Avezum Á, Sheth T, Pinilla-Echeverri N, Moreno R, Campo G, Wrigley B, Kedev S, Sutton A, Oliver R, Rodés-Cabau J, Stanković G, Welsh R, Lavi S, Cantor WJ, Wang J, Nakamya J, Bangdiwala SI, Cairns JA. Complete Revascularization with Multivessel PCI for Myocardial Infarction. *N Engl J Med* [Internet]. 2019 Sep 1;381(15):1411-21 [[acesso](#) em: 18 mar. 2021].
17. Vetter TR, Mascha EJ. Defining the Primary Outcomes and Justifying Secondary Outcomes of a Study: Usually, the Fewer, the Better. *Anesth Analg*. 2017 Aug;125(2):678-81.
18. Dahlöf B, Sever PS, Poulter NR, Wedel H, Beevers DG, Caulfield M, Collins R, Kjeldsen SE, Kristinsson A, McInnes GT, Mehlsen J, Nieminen M, O'Brien E, Ostergren J. Prevention of cardiovascular events with an antihypertensive regimen of amlodipine adding perindopril as required versus atenolol adding bendroflumethiazide as required, in the Anglo-Scandinavian Cardiac Outcomes Trial-Blood Pressure Lowering Arm (ASCOT-Lancet (London, England). 2005 Sep;366(9489):895-906.
19. Schulz KF, Altman DG, Moher D, Group C. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *Trials* [Internet]. 2010 Mar 24;11:32 [[acesso](#) em: 18 mar. 2021].
20. Altman DG. Comparability of randomised groups. *J R Stat Soc Ser D (The Stat*. 1985;34(1):125-36.
21. Packer DL, Mark DB, Robb RA, Monahan KH, Bahnson TD, Poole JE, Noseworthy PA, Rosenberg YD, Jeffries N, Mitchell LB, Flaker GC, Pokushalov E, Romanov A, Bunch TJ, Noelker G, Ardashev A, Revishvili A, Wilber DJ, Cappato R, Kuck K-H, Hindricks G, Davies DW, Kowey PR, Naccarelli G V, Reiffel JA, Piccini JP, Silverstein AP, Al-Khalidi HR, Lee KL. Effect of Catheter Ablation vs Antiarrhythmic Drug Therapy on Mortality, Stroke, Bleeding, and Cardiac Arrest Among Patients With Atrial Fibrillation: The CABANA Randomized Clinical Trial. *JAMA*. 2019 Apr;321(13):1261-74.
22. Shah PB. Intention-to-treat and per-protocol analysis. *CMAJ* [Internet]. 2011 Apr 5;183(6):696 [[acesso](#) em: 18 mar. 2021].



23. Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey Smith G. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. *Health Technol Assess*. 2001;5(33):1-56.
24. Bhatt DL, Fox KAA, Hacke W, Berger PB, Black HR, Boden WE, Cacoub P, Cohen EA, Creager MA, Easton JD, Flather MD, Haffner SM, Hamm CW, Hankey GJ, Johnston SC, Mak K-H, Mas J-L, Montalescot G, Pearson TA, Steg PG, Steinhubl SR, Weber MA, Brennan DM, Fabry-Ribaudo L, Booth J, Topol EJ. Clopidogrel and Aspirin versus Aspirin Alone for the Prevention of Atherothrombotic Events. *N Engl J Med* [Internet]. 2006 Apr 20;354(16):1706-17 [[acesso](#) em: 18 mar. 2021].
25. Pfeffer MA, Jarcho JA. The charisma of subgroups and the subgroups of CHARISMA. Vol. 354, *The New England journal of medicine*. United States; 2006. p. 1744-6.
26. A randomized trial of beta-blockade in heart failure. The Cardiac Insufficiency Bisoprolol Study (CIBIS). CIBIS Investigators and Committees. *Circulation*. 1994 Oct;90(4):1765-73.
27. The Cardiac Insufficiency Bisoprolol Study II (CIBIS-II): a randomised trial. *Lancet* (London, England). 1999 Jan;353(9146):9-13.
28. Fox K, Ford I, Steg PG, Tardif J-C, Tendera M, Ferrari R. Ivabradine in stable coronary artery disease without clinical heart failure. *N Engl J Med*. 2014 Sep;371(12):1091-9.
29. Fox K, Ford I, Steg PG, Tendera M, Ferrari R. Ivabradine for patients with stable coronary artery disease and left-ventricular systolic dysfunction (BEAUTIFUL): a randomised, double-blind, placebo-controlled trial. *Lancet* (London, England). 2008 Sep;372(9641):807-16.
30. Swedberg K, Komajda M, Böhm M, Borer JS, Ford I, Dubost-Brama A, Lerebours G, Tavazzi L. Ivabradine and outcomes in chronic heart failure (SHIFT): a randomised placebo-controlled study. *Lancet* (London, England). 2010 Sep;376(9744):875-85.
31. Pocock SJ, Stone GW. The Primary Outcome Fails — What Next? *N Engl J Med* [Internet]. 2016 Aug 31;375(9):861-70 [[acesso](#) em: 18 mar. 2021].
32. Sedgwick P. External and internal validity in clinical trials. *BMJ* [Internet]. 2012;344.
33. Wright JTJ, Williamson JD, Whelton PK, Snyder JK, Sink KM, Rocco M V, Reboussin DM, Rahman M, Oparil S, Lewis CE, Kimmel PL, Johnson KC, Goff DCJ, Fine LJ, Cutler JA, Cushman WC, Cheung AK, Ambrosius WT. A Randomized Trial of Intensive versus Standard Blood-Pressure Control. *N Engl J Med*. 2015 Nov;373(22):2103-16.
34. Berlin JA, Golub RM. Meta-analysis as Evidence: Building a Better Pyramid. *JAMA* [Internet]. 2014 Aug 13;312(6):603-6 [[acesso](#) em: 18 mar. 2021].
35. A co-operative trial in the primary prevention of ischaemic heart disease using clofibrate. Report from the Committee of Principal Investigators. *Br Heart J*. 1978 Oct;40(10):1069-118.
36. Echt DS, Liebson PR, Mitchell LB, Peters RW, Obias-Manno D, Barker AH, Arensberg D, Baker A, Friedman L, Greene HL. Mortality and morbidity in patients receiving encainide,

flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. *N Engl J Med.* 1991 Mar;324(12):781-8.

37. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med.* 2007 Jun;356(24):2457-71.

38. Beecher HK. The powerful placebo. *J Am Med Assoc.* 1955 Dec;159(17):1602-6.

39. Kienle GS, Kiene H. The powerful placebo effect: fact or fiction? *J Clin Epidemiol.* 1997 Dec;50(12):1311-8.

## 8. COMO INTERPRETAR UM TESTE DIAGNÓSTICO?

### Introdução

Alguns leitores podem estar se perguntando: “o que isso tem a ver com o livro?”. É que muitos conceitos usados nos capítulos anteriores servem pra nos ajudar a fazer a análise correta dos resultados de testes diagnósticos. E deixo como *spoiler* um fato: sensibilidade e especificidade, sozinhas, não ajudam, pelo contrário, atrapalham a elaboração do diagnóstico.

Não é culpa da graduação em si, mas culpa do pouco tempo que temos pra aprender tantas coisas. É que quando nos ensinam sobre uma doença, normalmente aprendemos qual a apresentação sintomática clássica e depois quais os exames e os achados desses exames. Uma pequena frase sobre a sensibilidade e a especificidade do teste é escrita, como se fosse para alívio do médico ou futuro médico: “sensibilidade de 90% e especificidade de 90%”. Ele pensa: “ainda bem, esse teste só vai errar em 10% das vezes”.

O livro está errado (ou impreciso) ao escrever isso e você está errado em interpretar assim. Esse conhecimento é crucial para a boa prática médica. Afinal, você não quer ser um desses médicos que pedem bateria de exames sem sentido, quer?

### Alguns conselhos

Você precisa entender que, da mesma forma como em uma metanálise, se os dados que você receber do paciente forem pobres, também pobre será a sua conclusão. Então, o primeiro conselho é que foque na qualidade e sistematização da sua anamnese e do seu exame físico.

O segundo conselho, antes de começar, é entender que tudo que será falado aqui neste capítulo é variável de acordo com uma série de fatores: variabilidade interobservador e intraobservador, a capacidade do médico de evitar vieses cognitivos<sup>1</sup> e da dinamicidade dos processos saúde-doença: o paciente pode responder bem à sua terapia e essa resposta acabar levando a uma outra doença, com novos achados, novos exames e novas condutas (por

exemplo: alguém congesto por insuficiência cardíaca que fica desidratado por ter sido bom respondedor à terapia diurética).

## Variabilidade inter e intraobservador

Variabilidade interobservador é a reprodutibilidade de um resultado de exame feito por diferentes observadores. Quanto maior a variabilidade, maior será a diferença entre os resultados – é o que se diz por aí dos “exames examinador-dependente”.

Variabilidade intraobservador é a repetitividade de um resultado de exame feito pelo mesmo médico em períodos diferentes, desde que o achado (ou a constelação de achados) ainda esteja lá.<sup>2</sup>

A concordância interobservador de alguns sinais do exame físico médico está descrita na tabela 1. A tabela 2 ilustra a concordância interobservador de alguns testes diagnósticos.<sup>3</sup>

**Tabela 1.**

Achado	Estatística k (concordância interobservador)
Cianose	0,36 – 0,7
Desidratação	0,44 – 0,53
Creptos	0,21 – 0,65
Sibilos	0,43 – 0,93
Sopro sistólico	0,19

Concordância interobservador de alguns sinais do exame físico.

**Tabela 2.**

Achado	Estatística k (concordância interobservador)
Infiltrado pulmonar no raio X de tórax	0,38
Presença ou ausência de lesão suspeita na mamografia de <i>screening</i>	0,47
Classificação de lesões coronárias no cateterismo	0,33

Achado	Estatística k (concordância interobservador)
Nódulo tireóideo na ultrassonografia	0,57 – 0,66
Gravidade da regurgitação mitral ao ecocardiograma	0,32 – 0,55

Concordância interobservador de alguns resultados possíveis em exames diagnósticos.

## Vieses cognitivos que podem atrapalhar na elaboração de um diagnóstico

1. Viés de disponibilidade: considerar possibilidades de mais fácil recordação como mais prováveis apesar de uma plausibilidade reduzida.
2. Viés de ancoramento: se prender a alguma informação e negligenciar outros dados e fatos a partir daí.
3. Negligência da taxa base: ignorar a plausibilidade, ou a probabilidade pré-teste de uma doença em um paciente específico ao interpretar achados do exame físico ou resultados de exames.
4. Viés de confirmação: é quando tendemos a procurar e valorizar apenas aquelas informações que vão na mesma direção das nossas expectativas preexistentes.
5. Viés de fechamento prematuro: quando o médico para de pensar em novas possibilidades por já haver um diagnóstico firmado, mesmo que esse diagnóstico tenha sido impreciso.

## Entendendo a sensibilidade e a especificidade de um teste

Se eu realizar um teste com 90% de sensibilidade e 90% de especificidade em uma pessoa aleatória para detectar uma doença relativamente prevalente, digamos 1% da população, e o resultado for positivo, qual a chance de esse resultado ser verdadeiro positivo?

Se você respondeu 90%, está enganado. A resposta correta é que a chance de esse teste ser verdadeiro positivo é de 8,3%. Se a sua resposta foi

errada, provavelmente sua linha de raciocínio foi:

1. Sensibilidade é a capacidade de detectar verdadeiros positivos.
2. Se a sensibilidade é 90%, então a chance de o teste ser verdadeiro positivo é 90%.

Esse raciocínio é muito comum entre estudantes de Medicina e médicos porque é a maneira como são dispostos os testes em nossos livros clássicos. Mas essa maneira de pensar é falha porque desconsidera uma variável importantíssima: a probabilidade pré-teste da doença. Ou você acha que eu mencionei a prevalência da doença à toa? Observe também que eu perguntei “qual a chance de o resultado ser verdadeiro positivo em uma pessoa aleatória?”. Então, em primeiro lugar, vamos fazer uma pequena, mas importantíssima, complementação ao conceito de sensibilidade:

- a. Sensibilidade é a capacidade de um teste de detectar **verdadeiros positivos entre quem tem a doença**.

Veja que diferença isso faz: nós não sabemos se a pessoa aleatória testada é doente ou não, portanto, o valor da sensibilidade, isoladamente, não serve pra ela.

Aproveitando, vamos definir o que é especificidade:

- b. Especificidade é a capacidade de um teste de detectar resultados verdadeiros-negativos entre quem não tem a doença – assim evitando resultados falsos-positivos. (Observe a figura 1 para compreender melhor).

**Figura 1.**

	A doença existe	A doença não existe
TESTE POSITIVO	VP - a	FP - b
TESTE NEGATIVO	FN - c	VN - d

Sensibilidade: positivo em quem tem a doença  
total de pessoas com a doença

$$\text{Sensibilidade} = \frac{a}{a + c}$$

Especificidade: negativo em quem não tem a doença  
total de pessoas sem a doença

$$\text{Especificidade} = \frac{d}{b + d}$$

Sensibilidade é a capacidade de um teste ser verdadeiro entre os que possuem a doença. Especificidade é a capacidade de um teste ser falso entre os que não possuem. Logo, um teste pouco sensível dará muitos resultados falsos-negativos e um teste pouco específico dará muitos resultados falsos-positivos.

Também é importante que você entenda o que acontece com os testes quando eles são pouco sensíveis ou pouco específicos:

- c. Um teste pouco sensível terá muitos resultados falsos-negativos.
- d. Um teste pouco específico terá muitos resultados falsos-positivos.

Um exemplo que pode ajudar é o da angiotomografia contrastada para o diagnóstico de tromboembolismo pulmonar, um exame que possui 94% de sensibilidade e 98% de especificidade.<sup>4</sup> Imaginando 100 pessoas doentes e 100 pessoas sem tromboembolismo pulmonar, a figura 2 demonstra como ficaria esse gráfico.

**Figura 2.**

	Pessoas com TEP	Pessoas sem TEP
TC POSITIVA	94	2
TC NEGATIVA	6	98

Sensibilidade: positivo em quem tem a doença  
total de pessoas com a doença

$$\text{Sensibilidade} = \frac{94}{94 + 6} = \frac{94}{100} = 94\%$$

Especificidade: negativo em quem não tem a doença  
total de pessoas sem a doença

$$\text{Especificidade} = \frac{98}{98 + 2} = \frac{98}{100} = 98\%$$

Sensibilidade e especificidade da angiotomografia contrastada para diagnóstico de tromboembolismo pulmonar.

## Para certos tipos de exames, quando a sensibilidade aumenta, a especificidade diminui

O resultado de uma angiotomografia é binário: os resultados possíveis deste exame no tocante à embolia pulmonar são “positivo” e “negativo”. Mas existem exames cujo resultado é descrito em forma de variáveis contínuas – como é a troponina, por exemplo. Veja bem, em exames com variáveis contínuas, alguém tem que definir um valor de limiar, concorda? É aquele valor acima do qual os pacientes serão considerados positivos.

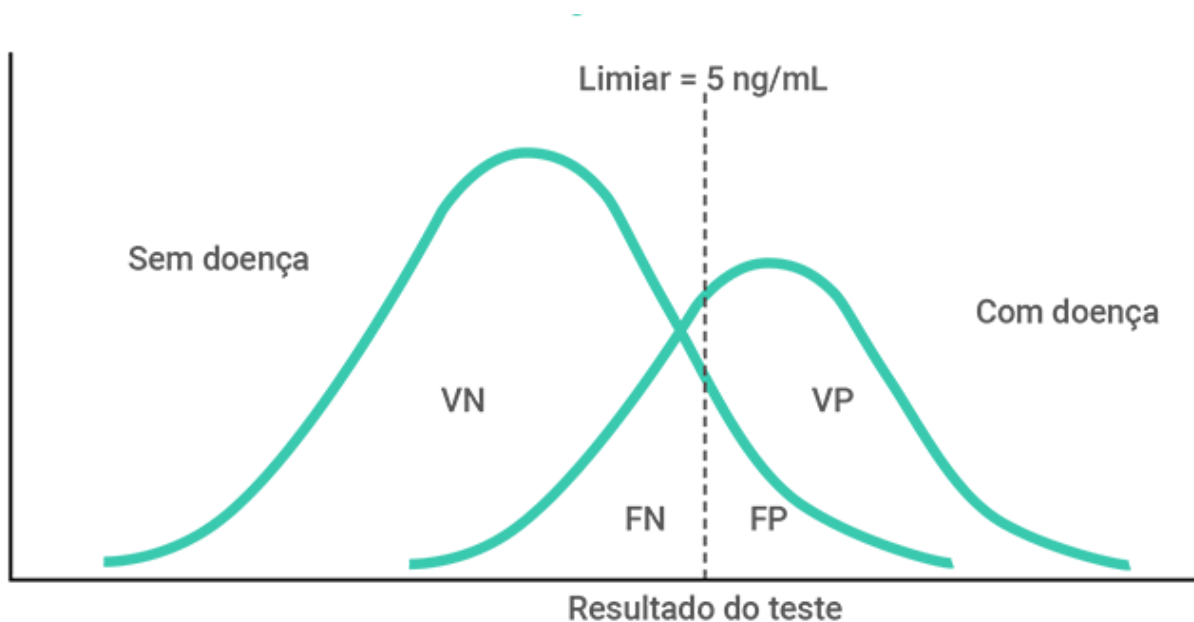
Então sigamos com o exemplo da troponina e imaginemos, para validação de um novo kit a ser lançado no mercado, os pesquisadores querem saber qual será esse valor de limiar. Se ele “escolher” esse valor como 5 ng/mL, significando que todos com valor acima disso serão considerados “positivos”, terá x% de pessoas com o diagnóstico de infarto. Contudo, se ele escolher 14 ng/mL, haverá muito menos pessoas infartando em hospitais, porque só valores acima de 14 é que serão positivos. Também é de se perceber que, uma vez atingindo esse valor, a especificidade (a qualidade de ver falsos positivos) é maior.

O que você acabou de ler é um fato: a quantidade de diagnósticos de várias doenças depende de um valor a ser definido pelos pesquisadores. Acontece assim, por exemplo, com o limiar do TSH (tireoide), da HbA1c



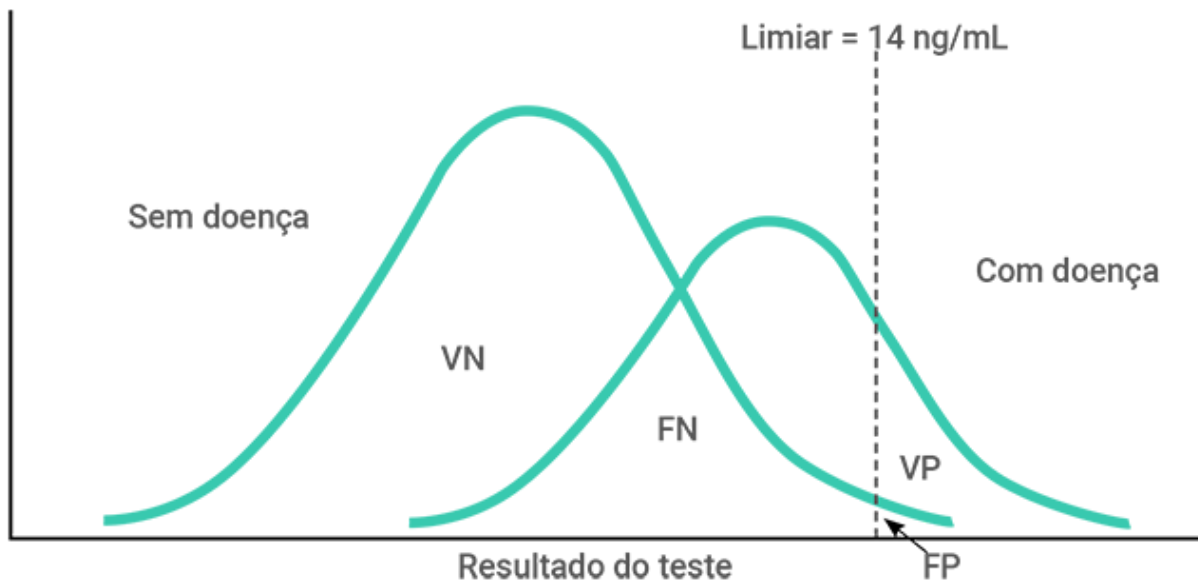
(diabetes), do d-dímero (emergências) e qualquer exame cujo resultado seja uma variável contínua (números) ao invés de binárias (“sim” e “não”). A sensibilidade e a especificidade dos exames de variáveis contínuas são interdependentes e inversamente correlatas. As figuras 3 e 4 ilustram isso.

**Figura 3.**



Exames com variáveis contínuas têm suas variáveis “verdadeiros-positivos” e “falsos-positivos” interdependentes. Isso significa que, ao aumentar ou reduzir o valor de referência do teste, esses valores serão modificados, modificando assim a sensibilidade e a especificidade do teste. Observe na figura 4 o que ocorre quando o limiar é modificado para 14 ng/mL.

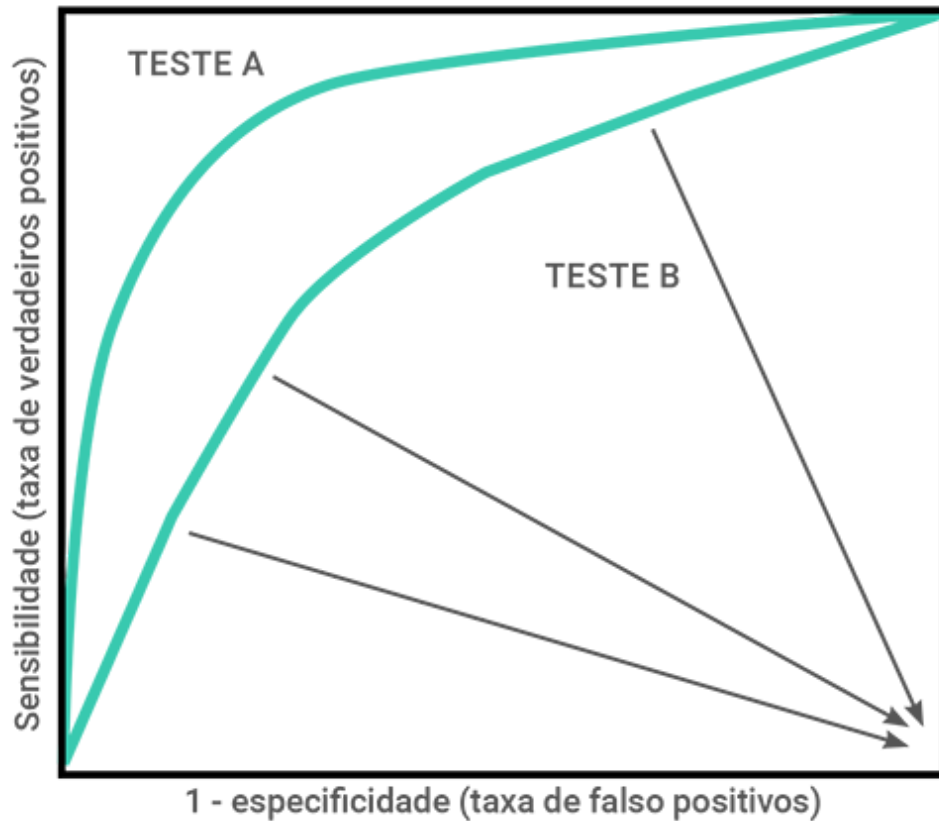
**Figura 4.**



Modificando o valor de referência do teste para 14, o que observamos é que haverá muito menos falsos-positivos (aumentando sobremaneira a especificidade), mas, como consequência, haverá muito mais falsos-negativos, o que diminui sua sensibilidade. A conclusão é que, em exames de variáveis contínuas, a sensibilidade e a especificidade são interdependentes e inversamente correlatas.

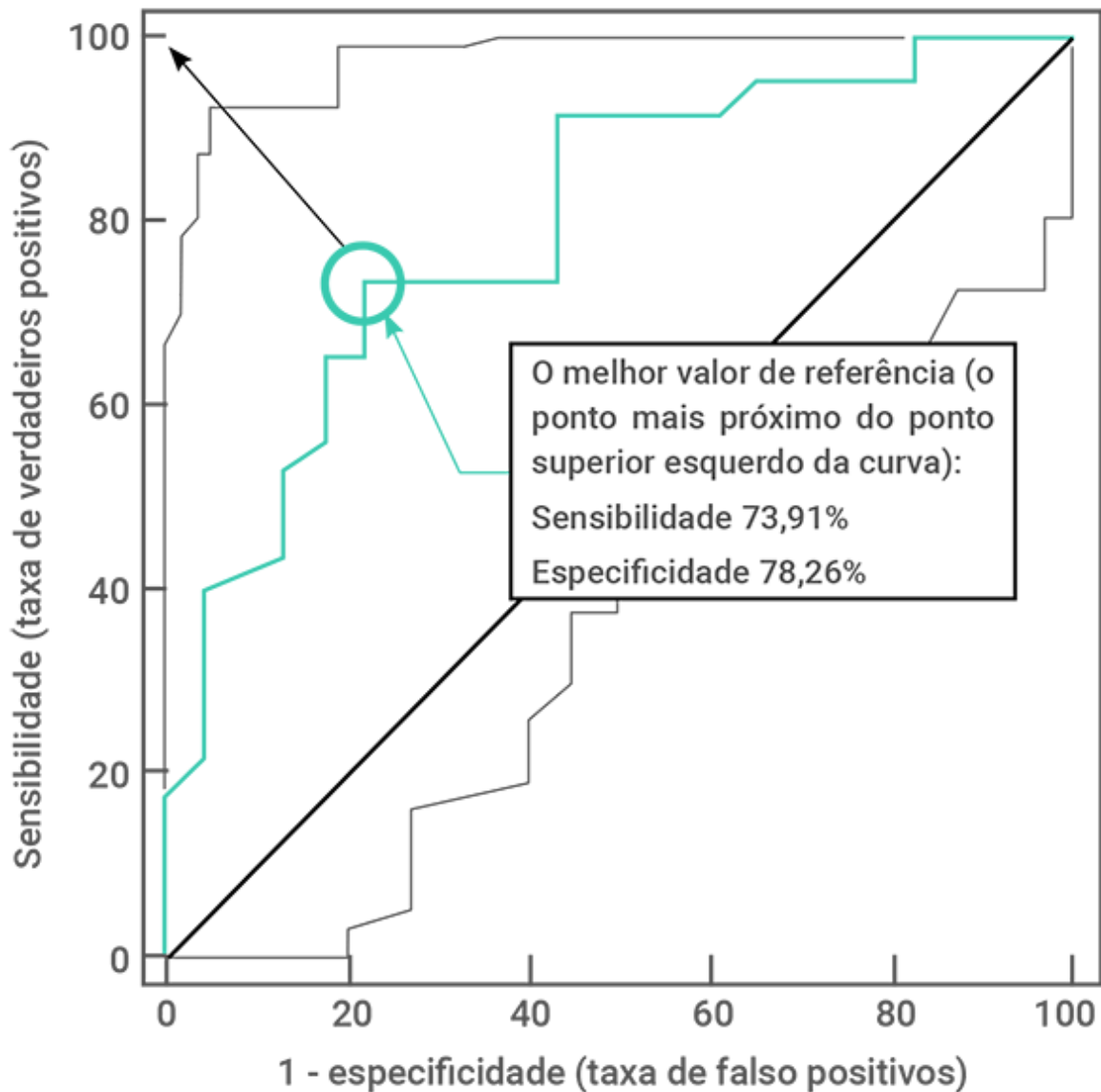
A sorte é que essa “escolha” de valores não é arbitrária. Para escolher o valor de referência de um teste, os pesquisadores lançam mão da **curva ROC** (*receiver operating characteristics*). É um gráfico em que se desenha a sensibilidade de um teste comparada à taxa de falsos positivos para cada valor de referência passível de ser encontrado. Por exemplo, se escolhermos o valor de referência 5 ng/mL, a sensibilidade será 96% e a especificidade 47% – então o pesquisador coloca esses dados no computador. Mas se o valor de referência for 6 ng/mL, então a sensibilidade será de 94% e a especificidade será de 50%. E assim vai. Desse modo, vários dados podem ser calculados, sendo os mais importantes a área abaixo da curva (AUC, em inglês) e o próprio valor de referência ideal para o teste. Quanto maior a AUC, maior a acurácia do teste. Observe a figura 5 para entender. E o valor de referência? Bem, o valor de referência ideal será o valor mais próximo do ponto superior esquerdo da curva – aquele em que o teste teria a maior sensibilidade e especificidade combinadas – isso está ilustrado na figura 6.

**Figura 5.**



Aqui se comparam áreas sob a curva de diferentes exames (teste A e teste B). A área abaixo da curva (AUC) do teste B foi desenhada pelas setas. Agora, imagine que eu também desenhei setas do teste A até o ponto inferior direito da curva, quais setas seriam maiores? As do teste A, sim? Então, o teste A é mais acurado porque tem uma AUC maior.

**Figura 6.**



O melhor valor de referência de um teste com variáveis contínuas é o ponto em que a curva ROC mais se aproxima do ponto superior esquerdo.

## A sensibilidade e a especificidade, quando apresentadas sozinhas, podem ser uma casca de banana

Como o médico, usuário final do produto “exame”, não sabe quem é doente e quem não é – aliás, foi para ter essa resposta que ele pediu o exame – os valores de sensibilidade e especificidade, sozinhos, são de pouca ajuda.

Voltando aos conceitos a e b vistos poucos parágrafos atrás, temos que:

- a. Sensibilidade é a capacidade de um teste de detectar **verdadeiros-positivos entre quem tem a doença**.
- b. Especificidade é a capacidade de um teste de detectar resultados **verdadeiros-negativos entre quem não tem a doença** – assim evitando resultados falsos-positivos.

Para resolver esse problema, precisamos nos converter ao raciocínio de Thomas Bayes.

## Precisamos nos converter ao raciocínio do reverendo Thomas Bayes

Já tivemos um capítulo de raciocínio bayesiano (capítulo 5). Lá, eu expliquei que esse raciocínio é a aplicação de uma fórmula para atualizar uma evidência, um dado ou uma hipótese sob a luz das novas evidências.<sup>5</sup> No caso dos testes diagnósticos, a nossa hipótese é: esta pessoa tem a doença que eu estou pesquisando? Ao passo que a nova evidência será o teste diagnóstico solicitado.

Alguém tem uma hipótese, solicita o teste e o seu resultado atualiza a hipótese – é basicamente assim que funciona.

Se a hipótese em questão é “esta pessoa tem a doença?”, nós precisamos estimar a chance de ele ter mesmo a doença. Isso se faz de duas maneiras e cada uma usa nomenclaturas diferentes:

1. Usando como ponto de partida a prevalência da doença naquela população específica a qual a pessoa faz parte. Nomenclatura usada quando se calcula a probabilidade pós-teste baseado na prevalência da doença: valor preditivo positivo.
2. Usando como ponto de partida a probabilidade clínica da doença naquela pessoa, baseado nos sinais e nos sintomas clínicos referidos e encontrados. Nomenclatura usada quando se calcula a probabilidade pós-teste baseado na probabilidade clínica: razão de verossimilhança (*likelihood ratio*).

Como se faz isso, então? Vamos lá.

### Valores preditivos positivo e negativo

**Valor preditivo positivo** é a proporção de pessoas com um teste positivo que verdadeiramente possuem a doença. Observe que isso é bem diferente da

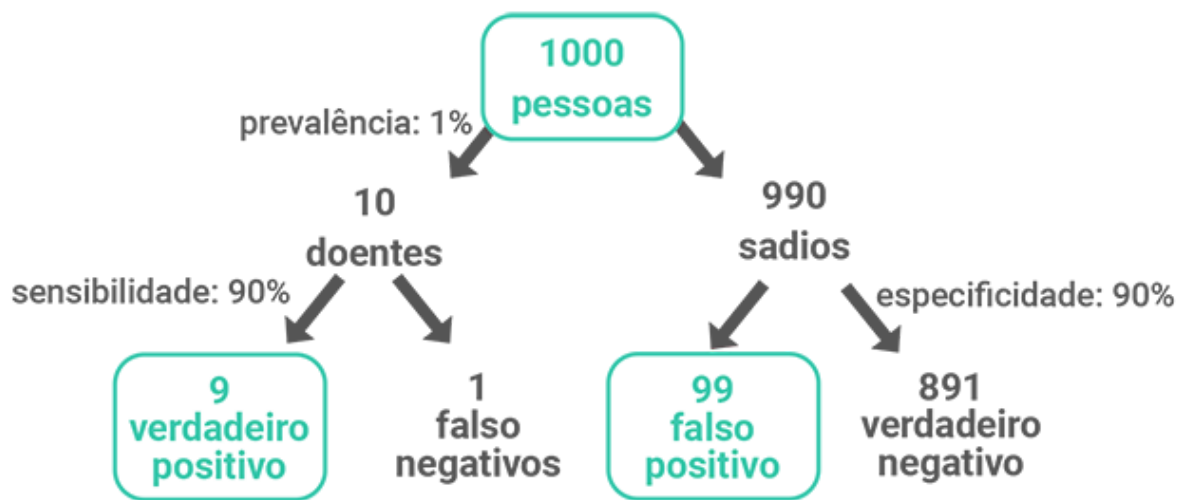
definição de sensibilidade.

**Valor preditivo negativo** é a proporção de pessoas com um teste negativo que verdadeiramente não possuem a doença.

Lembra do tal teste com 90% de sensibilidade, 90% de especificidade e 1% de prevalência na população? Pois bem, existe um teste que possui exatamente esses números, que é a mamografia como rastreamento do câncer de mama. Imagine que você é uma pessoa que faz o teste e esse teste é positivo, qual a proporção de pessoas como você que tem um teste verdadeiramente positivo? Pra isso, usaremos o valor preditivo positivo, que será desenhado na figura 7. Perceba que o gráfico da figura 7 é exatamente o mesmo gráfico usado nas figuras do capítulo 5.

Antes de passar adiante, reflita sobre a importância do que você aprendeu: um teste com 90% de especificidade pode dar tantos resultados falsos-positivos como 90% ou mais.

**Figura 7.**



$$\text{Probabilidade de ser verdadeiro positivo} = \frac{\text{Verdadeiro-positivo}}{\text{Todos os positivos}}$$

$$\text{Probabilidade de ser verdadeiro positivo} = \frac{9}{9 + 99}$$

$$\text{Probabilidade de ser verdadeiro positivo} = 8,3\%$$

O valor preditivo positivo é a proporção de verdadeiros positivos entre todos os positivos. Nessa situação, por haver uma prevalência de “apenas” 1% na população, os falsos-positivos serão muito mais prevalentes que os verdadeiros-positivos. Assim, a proporção de exames verdadeiros-positivos é baixa: apenas 8,3%. Apesar da especificidade de 90%.

## Razão de verossimilhança

Razões de verossimilhança são independentes da prevalência da doença na população porque se baseiam nas características clínicas da pessoa que está sendo atendida.

**Razão de verossimilhança positiva (RV+)** é a chance de um teste ser verdadeiro-positivo ao invés de falso-positivo. Quanto mais alto o número, melhor. Valores de RV+ acima de 10 causam mudanças importantes na probabilidade pós-teste. Se calcula dividindo os verdadeiros-positivos (isto é, sensibilidade) pelos falsos-positivos (1 – especificidade). Quanto maior a especificidade, menor o denominador e maior será o resultado. Exames com alta RV+ e boa especificidade são bons confirmadores.

**Razão de verossimilhança negativa (RV-)** é a chance de um teste ser falso negativo ao invés de verdadeiro negativo. Quanto mais baixo o número, melhor. Valores de RV- abaixo de 0,1 causam mudanças importantes na probabilidade pós-teste. Se calcula dividindo os falsos-negativos (isto é, 1 – sensibilidade) pelos verdadeiros-negativos (especificidade). Quanto maior a sensibilidade, menor o numerador e menor será o resultado. Exames com baixa RV- e boa sensibilidade são bons para excluir diagnósticos (Figura 8).

**Figura 8.**

$$RV+ = \frac{\%VP}{\%FP} = \frac{\text{Sensibilidade}}{1 - \text{especificidade}}$$

$$RV- = \frac{\%FN}{\%VN} = \frac{1 - \text{sensibilidade}}{\text{Especificidade}}$$

Cálculo das razões de verossimilhança positivas e negativas. Quanto maior a RV+ (especialmente se > 10), mais poder confirmatório este exame possui caso o exame seja positivo. Quanto menor a RV- (especialmente se < 0,1), mais poder de excluir um diagnóstico este exame possui caso o exame seja negativo.

Perceba que temos o mesmo conceito do valor preditivo, mas substituímos “proporção” (quando falávamos de valor preditivo) por “chance” (agora que falamos de razão de verossimilhança). Uma razão de verossimilhança positiva de 5, por exemplo, indica que, caso o exame seja positivo, há uma chance 5 vezes maior de que o paciente tenha a doença que está buscando.

Nem todas as doenças possuem razões de verossimilhança descritas para os variados sintomas da clínica médica. Quando é assim, o médico precisará usar da subjetividade e da sua experiência para elaborar a probabilidade pré-

teste da doença. Uma boa maneira de pesquisar e estudar isso é digitar no *Google* o sintoma seguido das palavras “*likelihood ratio*”. Não preciso dizer que você deve, sempre, procurar em bases confiáveis, como MEDLINE/PubMed.

Você também precisa saber que, usando esse método de raciocínio clínico (o método Bayesiano), suas hipóteses serão sempre atualizadas depois de cada resultado de exame ou cada novo sinal/sintoma apresentado/referido pelo paciente.

## Um modelo (adaptado) de raciocínio clínico diagnóstico usando o método bayesiano

Tenhamos como exemplo um caso que ilustra bem o esse método: é o caso de um paciente com **dor torácica que irradia para ambos os braços**.

### Passo 1 – Identifique o problema (queixa principal)

Após ouvir às queixas do paciente, pergunte-se: “qual a probabilidade pré-teste de que um paciente com dor torácica esteja passando por um evento de síndrome coronária aguda?”. Em um estudo clássico, se demonstrou que 29% dos pacientes que se apresentaram com dor torácica acabaram tendo exames positivos para isquemia em algum momento da internação.<sup>6</sup> Vamos usar 25%.

### Passo 2 – Elabore e organize as hipóteses diagnósticas em principais, “não posso deixar passar” e alternativas

Para as hipóteses principais, escolha testes com bom poder confirmatório: alta especificidade e, por conseguinte, alta RV+. Para as hipóteses do tipo “não posso deixar passar” – aquelas cuja gravidade preocupa – use testes com bom poder de exclusão: muito sensíveis e com muito baixa RV-.<sup>7</sup>

Em um paciente com dor torácica típica, a principal hipótese diagnóstica é síndrome coronária aguda. Mas o médico não pode deixar passar: tromboembolismo pulmonar, pneumotórax, entre outras. Além disso, as



hipóteses alternativas ficam entre pericardite, espasmo esofágico e até dor muscular.

### Passo 3 – Teste as suas hipóteses

Aqui você poderia usar o teorema de Bayes (capítulo 5). Mas duas maneiras mais rápidas que serão ensinadas.

A primeira é usar a fórmula simplificada: Chance pós-teste = chance pré-teste x razão de verossimilhança. A segunda direi daqui a pouco. Por enquanto, apenas observe que estamos falando de “chance” e não de “probabilidade” (já aprendemos no capítulo 6 que são diferentes).

Voltando à resolução do nosso caso, 25% é a probabilidade de que uma pessoa qualquer que dê entrada em um PA por dor torácica tenha isquemia. Mas, obviamente, alguém com dor torácica mais típica, como é o caso do nosso paciente hipotético, possui uma chance ainda maior. A razão de verossimilhança desse achado está descrita na literatura como 9.7 (IC95%: 4,6 – 20).<sup>8</sup> Observe, na figura 9, como lidaremos com esses dados.

**Figura 9.**

**Passo 1:** transforme a probabilidade pré-teste em “chance”:  $0,25 / (1 - 0,25) = 0,33$

**Passo 2:** use a fórmula “chance pós-teste = chance pré x RV+”  
Chance pós-teste =  $0,33 \times 9,7 = 3,23$

**Passo 3:** converta a chance pós-teste de novo em probabilidade:  $3,23 / (1 + 3,23) = 76\%$

Passo a passo da transformação de uma probabilidade pré-teste em chance pré-teste, e da aplicação do teorema simplificado de Bayes na avaliação de um paciente com dor torácica (probabilidade 25%) que irradia para ambos os membros (razão de verossimilhança positiva de 9,7).

### Passo 4 – Reorganize as suas hipóteses diagnósticas baseado nos novos dados

Como você viu na figura 9, a probabilidade de que um paciente com dor torácica que irradia para ambos os membros esteja sofrendo isquemia coronária aguda é de 76%. Uma probabilidade tão alta e, ainda por cima, acompanhada de tanta gravidade (este é um caso em que uma hipótese é, ao mesmo tempo, dos tipos “principal” e “não posso deixar passar”) deve ser imediatamente testada.

## Passo 5 – Teste as novas hipóteses sob a luz das novas evidências

É por isso que deve se realizar um eletrocardiograma em até 10 minutos da entrada do paciente com dor torácica no hospital. Antes de avaliar o ECG, lembre-se que a nova probabilidade pré-teste é de 76% (a antiga probabilidade pós-teste).

A razão de verossimilhança negativa do ECG é de 0,74.<sup>9</sup> Observe na figura 10 o que acontece quando calculamos a probabilidade pós-teste deste paciente cujo ECG foi normal.

**Figura 10.**

**Passo 1:** transforme a probabilidade pré-teste em “chance”:  $0,76 / (1 - 0,76) = 3,16$

**Passo 2:** use a fórmula “chance pós-teste = chance pré x RV-”  
Chance pós-teste =  $3,16 \times 0,74 = 2,34$

**Passo 3:** converta a chance pós-teste de novo em probabilidade:  $2,34 / (1 + 2,34) = 70\%$

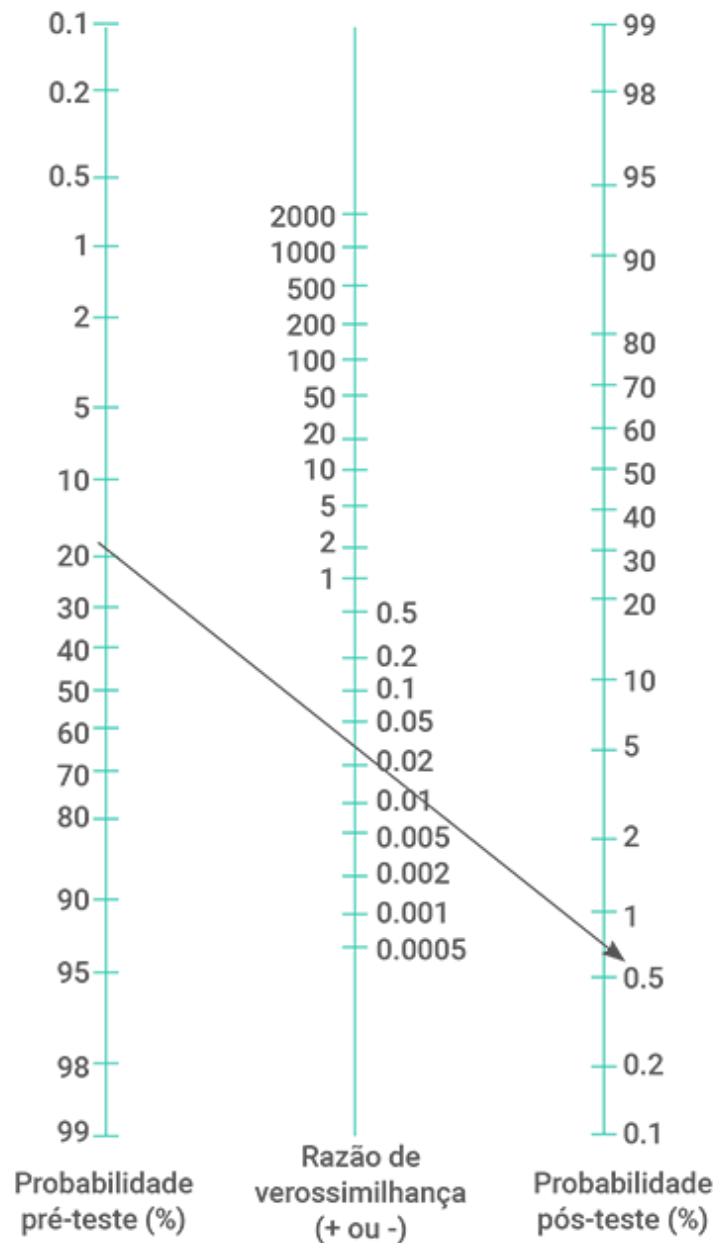
Passo a passo da transformação de uma probabilidade pré-teste em chance pré-teste, e da aplicação do teorema simplificado de Bayes na avaliação de um paciente com dor torácica que irradia para ambos os membros (probabilidade pré-teste de 76%) e com um ECG negativo para isquemia (razão de verossimilhança negativa de 0,74).

Aplicando o teorema de Bayes, assumimos haver uma probabilidade de 70% de que o paciente esteja tendo isquemia aguda, mesmo com o ECG negativo. Como essa hipótese está no time do “não posso deixar passar”, e continua com uma probabilidade muito alta, o médico deve correr contra o tempo para encontrar outras formas de dar esse diagnóstico.

## O nomograma de Fagan

Proposto em 1975, o nomograma pode ajudar porque contém, em um lado, a probabilidade pré-teste, no centro a razão de verossimilhança, e do outro lado a probabilidade pós-teste.<sup>10</sup> Você pode visualizá-lo na figura 11 e usar em seu próximo plantão.

**Figura 11.**



Nomograma de Fagan. Uma maneira fácil de aplicar o teorema de Bayes no raciocínio clínico.

## O modelo Bayesiano de atendimento

O modelo bayesiano de atendimento se fundamenta nos seguintes pressupostos:

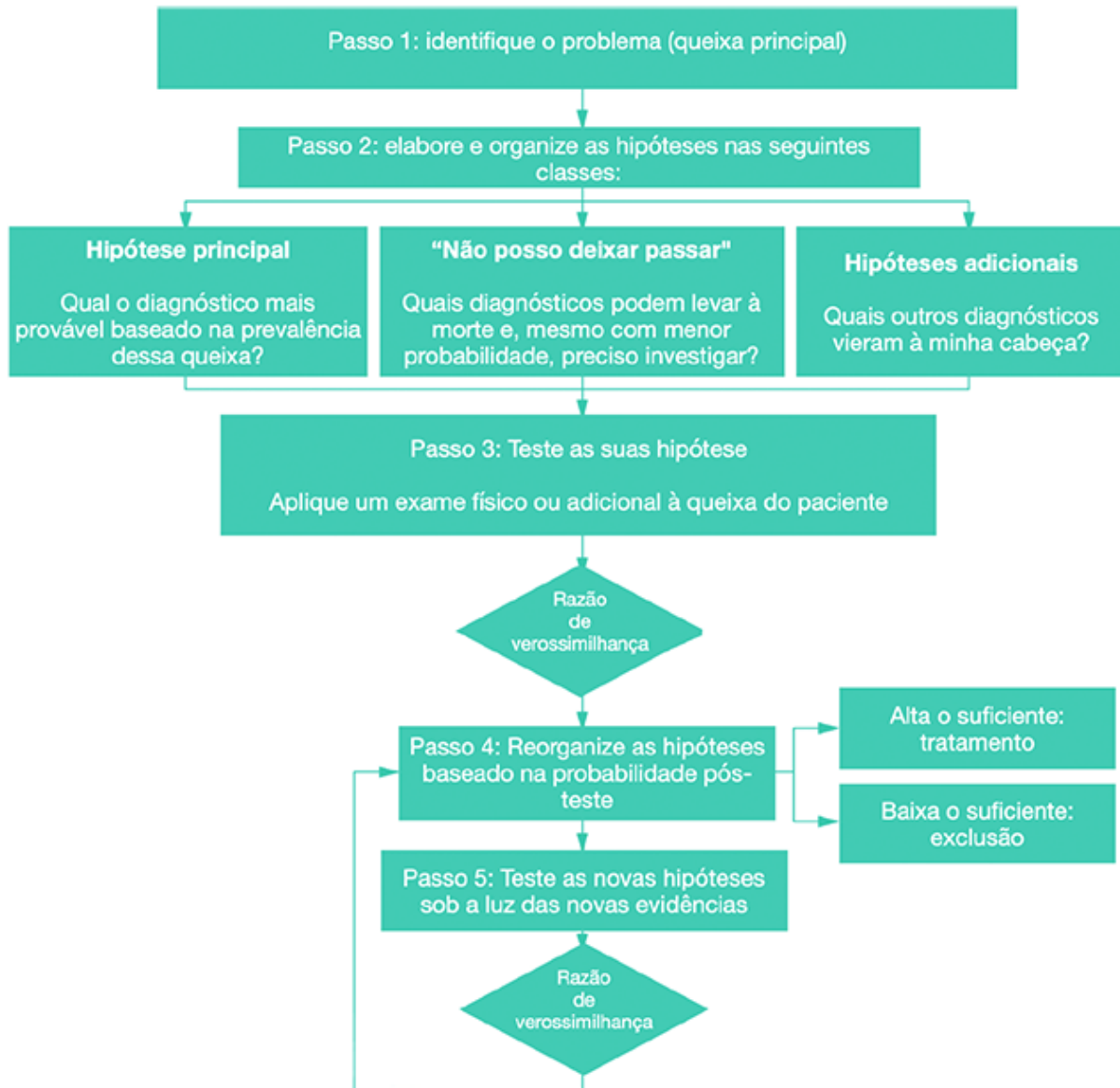
1. A sensibilidade e a especificidade, sozinhas, não são suficientes para a interpretação de um teste porque partem de um dado que o

médico que solicitou o exame não possui: o paciente é doente ou não?

2. Para interpretar corretamente o resultado de um exame, o médico precisa saber a probabilidade pré-teste que começa com a prevalência da doença, mas vai sendo atualizada no decorrer da consulta e dos novos exames.
3. Com probabilidades pós-teste suficientemente altas, o médico está autorizado a realizar o tratamento. E quando suficientemente baixas, o médico está autorizado a considerar aquela hipótese descartada.

Observe, na figura 12, o nosso modelo sugerido de atendimento de acordo com o raciocínio bayesiano.

**Figura 12.**



Modelo bayesiano de atendimento.

## Referências bibliográficas

1. Klein JG. Five pitfalls in decisions about diagnosis and prescribing. *BMJ* [Internet]. 2005 Apr 2;330(7494):781-3 [acesso em 31 mar. 2021].
2. Popović ZB, Thomas JD. Assessing observer variability: a user's guide. *Cardiovasc Diagn Ther* [Internet]. 2017 Jun;7(3):317-24 [acesso em 31 mar. 2021].
3. McGee S. Evidence-Based Physical Diagnosis E-Book [Internet]. Elsevier Health Sciences; 2016 [acesso em 31 mar. 2021].
4. Patel P, Patel P, Bhatt M, Braun C, Begum H, Wiercioch W, et al. Systematic review and meta-analysis of test accuracy for the diagnosis of suspected pulmonary embolism. *Blood*

Adv [Internet]. 2020 Sep 11;4(18):4296-311 [[acesso](#) em 31 mar. 2021].

5. Bittl JA, He Y. Bayesian Analysis: A Practical Approach to Interpret Clinical Trials and Create Clinical Practice Guidelines. *Circ Cardiovasc Qual Outcomes*. 2017 Aug;10(8).
6. Pozen MW, D'Agostino RB, Selker HP, Sytkowski PA, Hood WBJ. A predictive instrument to improve coronary-care-unit admission practices in acute ischemic heart disease. A prospective multicenter clinical trial. *N Engl J Med*. 1984 May;310(20):1273-8.
7. Stern SDC, Cifu AS, Altkorn D. Symptom to Diagnosis: An Evidence Based Guide, Second Edition [Internet]. McGraw-Hill Education; 2009. (LANGE Clinical Medicine) [[acesso](#) em 31 mar. 2021].
8. Berger JP, Buclin T, Haller E, Van Melle G, Yersin B. Right arm involvement and pain extension can help to differentiate coronary diseases from chest pain of other origin: a prospective emergency ward study of 278 consecutive patients admitted for chest pain. *J Intern Med*. 1990 Mar;227(3):165-72.
9. Zehtabchi S. Acute Coronary Syndrome – Diagnostics and Likelihood Ratios, Explained. 2017.
10. Fagan TJ. Letter: Nomogram for Bayes theorem. *N Engl J Med*. 1975 Jul;293(5):257.

## SEÇÃO 2

---

# Especialidades em Evidência

---

### **Autores**

Arn Migowski

Bruno Halpern

Bruno Robalinho Cavalcanti Barbosa

Cristina Van Blarcum de Graaff Mello

Edmond Raymond Le Campion

Felipe Nogueira

Guilherme Augusto Teodoro Athayde

Jonatas Brito de Alencar Neto

Luiza Caires

Márcio Bezerra Gadelha Lopes

Otavio Clark

Paulo Ricardo Criado

Thiago Bosco Mendes

### **O que você irá ver neste capítulo**

- › Cardiologia baseada em evidências
- › Endocrinologia baseada em evidências
  - › Oncologia baseada em evidências
- › Rastreamento de câncer baseado em evidências
  - › Cirurgia baseada em evidências
  - › Dermatologia baseada em evidências
  - › Ortopedia baseada em evidências
- › Jornalismo de saúde baseado em evidências

# 9. CARDIOLOGIA BASEADA EM EVIDÊNCIAS

## Introdução

A Cardiologia é uma das especialidades médicas que protagonizam o *ranking* de maior número de estudos científicos publicados, além grandes avanços científicos nos últimos anos. Contudo, também é lar de grandes exemplos de como o pensamento mecanicista é capaz de trazer certezas falaciosas. Conhecer os estudos que fomentam as evidências científicas atuais é fundamental, a fim de não cair em armadilhas cognitivas, reconhecer os pontos fortes de cada terapia e, principalmente, de trazer mais respaldo clínico e probabilístico às suas decisões.

## Tratamento das arritmias baseado em evidências

O termo “arritmia” referindo-se aos distúrbios do ritmo cardíaco foi introduzido na literatura médica em 1872, por Heidenhain, porém somente após a capacidade de registrar os sinais elétricos produzidos pelo coração, inicialmente por Waller, em 1887, e após por Einthoven, com o uso do galvanômetro de cordas, o estudo do ritmo cardíaco ganhou notoriedade.<sup>1</sup> Desde então, a evolução da terapêutica vem se desenvolvendo com rapidez, permitindo a melhora na qualidade de vida e na sobrevida dos pacientes com arritmias.

De maneira mecanicista e intuitiva, é fácil pensar que, se há um ritmo cardíaco normal, quaisquer alterações são prejudiciais e devem ser tratadas. Nas décadas de 1970 e 80, após a criação das unidades coronarianas, que reduziram significativamente a mortalidade intra-hospitalar dos pacientes pós-infarto agudo do miocárdio,<sup>2</sup> percebeu-se que, dentre os sobreviventes, muitos evoluíam com arritmias ventriculares frequentes e que estas, em vários estudos, mostravam-se como preditores de morte nesse cenário.<sup>3</sup>

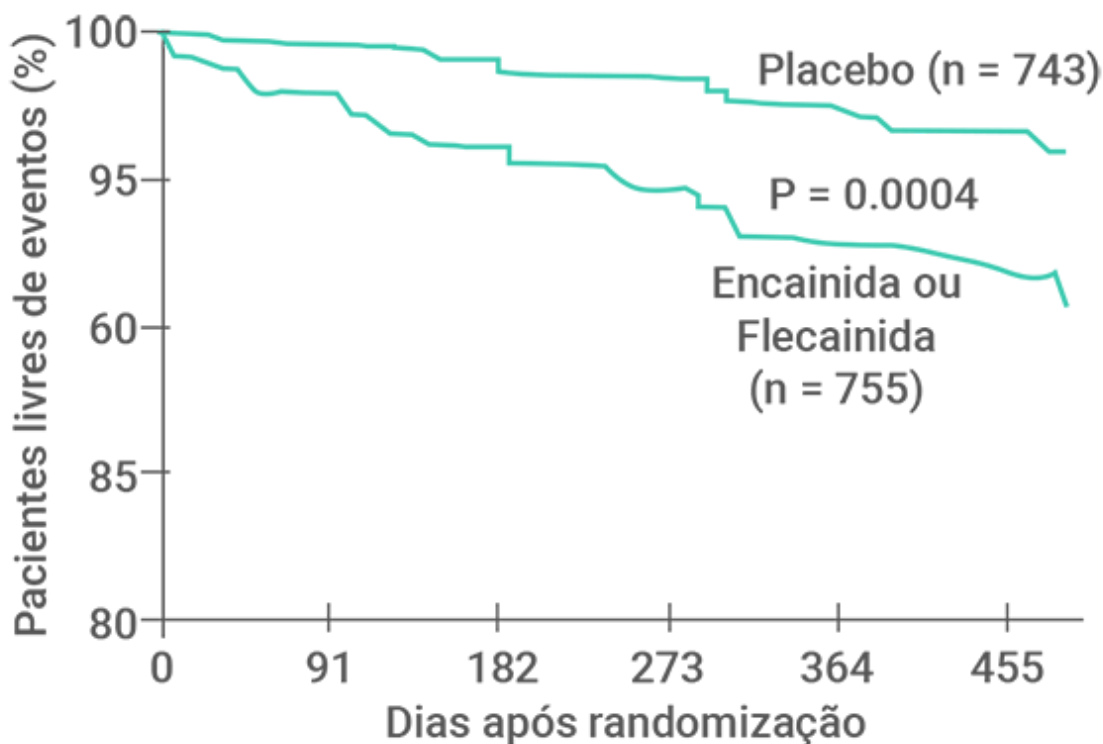
Um grupo de investigadores americanos, então, formulou a hipótese que, se as arritmias ventriculares estavam associadas a maior risco de morte em pacientes com miocardiopatia isquêmica, extingui-las reduziria tal risco, aumentando a sobrevida desses indivíduos. Em 1986, financiados pelo



National Heart, Lung, and Blood Institute, foi publicado o estudo CAPS (*The Cardiac Arrhythmia Pilot Study*),<sup>4</sup> em que se confirmava que flecainida, encainida e moricizina, drogas antiarrítmicas da classe IC (inibidores dos canais de sódio), eram capazes de praticamente suprimir as arritmias ventriculares nesses pacientes. Tudo parecia caminhar bem, porém, no percurso até as certezas geradas pelo pensamento mecanicista existe o rigor científico, capaz de testar adequadamente as hipóteses.

O estudo CAST (*The Cardiac Arrhythmia Suppression Trial*)<sup>5</sup> incluiu 1.498 pacientes, randomizados entre os grupos flecainida/encainida vs. placebo. Porém, após 10 meses de seguimento, o estudo foi interrompido por excesso de mortes (em especial por arritmias) no grupo que recebeu antiarrítmicos (Figura 1), tornando-se uma das maiores reviravoltas científicas do nosso tempo e limitando o uso de antiarrítmicos da classe IC em pacientes com cardiopatia estrutural.

**Figura 1.**



Probabilidade de sobrevivência a morte ou parada cardíaca devido a arritmia em 1498 pacientes recebendo Encainida / Flecainida ou Placebo. Adaptado de: Echt DS, Liebson PR, Mitchell LB, Peters RW, Obias-Manno D, Barker AH, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. *The Cardiac Arrhythmia Suppression Trial*. *N Engl J Med*. 1991 Mar 21;324(12):781-8.

Além disso, apesar de alguns estudos mostrarem benefícios prognósticos da amiodarona (classe III) nesse cenário,<sup>6</sup> tais achados não são consistentes,<sup>7</sup> com baixa validade externa, tornando a terapia medicamentosa baseada no tratamento da cardiopatia de base e ao uso de betabloqueadores e limitando o uso de antiarrítmicos a cenários específicos.

Ainda com o alto risco de morte súbita (MS) por arritmias em pacientes com cardiopatias estruturais em mente e sob o desempenho frustrante de drogas antiarrítmicas na redução de tal risco, surge o cardiodesfibrilador implantável (CDI). O CDI é um marca-passo capaz de diagnosticar arritmias ventriculares sustentadas potencialmente deletérias e aplicar terapias, quando necessário. De início, o dispositivo foi testado em sobreviventes de MS ou que tiveram arritmias ventriculares ameaçadoras à vida nos estudos CIDS<sup>8</sup> (*Canadian Implantable Defibrillator Study*), CASH<sup>9</sup> (*Cardiac Arrest Study Hamburg*) e AVID<sup>10</sup> (*The Antiarrhythmics versus Implantable Defibrillators*), mostrando redução de morte súbita arrítmica e de morte total (em meta-análise posterior<sup>11</sup>).

Já em pacientes que nunca foram acometidos por tais arritmias ou que não tiveram MS abortada prévia, o uso do CDI nos indivíduos com disfunção ventricular significativa (fração de ejeção menor ou igual a 35%) e sintomáticos (classe funcional II ou III da *New York Heart Association*) é embasado principalmente pelos estudos MADIT (*Multicenter Automatic Defibrillator Implantation Trial*),<sup>12</sup> MADIT II<sup>13</sup> e SCD-HeFT (*Sudden Cardiac Death in Heart Failure Trial*),<sup>7</sup> os quais mostraram redução significativa de mortalidade total com o uso do dispositivo.

Duas questões recentes surgiram, no que concerne ao uso do CDI em profilaxia primária: pacientes com cardiopatia dilatada, que classicamente têm menos fibrose e menor probabilidade de arritmias ventriculares por reentrada, teriam o mesmo benefício do dispositivo que pacientes isquêmicos? E os resultados encontrados em estudos da década de 1990, como o MADIT, seriam reproduzidos se o grupo controle fosse tratado com a terapia medicamentosa otimizada atual?

Para a primeira hipótese, o estudo DANISH,<sup>14</sup> incluindo apenas pacientes com cardiopatia dilatada, não encontrou diferenças quanto à mortalidade total ou cardíaca, levando à necessidade de individualização da terapia. Para a segunda hipótese, ao observarmos os estudos pivotais, em que havia baixo uso de betabloqueadores e de antagonistas mineralocorticoides (Tabela 1) e altas taxas de antiarrítmicos, como a amiodarona, isto é, longe do que é

orientado nos dias hodiernos, percebemos que novos estudos são necessários.

**Tabela 1.**

Trial	IECA/BRA	Betabloqueadores	Antagonistas Mineralocorticoides
MADIT I	57%	5%	não especificado
MADIT II	72%	70%	não especificado
SCD-HeFT	88%	79%	33%
DANISH	97%	92%	57%

Uso de terapia médica otimizada para insuficiência cardíaca em pacientes dos grupos controles dos grandes estudos<sup>7,12-14</sup> que embasam o uso do cardiodesfibrilador implantável (CDI) na profilaxia primária de morte súbita

Outra grande questão que permeia o vasto universo do tratamento das arritmias figura no tratamento da sua arritmia sustentada mais frequente: a fibrilação atrial (FA). É inegável o problema de saúde pública que a FA se tornou, representando cerca de 1,5% de todas as internações e 0,8% das mortes nas emergências norte-americanas.<sup>15</sup> Inquestionável, também, sua associação com eventos embólicos, principalmente o acidente vascular encefálico (AVE), e com mortalidade, sendo sempre necessária a avaliação de risco tromboembólico dos pacientes com FA, e a anticoagulação quando esse risco for aumentado (atualmente quando o escore CHA<sub>2</sub>DS<sub>2</sub>-VASc for maior ou igual 2 para homens e maior ou igual a 3 para mulheres) (Tabela 2). Contudo, em relação ao tratamento da arritmia em si, existe uma certeza de que se manter em ritmo sinusal é melhor?

**Tabela 2.**

CHA <sub>2</sub> DS <sub>2</sub> -VASc	Pontuação
Congestive heart failure/left ventricular dysfunction (Insuficiência cardíaca/ disfunção ventricular esquerda)	1
Hypertension (hipertensão)	1
Age ≥ 75 years (Idade ≥ 75 anos)	2
Diabetes mellitus	1

CHA <sub>2</sub> DS <sub>2</sub> -VASc	Pontuação
Stroke/transient ischaemic attack (histórico de AVC)	2
Vascular disease (prior myocardial infarction, peripheral artery disease or aortic plaque) (Doença Vascular)	1
Age 65–74 years (Idade)	1
Sex category (female gender) (Sexo feminino)	1

Escore CHA<sub>2</sub>DS<sub>2</sub>-VASc para a determinação de risco cardioembólico em pacientes com fibrilação atrial não valvar. O risco é significativo quando maior ou igual 2 para homens e maior ou igual a 3 para mulheres.

Em 2002, o estudo AFFIRM<sup>16</sup> (*The Atrial Fibrillation Follow-Up Investigation of Rhythm Management*) randomizou 4.060 pacientes para as estratégias de controle de frequência e controle de ritmo por meio de antiarrítmicos, avaliando o desfecho primário de mortalidade total. Ao final de uma média de 3,5 anos, o desfecho primário não foi diferente entre os grupos, assim como AVE, porém às custas de maior número de hospitalizações no grupo controle de ritmo. Além disso, era permitido interromper a anticoagulação no grupo controle de ritmo, caso o paciente permanecesse mais de 4 semanas em ritmo sinusal, algo proscrito nos dias de hoje. Também no mesmo ano, o estudo RACE<sup>17</sup> (*Rate Control versus Electrical Cardioversion for Persistent Atrial Fibrillation*), incluindo apenas pacientes com FA persistente, também randomizado, trouxe resultados semelhantes ao AFFIRM, equiparando as duas estratégias em relação a desfechos.

No entanto, uma nova variável nesta equação foi adicionada, quando o eletrofisiologista francês Michel Haissaguerre desenvolveu a técnica de ablação por cateter da FA, ao isolar as veias pulmonares, locais onde com maior frequência se encontravam os gatilhos para o início da arritmia. A ablação da FA vem se mostrando superior à terapia com antiarrítmicos no controle de sintomas e de recorrência da arritmia,<sup>18</sup> principalmente quando realizada em pacientes com FA paroxística. Ainda, em pacientes com insuficiência cardíaca de fração de ejeção reduzida (ICFER), mostrou redução de mortalidade, quando comparada com as estratégias medicamentosas convencionais.<sup>19</sup>

O estudo CABANA,<sup>20</sup> publicado em 2019, trouxe ainda mais discussão sobre o tema. Foram incluídos 2.204 pacientes, randomizados entre os grupos ablação por cateter ou tratamento medicamentoso, sendo avaliado o desfecho primário composto de morte, AVE desabilitante, sangramento significativo ou parada cardiorrespiratória. Após a análise por intenção de tratar, não houve diferença entre os grupos acerca do desfecho primário, porém às custas de um grande crossover entre os braços (27,5% do grupo medicamentoso foi submetido à ablação), fato que reduz o poder do estudo.

E, por fim, o estudo EAST-AFNET 4,<sup>21</sup> randomizando 2.789 pacientes com FA recém-descoberta para os grupos controle de ritmo (incluindo ablação) ou controle de frequência, encontrou uma redução do desfecho primário composto de morte cardiovascular, AVE, hospitalização por insuficiência cardíaca ou síndrome coronariana aguda, favorecendo a estratégia de controle de ritmo. Após tantos resultados conflitantes, que fazer? Individualizar à luz das evidências

## Tratamento das valvopatias baseado em evidências

O tratamento das doenças valvares se baseia classicamente na intervenção cirúrgica em valvopatias classificadas ecocardiograficamente como importantes, quando associadas a sintomas, disfunção ventricular ou fatores complicadores específicos, como aumento dos diâmetros do ventrículo esquerdo, baixa tolerância ao teste de esforço, desenvolvimento de FA ou hipertensão arterial pulmonar significativa, entre outros.<sup>22</sup> Tais indicações cirúrgicas vêm principalmente de estudos de coorte e de experiências de grandes centros, que já realizam tais procedimentos há décadas.

A estenose mitral (EMi) tem, como principal etiologia, a doença reumática, em especial nos países em desenvolvimento, como o Brasil.<sup>22</sup> Seu tratamento, intervencionista, é indicado para pacientes com valvopatia significativa sintomática ou para aqueles com complicadores, como fibrilação atrial ou hipertensão pulmonar, e ele pode ser realizado a partir da comissurotomia valvar cirúrgica ou da valvuloplastia por cateter-balão (VPCB), a depender do grau de acometimento mitral, isto é, quanto menor o envolvimento do aparelho subvalvar, espessamento e calcificação das cúspides e maior o grau de mobilidade, maior o benefício da VPCB.<sup>23</sup>

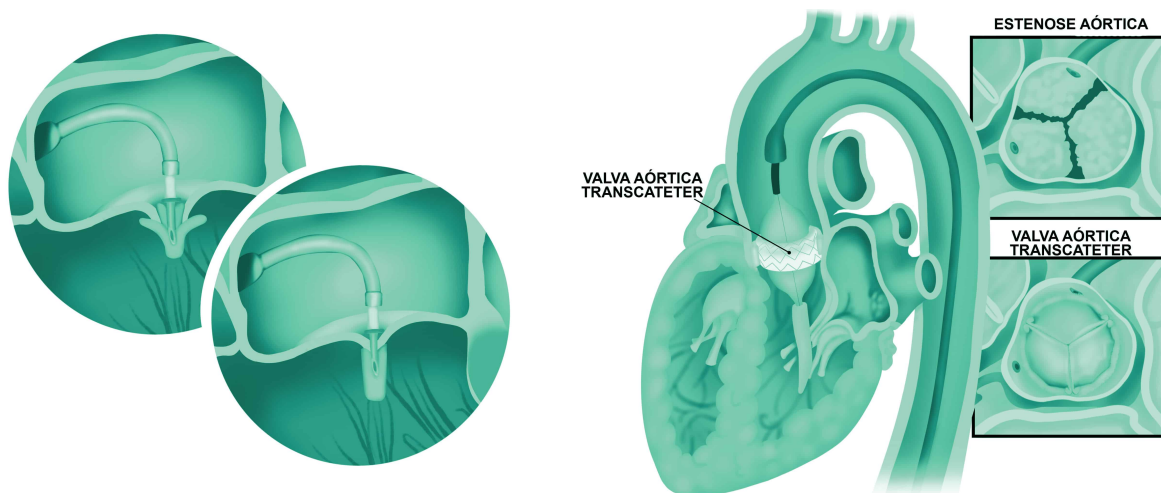
O conhecimento que levou a perceber que seria necessária uma intervenção nesses doentes vem de coortes antigas, em uma época em que

não existia a cirurgia, as quais mostravam que, mesmo com tratamento clínico medicamentoso, a mortalidade era muito elevada, principalmente nos sintomáticos.<sup>24</sup> Em 1948, a cirurgia de comissurotomia mitral foi desenvolvida, sendo indicada em especial para aqueles com sintomas, com melhora desses após o procedimento, em uma época em que a avaliação da valvopatia não era feita por ecocardiograma (pois este nem existia).<sup>25</sup> Já a VPCB foi testada *versus* a cirurgia em estudos randomizados pequenos,<sup>26-27</sup> mostrando-se não inferior ao tratamento cirúrgico, com indicações aperfeiçoadas em novos estudos de coorte.<sup>28</sup>

Já a insuficiência mitral (IMi), que também tem a doença reumática, no nosso meio, como principal etiologia, apresenta-se, mundo afora, como consequência da degeneração valvar ou pelo prolapso da válvula mitral (PVM). A IMi tem sua correção indicada naqueles com valvopatia importante pelo ecocardiograma transtorácico aliados a sintomas, disfunção ventricular esquerda ou a dilatação do ventrículo esquerdo (VE).<sup>29</sup> Nos pacientes assintomáticos, existe controvérsia se a estratégia do *watchful waiting* é segura, quando comparada à cirurgia precoce, mas novamente, baseando-se em estudos de coorte.<sup>29-30</sup> Tal tratamento pode ser realizado a partir da plastia da válvula (preferível, quando viável) ou mesmo de sua troca por próteses, sejam biológicas ou mecânicas.<sup>31</sup>

Recentemente, o reparo de valva mitral percutâneo utilizando o MitraClip foi adicionado à miríade terapêutica da IMi, questionando a soberania da cirurgia (Figura 2). O MitraClip foi inicialmente testado no cenário da insuficiência mitral (IM) primária importante, como uma alternativa à cirurgia de troca/plastia valvar. O estudo EVEREST II<sup>32</sup> (*Endovascular Valve Edge-to-Edge Repair Study II*) randomizou 279 pacientes com insuficiência mitral importante (graus 3+ ou 4+), em uma razão 2:1, entre os grupos reparo percutâneo vs. cirurgia convencional, sob o desfecho primário composto de morte, necessidade de cirurgia por disfunção da valva mitral e regurgitação mitral significativa aos 12 meses de seguimento.

## Figura 2.



Reparo percutâneo mitral por meio de MitraClip (à esquerda) e implante transcater de valva aórtica (à direita). Adaptado de: Feldman T, Foster E, Glower DD, et al; EVEREST II Investigators. Percutaneous repair or surgery for mitral regurgitation. *N Engl J Med.* 2011 Apr 14;364(15):1395-406 e Smith CR, Leon MB, Mack MJ, et al; PARTNER Trial Investigators. Transcatheter versus surgical aortic-valve replacement in high-risk patients. *N Engl J Med.* 2011 Jun 9;364(23):2187-98.

O reparo percutâneo teve pior desempenho neste cenário, principalmente às custas da necessidade de cirurgia por disfunção da valva mitral, porém apresentou-se com baixa taxa de complicações e com melhora significativa de classe funcional e redução de diâmetros ventriculares ao final do seguimento, além de mortalidade total semelhante entre os grupos, achados que foram ratificados em meta-análise posterior.<sup>33</sup> Devido a isso, no cenário da IM primária, o MitraClip tornou-se uma opção para aqueles pacientes com alto risco cirúrgico (ou impeditivo).

Tendo como base a segurança do MitraClip e sua indicação para pacientes com alto risco cirúrgico ou impedidos de realizar a cirurgia de reparo/troca valvar, foi aventada a hipótese de que o reparo percutâneo seria uma boa alternativa em pacientes com IM secundária, classicamente mais graves e com resultados cirúrgicos pobres. Para testar essa hipótese, foram formulados 2 estudos semelhantes, o COAPT<sup>34</sup> (*Cardiovascular Outcomes Assessment of the MitraClip Percutaneous Therapy for Heart Failure Patients with Functional Mitral Regurgitation*) e o MITRA-FR<sup>35</sup> (*Percutaneous Repair with the MitraClip Device for Severe Functional/Secondary Mitral Regurgitation*), ambos publicados em 2018.

Ambos randomizaram pacientes entre os grupos tratamento clínico e tratamento clínico associado ao reparo valvar percutâneo, com critérios de inclusão e exclusão parecidos entre si. No COAPT, o desfecho primário foi

de hospitalização por IC ao final de 1 ano, e, no MITRA-FR, foi de morte ou hospitalização por IC, também ao final de 1 ano.

Apesar das semelhanças, os resultados foram bem distintos: no COAPT, houve menos desfechos primários no grupo MitraClip, além de redução na mortalidade total, enquanto no MITRA-FR, não houve diferenças entre os grupos. Analisando os dados de cada um dos *trials*, percebemos que, no COAPT, os pacientes tinham maior grau de doença valvar e menor grau de disfunção ventricular, o que pode ter trazido melhores resultados na correção da IM secundária. Estudar as nuances dos estudos, avaliar com detalhes sua metodologia e saber interpretar os resultados é fundamental para individualizar seus achados, com base a proporcionar o melhor ao paciente.

A estenose aórtica (EAo), devido ao envelhecimento populacional, tem prevalência crescente, uma vez que a degeneração/calcificação aórtica é sua principal etiologia. O tratamento intervencionista geralmente é indicado nas EAo importantes sintomáticas, ou naqueles assintomáticos com complicadores, como disfunção ventricular esquerda, áreas valvares muito diminutas ou baixa performance no teste ergométrico.<sup>22</sup> A estratégia da cirurgia de troca valvar precoce em pacientes assintomáticos com área valvar diminuta versus o acompanhamento clínico foi recentemente testada em ensaio clínico randomizado envolvendo 145 pacientes, mostrando benefício para a cirurgia, no que concerne ao desfecho de morte nos primeiros 30 dias após a cirurgia ou morte cardiovascular no período total de acompanhamento.<sup>36</sup> Apesar de randomizado e com um desfecho duro (morte), tal estudo merece observações. A mortalidade cirúrgica foi de zero, algo que não representa o mundo real. Além disso, a idade média dos pacientes incluídos é relativamente baixa, comparando-se os estudos mais recentes. Logo, a indicação da troca valvar cirúrgica nesses pacientes assintomáticos deve ser individualizada de acordo com o perfil e desfechos cirúrgicos de onde se trabalha.

Uma vez que a idade avançada é fator de maior risco para as cirurgias cardíacas e que esses pacientes frequentemente têm comorbidades associadas, a troca valvar cirúrgica muitas vezes era contraindicada, devido a altas taxas de mortalidade. Com esse racional, foi desenvolvida o implante transcater da válvula aórtica (TAVI) (Figura 2).

Inicialmente, o implante transcater foi testado em pacientes com risco cirúrgico proibitivo para a troca valvar convencional no estudo PARTNER<sup>37</sup> (*Placement of Aortic Transcatheter Valves*), sendo randomizados 358



pacientes entre os grupos TAVI vs. terapia clínica (os pacientes poderiam ser submetidos a valvoplastia por balão). O desfecho primário foi o de morte por qualquer causa. Houve diferença significativa entre os grupos em favor da TAVI com uma redução importante de mortalidade (50,7% x 30,7%), às custas de maior taxa de complicações vasculares (lembrando que estava sendo comparada com tratamento clínico).

A TAVI também foi avaliada em pacientes de risco cirúrgico alto<sup>38</sup> e intermediário,<sup>39</sup> comparada à cirurgia convencional, em estudos randomizados cujos desfechos principais foram morte por todas as causas em 1 ano – no estudo de pacientes com risco alto, e morte por todas as causas e AVE desabilitante em 2 anos – no de pacientes com risco intermediário, mostrando-se não inferior ao tratamento convencional.

Além disso, mais recentemente, em 2019, a TAVI foi testada frente ao tratamento cirúrgico em pacientes de baixo risco no estudo PARTNER 3.<sup>40</sup> Foram randomizados 1.000 pacientes entre os grupos, e o desfecho primário foi de morte, AVE e rehospitalização em 1 ano de seguimento médio, sendo menor no grupo TAVI. Logo, a cirurgia está morta para a EAo, certo? Não é bem assim.

Precisamos perceber que o procedimento cirúrgico traz maiores benefícios quanto mais longo é o seguimento, portanto estudos que avaliam desfechos em 1 ano geralmente são desfavoráveis à cirurgia. Além disso, apesar de vários estudos mostrarem que as próteses da TAVI são duradouras, precisamos de avaliações mais longas, principalmente em pacientes mais jovens, que tenham maior expectativa de vida. Por fim, o desfecho composto utilizado no PARTNER 3 não é tão “duro” quanto os dos estudos anteriores, com a variável hospitalização claramente favorecendo o grupo TAVI. Entretanto, é inegável que a TAVI veio para ficar.

Já a insuficiência aórtica (IAo) tem como principais etiologias a reumática, a degenerativa e a bicúspide. Caracteriza-se por ser uma valvopatia com sintomas mais indolentes, de progressão mais arrastada, em que o paciente permanece assintomático por longos períodos.<sup>41</sup> Contudo, em pacientes com IAo importante, a taxa de mortalidade é elevada e a cirurgia frequentemente requerida, como mostrado por Dujardin et al., em sua coorte com seguimento médio de 7 anos, sendo os sintomas a disfunção ventricular esquerda e diâmetros elevados do VE marcadores de pior prognóstico.<sup>42</sup> O tratamento cirúrgico é indicado em pacientes com IAo importante sintomática ou em pacientes assintomáticos com complicadores (FEVE

reduzida e diâmetros aumentados) e se baseia principalmente na troca valvar com próteses mecânicas ou biológicas.<sup>43</sup>

## Tratamento da insuficiência cardíaca baseado em evidências

O tratamento medicamentoso da insuficiência cardíaca com fração de ejeção reduzida (ICFER), com comprovada eficácia e maior probabilidade de redução de mortalidade, é composto por um inibidor da enzima conversora de Angiotensina (iECA)<sup>44</sup> ou um bloqueador do receptor da Angiotensina II (BRA),<sup>45</sup> quando o iECA não puder ser usado por efeitos adversos; por um betabloqueador, mais precisamente carvedilol,<sup>46-47</sup> succinato de metoprolol<sup>48</sup> ou bisoprolol;<sup>49</sup> e por um diurético antagonista do receptor mineralocorticoide<sup>50</sup> há décadas.

Entretanto, recentemente, duas novas classes de drogas foram introduzidas no cenário clínico buscando um lugar nas indicações rotineiras de medicamentos para ICFER. São elas o inibidor da neprilisina e os inibidores do cotransportador sódio-glicose 2 (iSGLT-2).

O inibidor da neprilisina foi avaliado no contexto da ICFER pelo estudo PARADIGM-HF<sup>51</sup> (*Prospective Comparison of ARNI [Angiotensin Receptor–Neprilysin Inhibitor] with ACEI [Angiotensin-Converting–Enzyme Inhibitor] to Determine Impact on Global Mortality and Morbidity in Heart Failure*), publicado em 2014. Nele, 8.442 pacientes com insuficiência cardíaca sintomática e com fração de ejeção do ventrículo esquerdo menor ou igual a 40% foram randomizados entre os grupos inibidor da neprilisina e enalapril, após uma fase de *run-in* para cada droga. O estudo foi duplo-cego, e os pacientes seriam seguidos por 34 meses. Porém, ao final de 27 meses, o estudo foi interrompido pela redução de desfecho primário (morte e internação por IC) no grupo inibidor da neprilisina (26,5% vs. 21,8%), assim como de mortalidade total.

Para o leitor desavisado, temos um estudo perfeito, isento de críticas. No entanto, existem alguns pontos que merecem discussão e reflexão. A fase de *run-in* é usada para avaliar se o medicamento em questão é tolerável pelo paciente. Uma vez que não o seja, o paciente é excluído do estudo. Perceba, caro leitor, que a avaliação de efeitos adversos provocada por um medicamento é vital para a avaliação da segurança de seu uso e de risco vs. benefício. Além disso, na “vida real” não existe fase de *run-in*: se o paciente

desenvolver algum efeito deletério pelo medicamento, não há como excluí-lo.

Em segundo lugar, uma crítica ao estudo PARADIGM-HF é a de que a dose máxima de inibidor da neprilisina foi comparada à metade da dose máxima do enalapril (10 mg, duas vezes ao dia), uma vez que, frequentemente, chega-se a essa no tratamento dos pacientes com IC. Embora, no estudo SOLVD,<sup>44</sup> que introduziu o enalapril no rol do tratamento da IC, a dose usada do medicamento *versus* o placebo tenha sido de 10 mg, duas vezes ao dia, o ideal seria que, em um estudo que tenta introduzir uma nova droga que também tem ação tanto inibidora do sistema renina-angiotensina-aldosterona (SRAA) quanto vasodilatadora em comparação ao tratamento padrão, ambas sejam contrapostas em sua ação máxima. Por fim, estudos truncados, isto é, aqueles que são interrompidos de maneira precoce pela demonstração do benefício da terapia, carregam o viés de superestimar a magnitude do benefício, além de privar a avaliação de efeitos adversos tardios. Logo, a análise criteriosa do estudo foi capaz de encontrar falhas metodológicas que comprometem sua validade. Os iSGLT-2 são medicamentos antidiabéticos orais que atuam no túbulo contorcido proximal, aumentando a glicosúria. Seu uso foi idealizado para ser mais uma opção no arsenal contra a diabetes mellitus tipo 2 (DM2), porém tudo mudou quando se começou a observar uma redução de desfechos cardiovasculares, liderados pela diminuição no número de internações por IC, nos estudos que avaliavam a droga em pacientes com DM2.<sup>52-54</sup>

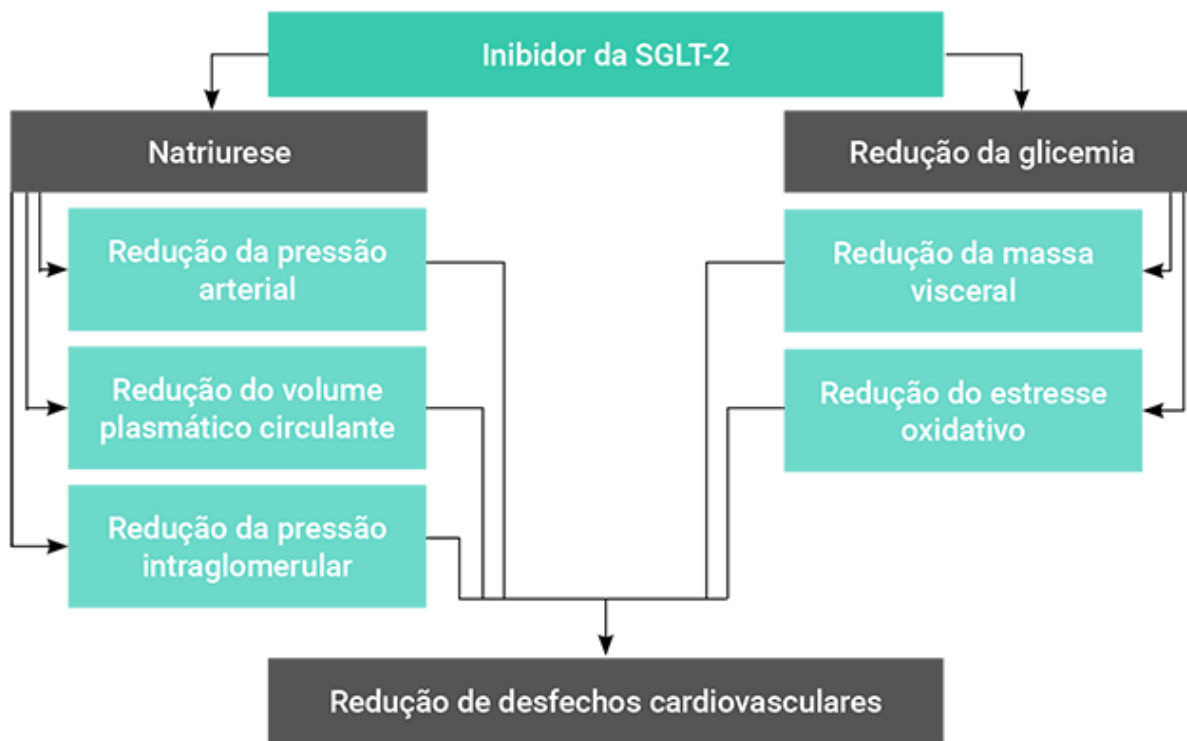
Baseados nestes achados, investigou-se o impacto dos iSGLT-2 em pacientes com ICFER, com ou sem DM2. No DAPA-HF<sup>55</sup> (*Dapagliflozin in Patients with Heart Failure and Reduced Ejection Fraction*), publicado em 2019, foram incluídos 4.744 pacientes com IC sintomáticos, fração de ejeção menor ou igual a 40%, com ou sem DM2, randomizados entre os grupos placebo e dapaglifozina, sendo avaliado o desfecho primário composto de piora da IC ou morte cardiovascular. Nesse ensaio clínico, houve redução do desfecho composto e de seus componentes, favorecendo o grupo dapaglifozina, com benefício mantido mesmo naqueles sem DM2.

Os resultados do DAPA-HF foram reproduzidos no EMPEROR-Reduced<sup>56</sup> (*Cardiovascular and Renal Outcomes with Empagliflozin in Heart Failure*), dessa vez com outro iSGLT-2, a Empaglifozina, mostrando uma tendência a acreditarmos se tratar de um efeito de classe. Ainda, como vem sendo retratado, há também benefício do uso dessas medicações no

âmbito de redução da progressão da doença renal crônica,<sup>57</sup> com consistência louvável em vários cenários. Devido aos empolgantes resultados, os iSGLT-2 vêm figurando como a quarta droga para o tratamento da ICFER, principalmente naqueles pacientes que já estão com a terapia medicamentosa otimizada.

A grande questão que envolve os iSGLT-2 é o mecanismo (ou os mecanismos) de ação relacionado à redução de desfechos cardiovasculares (Figura 3). Uma vez que os benefícios ocorreram mesmo em pacientes sem DM2, propõe-se que os efeitos fisiopatológicos vão além daqueles na glicemia. A natriurese provocada pela inibição do cotransportador sódio-glicose no túbulo contorcido proximal, associada aos efeitos de redução da pressão arterial, do volume plasmático circulante e da pressão intraglomerular; e, por outro lado, seu poder hipoglicemiante, redutor da massa visceral e do estresse oxidativo, em conjunto, são os prováveis responsáveis pelos excelentes resultados encontrados em pacientes com risco cardiovascular elevado.<sup>58</sup>

**Figura 3.**



Mecanismos de ação propostos para a redução de desfechos cardiovasculares provocada pelos inibidores da SGLT-2 (adaptado de Cowie MR, Fisher M. SGLT2 inhibitors: mechanisms of cardiovascular benefit beyond glycaemic control. *Nat Rev Cardiol.* 2020 Dec;17(12):761–72).

## Tratamento da doença arterial coronariana baseado em evidências

A doença arterial coronariana (DAC) é talvez uma das entidades mais estudadas na medicina. Até a década de 1950, a mortalidade intra-hospitalar nos pacientes com infarto agudo do miocárdio (IAM) era de 35%-45%. Um dos grandes responsáveis pela mudança na mentalidade vigente foi o médico Desmond Julian, com o advento da unidade coronariana no *Edinburgh Royal Infirmary*, conseguindo uma redução de 50% na mortalidade.<sup>59</sup> Desde então, várias condutas foram testadas em ensaios clínicos randomizados buscando reduzir a ainda significativa mortalidade dessa doença (Figura 4).

**Figura 4.**

Ano	Conduta	Mortalidade / Morbidade	Tratamento medicamentoso
60s	<ul style="list-style-type: none"> <li>• IAM: artéria coronária ocluída</li> <li>• Cateterismo cardíaco / angiografia coronária contraídicada</li> </ul>	<p>&gt; 30%</p> <p>5%</p>	<ul style="list-style-type: none"> <li>• "Código azul" na UCO</li> <li>• Enfermeiras usam desfibrilador;</li> </ul>
70s			<ul style="list-style-type: none"> <li>• Beta-Bloqueadores</li> </ul>
80s	<ul style="list-style-type: none"> <li>• Oclusão coronária aguda -&gt; IAMCSST</li> <li>• Abrir artéria culpada com trombólise melhora desfechos</li> <li>• ATC na lesão culpada não é recomendada após trombólise com sucesso</li> </ul>		<ul style="list-style-type: none"> <li>• Trombólise</li> <li>• AAS</li> <li>• Reabilitação</li> </ul>
90s	<ul style="list-style-type: none"> <li>• ATC primária recomendada</li> <li>• ATC em lesões não culpadas não recomendada</li> <li>• Lesões não-culpadas relacionadas a pior desfecho</li> </ul>		<ul style="list-style-type: none"> <li>• Estatinas</li> <li>• AAS + Clopidogrel</li> <li>• Anticoagulação melhor</li> </ul>
2000s	<ul style="list-style-type: none"> <li>• Trabalho em equipe, padronização das condutas no IAM</li> </ul>		<ul style="list-style-type: none"> <li>• Novos Inibidores da P2y<sub>12</sub></li> <li>• Hipotermia</li> </ul>
2019 2020	<ul style="list-style-type: none"> <li>• Revascularização completa melhora desfechos?</li> <li>• Revascularização em pacientes crônicos estáveis?</li> </ul>		<ul style="list-style-type: none"> <li>• Nova prevenção em pacientes de alto risco</li> </ul>

Evolução da revascularização, terapias medicamentosas e desfechos no manejo do IAMCSST. SCA: síndrome coronariana aguda; IAM: infarto agudo do miocárdio, ASA: ácido acetilsalicílico; UCo, unidade coronariana, ATC: angioplastia transluminal coronária; IAMCSST: infarto com supradesnívelamento de ST. Adaptado de: Sendón et al. REC Interv Cardiol. 2019;3:203-206.

## Estratégias de revascularização na doença arterial coronariana crônica

A cirurgia de revascularização do miocárdio (CRVM), idealizada no início do século XX, tornou-se a conduta padrão após Kolesov na Rússia, em 1964, utilizando anastomose de artéria torácica interna, e Favaloro em 1967, na Cleveland Clinic, com enxertos de veia safena, publicarem seus resultados.<sup>60-61</sup> Até a década de 1990 a CRVM era a principal estratégia de revascularização na DAC crônica.

Duas décadas antes, Grüntzig iniciou testes para uma nova estratégia de tratamento das estenoses coronarianas, por meio da angioplastia por cateter-balão (ATCP).<sup>62-63</sup> Como havia uma taxa significativa de reestenoses (20%-40%), estudos com implante de um dispositivo metálico chamado *stent* mostraram redução na taxa de reestenose, porém com aumento na taxa de trombose intrastent.<sup>64</sup> Foi com uso da dupla antiagregação plaquetária (DAPT) e dos *stents* farmacológicos que essas taxas foram reduzidas, como demonstrado nos estudos SESAMI (*Randomized trial of Sirolimus-Eluting Stent Versus Bare-Metal Stent in Acute Myocardial Infarction*) e TYPHOON (*Trial to assess the use of the cYPHer sirolimus-eluting stent in acute myocardial infarction treated with balloon angioplasty*).<sup>65-66</sup>

O primeiro estudo que comparou estratégias diferentes na DAC foi o MASS (*The Medicine, Angioplasty or Surgery Study*), publicado em 1995, que comparou 214 pacientes uniarteriais randomizados para tratamento médico otimizado (composto na época por betabloqueador, nitrato e antiplaquetário), ATCP e CRVM. O resultado evidenciou taxa de mortalidade similar entre os três grupos após 3 anos, porém com maior necessidade de novas intervenções no grupo da ATCP e mais angina no grupo do tratamento clínico.<sup>67</sup> O estudo MASS II, com 611 pacientes, dessa vez multiarteriais, publicou em 2010 sua análise de 10 anos e corroborou os resultados de outros estudos como BARI (*Bypass Angioplasty Revascularization Investigation*) e BARI2D (*Bypass Angioplasty Revascularization Investigation 2 Diabetes*) (que demonstrou CRVM melhor que TMO em diabéticos).<sup>68-71</sup> Com a disseminação do uso do *stent* metálico e depois farmacológico, os resultados desses estudos não mais representavam a realidade. Outros estudos foram então realizados comparando *stents* mais modernos com tratamento médico otimizado (TMO) ou CRVM.

O estudo COURAGE (*Clinical Outcomes Utilizing Revascularization and Aggressive Drug Evaluation*), publicado em 2007, selecionou mais de 35.000 pacientes, porém apenas 2.287 pacientes foram recrutados para TMO ou TMO e ATC, selecionando pacientes com isquemia leve ou mesmo ausente, portanto, muitos sem indicação de angioplastia. Os resultados não encontraram diferenças em relação à mortalidade. Esse estudo ainda utilizou em sua maioria *stents* metálicos que, no momento da publicação, já não representavam o melhor dispositivo para ATC, havendo uma taxa de *crossover* importante entre os grupos devido à necessidade de intervenção, sendo uma das principais críticas ao estudo, mas que representa a vida real.<sup>72</sup>

Comparando CRVM com TMO em pacientes com disfunção ventricular esquerda, foi publicado o STICH (*Surgical Treatment for Ischemic Heart Failure*) em 2011 e o STICHES (*Surgical Treatment for Ischemic Heart Failure Extension Study*) em 2016. No primeiro, o desfecho morte por todas as causas não mostrou diferença estatisticamente significativa em 5 anos entre os grupos.<sup>73</sup> Na análise de 10 anos o valor de  $P = 0,02$ , demonstrando significância estatística, com NNT de 14, favorecendo a cirurgia.<sup>74</sup> Seriam resultados discordantes? Precisamos analisar a magnitude do efeito. Então a melhor forma seria utilizar o NNT do Hazard Ratio com a fórmula de Altman, considerando assim o tempo de exposição e a probabilidade cumulativa do desfecho.<sup>75</sup> Chegamos então a um NNT de 25 podendo variar de 14 a 129. Então ambos os estudos são negativos, o STICH não demonstrou benefício relevante enquanto o STICHES demonstrou benefício pouco relevante.

Outros estudos compararam ATC versus CRVM em multiarteriais e/ou com estenoses de tronco de coronária esquerda (TCE), como o SYNTAX trial em 2009, que contribuiu mostrando que o tratamento com ATC com *stents* farmacológicos de primeira geração (paclitaxel) versus CRVM são similares em pacientes de baixa complexidade anatômica. Nos pacientes de complexidades intermediária ou alta, a CRVM permaneceu como tratamento padrão. Os pacientes tratados com ATC tiveram uma maior taxa de reintervenção.<sup>76</sup> O estudo FREEDOM (*Future Revascularization Evaluation in Patients with Diabetes Mellitus: Optimal Management of Multivessel Disease*) publicado em 2012 foi o primeiro a realmente demonstrar que CRVM diminui mortalidade. Ele evidenciou resultados similares ao SYNTAX trial (*The SYNergy between Percutaneous Coronary Intervention*

*and Cardiac Surgery Study*), porém com uma taxa significativamente maior de AVC no grupo da CRVM.<sup>77</sup>

Mais recentemente, o estudo EXCEL (*Evaluation of XIENCE versus Coronary Artery Bypass Surgery for Effectiveness of Left Main Revascularization*) publicado em 2016 comparou CRVM versus ATC com *stent* farmacológico de segunda geração (everolimus) para o desfecho composto de morte, infarto e AVC (inadequado, visto que sabidamente a CRVM causa mais AVC comparado a ATC, trazendo desfechos em direção opostas) em 3 anos, em pacientes com estenose de TCE e complexidade anatômica baixa ou intermediária. O estudo demonstrou não inferioridade do tratamento percutâneo, porém com alguns vieses importantes, como a análise *intention-to-treat*, quando seria mais adequado nesse caso *per-protocol*. Houve ainda críticas relacionadas às definições de infarto e o uso de infarto periprocedimento como composto do desfecho primário, visto que combinado com infarto espontâneo têm direções opostas, quando se compara CRVM e ATC.<sup>78</sup> Mais adequado seria deixar o primeiro como um desfecho de segurança e apenas o último como desfecho primário, como realizado no estudo NOBLE (*The Nordic-Baltic-British left main revascularisation study*).<sup>79</sup>

O ORBITA (*The Objective Randomised Blinded Investigation with optimal medical Therapy of Angioplasty in stable angina*) foi um estudo que comparou de forma elegante 200 pacientes submetidos a cateterismo cardíaco e reserva de fluxo fracionada (FFR). Se FFR positivo para isquemia, os pacientes sedados eram randomizados para ATC ou um sham procedure (procedimento simulado como efeito placebo). O resultado não demonstrou diferença significativa entre os grupos para o desfecho de aumento no tempo de exercício.<sup>80</sup> Cada vez mais se provava que os benefícios da revascularização se dão quando há presença de isquemia importante, como tentou demonstrar o estudo FAME-2, que apesar de truncado e com potencial viés de aferição, utilizou o FFR na definição de quais artérias coronárias tratar, não demonstrando redução de mortalidade ou infarto, apenas redução de hospitalização por sintomas e nova revascularização urgente, corroborando os resultados do estudo COURAGE.<sup>81</sup>

Para tentar responder essa questão definitivamente surgiu então o ISCHEMIA trial (*International Study of Comparative Health Effectiveness with Medical and Invasive Approaches*), estudo que custou 100 milhões de



dólares, o mais caro da história, e foi publicado em 2020. Foram randomizados 5.179 pacientes com isquemia moderada ou grave para TMO ou cateterismo e ATC se necessário. O desfecho composto inicial foi morte cardiovascular e infarto. Os autores, no entanto, acrescentaram hospitalização e insuficiência cardíaca ao desfecho primário devido à baixa incidência do desfecho original. Os resultados não demonstraram diferença entre os grupos.<sup>82</sup> Modificar o desfecho primário do estudo, mesmo que definido *a priori*, foi uma decisão ruim, já que tirou confiabilidade, misturando desfechos objetivos (morte e IAM) com desfechos subjetivos (angina instável, insuficiência cardíaca), além de fugir do cerne da questão: mesmo em doentes com isquemia moderada/grave, a DAC crônica tem bom prognóstico.<sup>83</sup>

A DAC estável é uma doença em que as intervenções testadas nesses estudos parecem demonstrar pouco impacto nos desfechos clínicos, evidenciando que não há diferença entre pequeno benefício encontrado na maioria dos estudos e ausência de benefício.

## Drogas e estratégias de revascularização no infarto agudo do miocárdio

No infarto do miocárdio (IAM), com o advento da angioplastia por balão e, posteriormente, com *stent*, iniciou-se uma era de grande modificação nas condutas vigentes, por conseguinte, nos desfechos dessa doença. Primeiramente associados à trombólise química, com resultados ruins, foram então testados sozinhos, tendo demonstrado sucesso em grandes ensaios como o PAMI (*Primary Angioplasty in Myocardial Infarction*), estudo multicêntrico, publicado em 1993, com quase 400 pacientes randomizados para ATCP ou alteplase, que evidenciou benefício para o desfecho combinado de IAM não-fatal ou morte.<sup>84</sup> Antes da era da reperfusão algumas condutas farmacológicas foram testadas como a prescrição de betabloqueadores, antiagregantes plaquetários, anticoagulantes e trombolíticos.

Os betabloqueadores foram testados no IAM em dois estudos principais: o ISIS-1 e o MIAMI. O ISIS-1 (*International Study of Infarct Survival*) foi um estudo aberto, publicado em 1986, com mais de 16.000 pacientes, não controlado e foi negativo, isto é, não houve significância estatística para o desfecho mortalidade ( $p = 0,07$ ).<sup>85</sup> Já o MIAMI (*Metoprolol In Acute Myocardial Infarction*) foi um estudo randomizado duplo-cego, com mais de

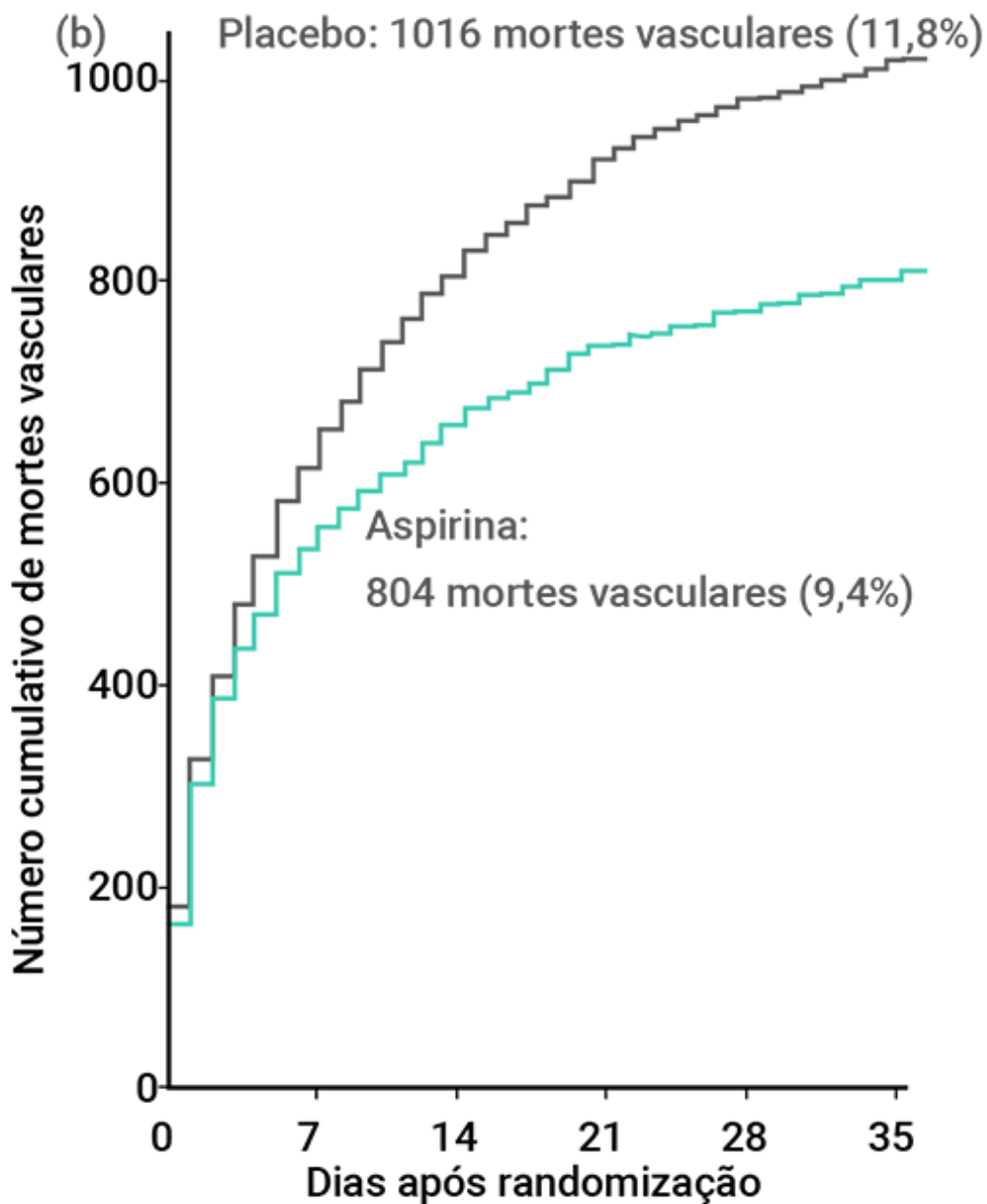
8.600 pacientes, controlado com placebo, publicado um ano antes, que também não mostrou diminuição de mortalidade com Metoprolol ( $p = 0,29$ ).<sup>86</sup> Nesse contexto, um estudo mais bem desenhado (MIAMI), com menos risco de viés foi erroneamente contestado por um estudo sujeito a um viés de desempenho enorme, mas que foi aceito com entusiasmo exagerado pela comunidade científica na época e que até os dias atuais se propaga.

Outras drogas testadas foram os antiagregantes plaquetários e trombolíticos. O ácido acetilsalicílico (AAS) é utilizado na medicina há muitos anos. Ele age sobre a ciclo-oxigenase 1 (COX-1), inibindo a formação de tromboxano A<sub>2</sub> (TXA<sub>2</sub>) e induzindo inibição da função plaquetária. O TXA<sub>2</sub> é produzido em resposta à ativação plaquetária e induz a agregação plaquetária, mediando a expressão do complexo glicoproteína IIb/IIIa. Nos anos de 1960-70, alguns estudos pequenos foram realizados com fibrinolíticos como a estreptoquinase, uma proteína secretada pelos estreptococos que se ligam e ativam o plasminogênio humano, sendo utilizada como trombolítico.

Nas décadas de 1970 e 1980, estudos publicados por Lewis et al. em 1983 e Cairns et al. em 1985 estudaram o AAS no contexto de angina instável, com ambos obtendo redução de mortalidade, porém com algumas limitações metodológicas (cálculo amostral não descrito, amostra pequena).<sup>87-88</sup>

No IAM, seu uso ficou estabelecido a partir de 1988 com a publicação do estudo ISIS-2, um ensaio clínico randomizado, placebo controlado, 2x2, com mais de 17 mil pacientes, que testou estreptoquinase, 1 mês de AAS na dose de 162,5 mg/dia. Foi demonstrada uma redução de mortalidade vascular de 20% com AAS, 23% com estreptoquinase e 40% com as duas drogas combinadas, em 5 semanas. Com uma redução de risco absoluta de 2,4 mortes em cada 100 pacientes tratados com AAS, o NNT foi calculado em 42, estabelecendo uma conduta de grande impacto (Figura 5).<sup>89</sup>

**Figura 5.**



Mortalidade vascular cumulativa para todos os pacientes alocados para Aspirina vs. Placebo em 0-35 dias. Adaptado de: Lancet. 1988 Aug 13; 2(8607):349-60.

Outros estudos como o GISSI 2 (*Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico*), GUSTO (*The Global Utilization of Streptokinase and Tissue Plasminogen Activator for Occluded Coronary Arteries*) e ASSENT (*The Assessment of the Safety and Efficacy of a New Thrombolytic Regimen*), todos com grande número de pacientes, testaram outros trombolíticos (alteplase e tenecteplase), demonstrando bons resultados, sem, no entanto, evidenciar diferenças significativas entre eles.<sup>90-</sup>

Nos anos 2000 foi testada uma estratégia de reperfusão no IAM com supradesnivelamento de ST (IAMCSST) com trombolíticos seguido de angioplastia com stent, chamada angioplastia facilitada, porém os resultados do estudo ASSENT-4 PCI, desenhado para randomizar 4.000 pacientes e truncado após 1.667 pacientes apenas devido a ocorrência do desfecho primário (composto de morte ou insuficiência cardíaca congestiva ou choque em 90 dias) significativamente maior no grupo da angioplastia facilitada, já presente nos primeiros 30 dias, além de mais eventos adversos como AVC e reinfarto.<sup>93</sup> Atualmente os trombolíticos são drogas de escolha no IAMCSST apenas em lugares em que a estratégia de angioplastia primária não está disponível em tempo hábil (> 120 minutos). A estratégia de ATC com *stent* farmacológico constitui a principal forma de reperfusão miocárdica, sendo o padrão-ouro no IAMCSST (Figura 6).<sup>36</sup>

**Figura 6.**



Apresentação dos sintomas **0**

Fase precoce do IAMCSST

3 horas

12 horas

IAMCSST evoluído

48 horas

IAMCSST recente

ATC PRIMÁRIA

I A



I A

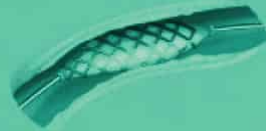


Fibrinólise

(apenas se a ATC não puder ser realizada dentro de 120 min do diagnóstico do IAMCSST)

ATC PRIMÁRIA

I A



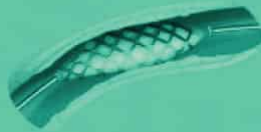
I A



Fibrinólise

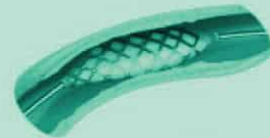
(apenas se a ATC não puder ser realizada dentro de 120 min do diagnóstico do IAMCSST)

ATC primária  
(se sintomas, instabilidade hemodinâmica ou arritmias)



I C

ATC primária  
(pacientes assintomáticos estáveis)



IIa B

ATC rotina  
(pacientes assintomáticos estáveis)

III A



Estratégias de reperfusão na artéria relacionada ao infarto de acordo com o tempo de início dos sintomas e nível de evidência / classe de recomendação. IAMCSST: infarto agudo do miocárdio com supradesnivelamento de ST; ATC: angioplastia transluminal coronária. Adaptado de: EurHeartJ, 2018;39(2):119-177.

Uma das condutas adjuvantes no IAM é a utilização precoce da anticoagulação com heparina. A heparina não fracionada (HNF) foi utilizada no estudo de Theroux et al. publicado em 1988 e demonstrou diminuição de IAM em pacientes com angina instável. Desde então, buscou-se substituir a HNF devido às suas desvantagens clínicas e práticas, como necessidade de monitorizar o tempo de tromboplastina parcial ativado, resposta anticoagulante pouco previsível e a indução de trombocitopenia.<sup>95</sup> As heparinas de baixo peso molecular (HBPM) foram então testadas contra HNF no cenário de síndrome coronariana aguda. O estudo ESSENCE (*Efficacy and Safety of Subcutaneous Enoxaparin in Non-Q-Wave Coronary Events*), publicado em 1997, com mais de 3.000 pacientes randomizados, mostrou superioridade ante a HNF no desfecho combinado de morte, IAM e angina recorrente.<sup>96</sup> Um estudo similar, o TIMI 11B (*Thrombolysis in Myocardial Infarction*), publicado em 1998, também demonstrou os mesmos resultados.<sup>97</sup> Outras HBPM como a dalteparina e nadroparina não demonstraram superioridade, como visto nos estudos FRIC (*Fragmin in unstable coronary artery disease study*) e FRAX.I.S. (*Fraxiparin in Ischaemic Syndrome*).<sup>98-99</sup>

Uma classe de drogas que também modificou a sobrevida da DAC foram os antiagregantes plaquetários inibidores da P2Y<sub>12</sub>. O clopidogrel foi avaliado no estudo CURE (*The Clopidogrel in Unstable Angina to Prevent Recurrent Events*), publicado em 2001, que randomizou mais de 12.500 pacientes com síndrome coronariana aguda sem supradesnivelamento de ST (SCASSST) para clopidogrel + AAS versus placebo + AAS, tendo um desfecho composto de morte cardiovascular, IAM ou AVC. Os resultados mostraram benefício no braço do clopidogrel, porém a redução absoluta de risco foi de apenas 2,1%, com NNT modesto de 47,6.<sup>100</sup> Já no estudo COMMIT publicado em 2005, com mais de 45.000 pacientes, o clopidogrel mostrou redução de mortalidade por todas as causas e redução no desfecho composto de morte, IAM ou AVC, tornando assim a droga padrão associada ao AAS nas SCA.<sup>101</sup>

Posteriormente, o clopidogrel foi comparado com duas outras drogas: Prasugrel, no estudo TRITON TIMI 38 (*Trial to Assess Improvement in Therapeutic Outcomes by Optimizing Platelet Inhibition with Prasugrel-*

*Thrombolysis in Myocardial Infarction*), e Ticagrelor, no estudo PLATO (*PLATelet inhibition and patient Outcomes*). No primeiro estudo, publicado em 2007, com mais de 12.000 pacientes o Prasugrel mostrou benefício no desfecho composto de morte cardiovascular, IAM não fatal e AVC não fatal, porém sem demonstrar diminuição de mortalidade.<sup>102</sup> Já o Ticagrelor, no segundo estudo, publicado em 2009 com mais de 18.000 pacientes, mostrou redução de desfecho composto de morte vascular, IAM ou AVC. Houve redução de morte cardiovascular isoladamente, estatisticamente significativa. Esses estudos elevaram o nível de evidência dessas duas drogas nas diretrizes de IAM, porém o NNT do PLATO (52 para prevenção de evento cardiovascular e 91 para prevenir 1 óbito) mostrou um benefício de magnitude moderada.<sup>103</sup>

Outra grande questão é se a hipótese de revascularização percutânea completa no momento do infarto (ou nos primeiros dias posteriores ao IAM) traria mais benefício versus tratar apenas a artéria culpada, como era estabelecido nas diretrizes. Os estudos PRAMI (*Preventive Angioplasty in Acute Myocardial Infarction*), CvLPRIT (*Complete versus Lesion-only Primary PCI trial*), DANAMI-3 PRIMULTI (*Complete revascularisation versus treatment of the culprit lesion only in patients with ST-segment elevation myocardial infarction and multivessel disease*) e COMPARE-ACUTE (*Fractional Flow Reserve–Guided Multivessel Angioplasty in Myocardial Infarction*), mostraram evidências de benefício em tratar todas as artérias (que tivessem estenoses anatômica- ou fisiologicamente graves), em desfechos combinados, às custas de novas revascularizações.<sup>104-107</sup> Em 2019 foi publicado o estudo COMPLETE (*The Complete versus Culprit-Only Revascularization Strategies to Treat Multivessel Disease after Early PCI for STEMI*), ensaio clínico randomizado, multicêntrico, mostrou diminuição no desfecho composto de morte e IAM (às custas de mais IAM) no braço da revascularização completa. Apesar de não demonstrar diminuição de mortalidade, essa conduta modificou as diretrizes, dando uma classe de recomendação maior a favor da revascularização completa, provavelmente devido à quantidade de ensaios clínicos randomizados apontando nessa direção, e não à qualidade da evidência.<sup>108-109</sup>

O pensamento mecanicista nos inflige a afirmar que abrir uma artéria ocluída prolongadamente seria benéfico (reflexo óculo-estenótico). Essa questão foi estudada no OAT trial, que randomizou pacientes para ATC + TMO versus TMO em 72h até 28 dias após o IAM. Para o desfecho morte,

reinfarto e insuficiência cardíaca classe IV (NYHA) não houve diferença entre os grupos, mostrando que não há benefício em realizar ATC nessa população.<sup>110</sup>

## Prevenção primária e secundária com AAS e estatinas na doença arterial coronariana

O uso de AAS na prevenção secundária de eventos cardiovasculares é bem consolidado. Porém, na prevenção primária, seus benefícios não são bem estabelecidos. Em 2018 foram publicados três estudos que ajudaram a jogar luz sobre a tomada de decisão clínica: ASPREE, ARRIVE e ASCEND. O ASPREE (*Aspirin in Reducing Events in the Elderly*) foi um estudo que comparou AAS 100 mg versus placebo em cerca de 19.000 idosos com DAC. O desfecho primário era um composto de morte, demência ou incapacidade. Após 5 anos o estudo foi truncado por futilidade, o número de eventos foi muito baixo na população estudada, o resultado evidenciou que o AAS não demonstrou benefício no desfecho primário e aumentou significativamente sangramentos maiores.<sup>111</sup>

O estudo ARRIVE (*Use of aspirin to reduce risk of initial vascular events in patients at moderate risk of cardiovascular disease*) também comparou AAS 100 mg versus placebo em mais de 12.000 pacientes maiores de 55 anos sem DAC estabelecida, porém com fatores de risco. O desfecho combinado era morte cardiovascular, IAM, angina instável, AVC e AIT, e o braço do AAS não demonstrou redução do desfecho composto, porém aumentou sangramento gastrointestinal significativamente.<sup>112</sup>

O estudo ASCEND (*A Study of Cardiovascular Events in Diabetes*) randomizou mais de 15.000 pacientes com mais de 40 anos, diabéticos, sem DAC manifesta. Com desfecho composto de IAM, AVC, AIT ou morte decorrente de tais eventos, o grupo do AAS demonstrou redução de 12% no desfecho primário, porém com 29% mais risco de sangramento.<sup>113</sup> Com esses achados podemos afirmar que os escores de risco como Framingham e ASCVD (*atherosclerotic cardiovascular disease*) parecem superestimar o risco de evento cardiovascular e o uso preventivo de AAS tem pouco impacto na redução desses eventos.

A dislipidemia foi identificada como fator de risco para DAC após a publicação do Framingham Study em 1961.<sup>114</sup> As estatinas são drogas inibidoras da enzima HMG-CoA-redutase e foram utilizadas a partir dos



anos 80 como redutoras do colesterol LDL e nos anos 90 foram iniciados estudos para avaliar prevenção primária e secundária no contexto de DAC.

Na prevenção primária (pacientes sem DAC manifesta), o primeiro grande estudo publicado foi o WOSCOPS (*The West of Scotland Coronary Prevention Study*), publicado em 1995 com mais 6.500 pacientes randomizados para pravastatina ou placebo, havendo uma redução de risco absoluto de 2,4%, gerando um NNT de 41 para prevenir um evento coronariano em 5 anos, um efeito considerado modesto. Este estudo não demonstrou redução de mortalidade (não era o objetivo primário).<sup>115</sup> Em 1998 foi publicado o AFCAPS/TexCAPS (*The Air Force/Texas Coronary Atherosclerosis Prevention Study*), que avaliou mais de 6.000 pacientes (homens eram grande maioria) com colesterol alto e baixo risco cardiovascular, randomizados para lovastatina ou placebo. O resultado, evidenciou uma redução de risco absoluta de apenas 0,82% para o desfecho de IAM não fatal, angina instável ou morte súbita cardíaca, muito provavelmente devido ao fato de a população ser de baixo risco, gerando um NNT de 122 para prevenir um evento coronariano por ano.<sup>116</sup>

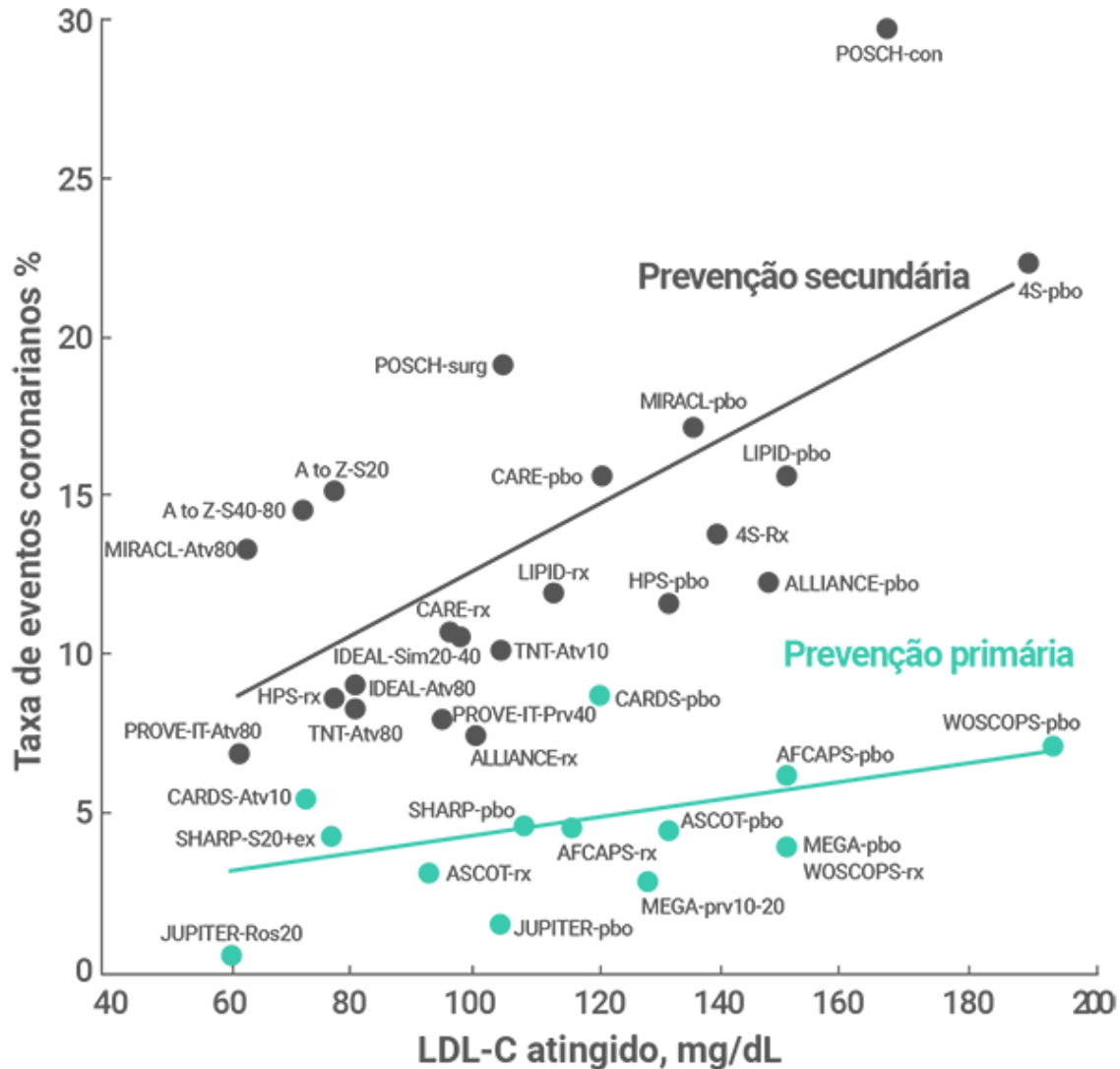
Dez anos após, o estudo JUPITER (*Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin*) foi publicado, tendo randomizado mais de 17.000 pacientes para uma estatina de alta potência, a rosuvastatina ou placebo. Os resultados evidenciaram que, para o desfecho primário de IAM não fatal, AVC não fatal, hospitalização por angina instável, revascularização ou morte por causas cardíacas, o grupo que utilizou foi significativamente menor. Nesse estudo, chama atenção o baixo número de eventos e o truncamento, realizado com menos de 500 eventos, que pode superestimar o efeito do tratamento (provavelmente influenciando na redução de mortalidade por todas as causas, que foi maior no grupo da estatina – efeito muito alardeado e pouco explicado). A redução absoluta de risco foi de 0,2%, muito baixa. O NNT calculado foi de 169 em um ano de tratamento para o desfecho composto primário de 500 para reduzir um evento coronariano.<sup>117</sup>

Devemos levar em consideração que morte não é especificamente um problema em uma população que tem colesterol alto e é assintomática, já que esses pacientes têm uma mortalidade muito baixa no momento do estudo e um grande estudo para ter poder de detectar benefício nesse desfecho teria que ter mais de 54.000 pacientes. Os estudos então deveriam ter focado em desfechos como IAM, AVC e angina, que compõem a história natural da

aterosclerose, além de análises de custo-efetividade, que são importantes para determinar se uma conduta pode ser empregada em determinado sistema de saúde.

Na prevenção secundária, o primeiro estudo publicado foi o 4S (*Scandinavian Simvastatin Survival Study*) em 1994, que avaliou sinvastatina versus placebo em 4.444 pacientes que tinha tido IAM ou angina estável, com desfecho primário de morte por todas as causas, e foi interrompido precocemente (gerando risco de superestimação do efeito do tratamento) devido a redução de risco absoluto de 3,3% na mortalidade por todas as causas com sinvastatina, gerando um NNT de 30 (para eventos coronarianos maiores o NNT foi de 15), mostrando grande benefício, sendo corroborado por estudos subsequentes como o CARE (*The Cholesterol And Recurrent Events*) e o HPS (*The Heart Protection Study*) (Figura 7).<sup>118-121</sup>

### **Figura 7.**



Nível de LDL-C atingido e taxa de eventos coronarianos em prevenção primária e secundária na DAC. LDL-C: lipoproteína de baixa densidade; DAC: doença arterial coronariana. Adaptado de: Cleveland Clinic Journal of Medicine March 2016;83(3):181-186

## Tratamento da hipertensão arterial sistêmica baseada em evidências

A hipertensão arterial sistêmica (HAS) é uma doença muito prevalente em nosso meio e seu tratamento sempre foi alvo de muitos ensaios clínicos, tendo muitas classes de drogas estudadas (algumas já estabelecidas há algumas décadas como betabloqueadores, diuréticos), além de medidas não farmacológicas (altamente sujeitas a vieses de confusão) como dieta e atividade física.

O estudo DASH (*Dietary Approaches to Stop Hypertension*) foi publicado em 1997 e avaliou os efeitos de uma dieta rica em vegetais, fibras e baixos níveis de gordura saturada na pressão arterial. Os pacientes foram randomizados para uma dieta habitual da população americana e a dieta DASH, após uma fase de 3 semanas utilizando uma dieta padrão e foram acompanhados por mais 8 semanas. A quantidade de sódio era similar nas duas dietas. Houve redução significativa da pressão arterial diastólica (desfecho primário) utilizando a dieta DASH em relação a uma dieta típica americana. A magnitude da dieta, apesar de modesta, foi similar à monoterapia medicamentosa em pacientes com HAS estágio I. Como o período de seguimento foi curto, não se pode avaliar os benefícios a médio e longo prazo.<sup>122</sup> Da mesma forma, o estudo PREDIMED (*Prevención con Dieta Mediterránea*) (retratado 5 anos após sua publicação devido a falhas na randomização) mostrou redução de eventos cardiovasculares maiores com a dieta mediterrânea.<sup>123</sup> A magnitude do efeito, porém, demonstrado nesses ensaios clínicos, é muito pequena, conforme mostra Ioannidis,<sup>124</sup> tornando os resultados desses ensaios de nutrição pouco verídicos.

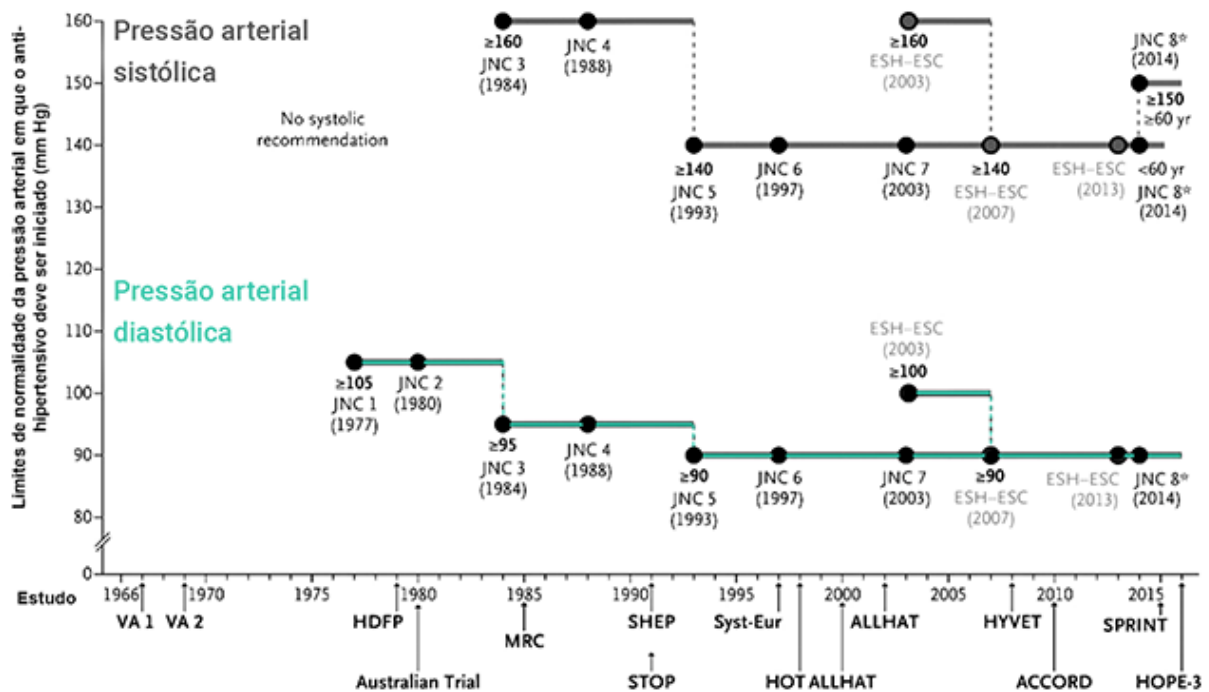
O estudo HOT (*Hypertension Optimal Treatment*), publicado em 1998, foi um dos ensaios que demonstrou que a redução intensiva da pressão arterial é associada com menor taxa de eventos cardiovasculares.<sup>125</sup> Em 2000 foi publicado o estudo HOPE (*The Heart Outcomes Prevention Evaluation*), que randomizou mais de 9.000 pacientes para ramipril ou placebo. O resultado mostrou benefício com redução significativa e clinicamente relevante de eventos cardiovasculares. Foi o primeiro estudo a mostrar evidências de benefício dos inibidores da ECA na prevenção de desfechos cardiovasculares.<sup>126</sup> Outro estudo que mostrou benefício com uma droga que age no sistema renina-angiotensina foi o LIFE (*Losartan Intervention For Endpoint reduction in hypertension study*), publicado em 2002. Nele, o bloqueador do receptor do angiotensina II (BRA), losartana versus o atenolol, demonstrou redução de AVC e DM2.<sup>127</sup>

O estudo VALUE (*The Valsartan Antihypertensive Long-term Use Evaluation*), que comparou o valsartana com anlodipino, não demonstrou diferenças significativas em relação a desfechos cardiovasculares.<sup>128</sup> Publicado em 2002, o estudo ALLHAT (*The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial*), comparou Lisinopril ou Anlodipino versus Clortalidona não demonstrando diferença entre os grupos no tocante ao desfecho composto de IAM não fatal e morte por DAC.<sup>129</sup> Em

2008 foi testada a hipótese se um duplo bloqueio no sistema renina-angiotensina traria mais benefício aos hipertensos no estudo ONTARGET (*Telmisartan, Ramipril, or Both in Patients at High Risk for Vascular Events*). Foram randomizados mais de 24.000 pacientes em três grupos: ramipril, telmisartan e combinação de ambos, mostrando não inferioridade do telmisartan frente ao ramipril e que a associação de ambos foi associada a mais eventos adversos.<sup>130</sup> No mesmo ano foi publicado o estudo ACCOMPLISH (*The Avoiding Cardiovascular Events through Combination Therapy in Patients Living with Systolic Hypertension*), que testou a combinação de benazepril e anlodipino versus benazepril e hidroclorotiazida. Para o desfecho primário de morte cardiovascular, IAM não fatal, AVC não fatal, angina instável, PCR ou revascularização, o grupo do benazepril e anlodipino mostrou redução do desfecho primário, porém às custas apenas de eventos cardiovasculares. Além disso o grupo comparativo poderia ter usado clortalidona que tem um perfil de tolerabilidade melhor e potência anti-hipertensiva maior.<sup>131</sup>

Analisando a intensidade de controle de pressão, foi testada a hipótese se a meta de pressão arterial sistólica mais baixa (< 120 mmHg) reduziria eventos clínicos comparado com a meta padrão (< 140 mmHg) em pacientes não diabéticos, no estudo SPRINT (*The Systolic Blood Pressure Intervention Trial*), publicado em 2015. O estudo foi truncado com 3 anos após análise interina mostrar superioridade do tratamento intensivo para o desfecho composto de IAM, SCA, AVC, IC descompensada ou morte cardiovascular. A magnitude do efeito foi baixa, com uma redução de risco absoluta de 1,6%, resultando em um impacto modesto (NNT de 68).<sup>132</sup> Outros estudos também mostraram pouco benefício do controle intensivo em diabéticos (estudo ACCORD – *Action to Control Cardiovascular Risk in Diabetes* – sem melhora em desfechos cardiovasculares) e em pacientes com AVC lacunar prévio, não encontrando redução no risco de AVC recorrente (estudo SPS3-BP - *Secondary Prevention of Small Subcortical Strokes-Blood Pressure*).<sup>133-134</sup> O estudo SPRINT deve ser considerado de forma Bayesiana. A pressão sistólica menor que 120 mmHg não é o valor “normal”, apenas esse foi ponto de corte estudado no ensaio, e não, de forma mecanicista, inferir que com uma pressão sistólica de 125 mmHg, por exemplo, o paciente seria hipertenso (Figura 8).<sup>135</sup>

### **Figura 8.**



Ensaio clínico que influenciaram os limites de normalidade da pressão arterial em que o anti-hipertensivo deve ser iniciado. ACCORD (Action to Control Cardiovascular Risk in Diabetics), ALLHAT (Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial), Australian Trial (the Australian Therapeutic Trial), ESH-ESC (the European Society of Hypertension-European Society of Cardiology Guidelines), HDFP (Hypertension Detection and Follow-up Program), HOPE (Heart Outcomes Prevention Evaluation), HOT (Hypertension Optimal Treatment), HYVET (Hypertension in the Very Elderly Trial), JNC (Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure), MRC (Medical Research Council), SHEP (Systolic Hypertension in the Elderly Program), SPRINT (Systolic Blood Pressure Intervention Trial), STOP (Swedish Trial in Old Patients with Hypertension), Syst-Eur (Systolic Hypertension in Europe), and VA (Veterans Administration Cooperative Study Group). Adaptado de: N Engl J Med 2016; 375:1756-1766.

## Conclusões

Conhecer as evidências científicas que baseiam o tratamento de pacientes cardiopatas permite uma maior probabilidade de benefício a esses doentes. Além disso, saber as particularidades dos estudos, suas indicações e metodologia, assim como interpretá-los, diminui a quantidade de vieses cognitivos, trazendo o que há de melhor para os que necessitam das terapias propostas. Praticar a Cardiologia baseada em evidências significa nortear suas ações a partir da ciência, tão vastamente produzida pela especialidade, além de identificar a incerteza como pedra fundamental do pensamento médico e, do ponto de vista “bayesiano”, usá-la a seu favor.

## Referências bibliográficas

1. Janse MJ, Rosen MR. History of arrhythmias. *Handb Exp Pharmacol*. 2006;(171):1-39.
2. Khush KK, Rapaport E, Waters D. The history of the coronary care unit. *Can J Cardiol*. 2005 Oct;21(12):1041-5.
3. Juul-Möller S, Lilja B, Johansson BW. Ventricular arrhythmias and left ventricular function: one-year follow-up after myocardial infarction. *Eur Heart J*. 1988 Nov;9(11):1181-7.
4. The Cardiac Arrhythmia Pilot Study. The CAPS investigators. *Am J Cardiol*. 1986 Jan;57(1):91-5.
5. Echt DS, Liebson PR, Mitchell LB, Peters RW, Obias-Manno D, Barker AH, Arensberg D, Baker A, Friedman L, Greene HL. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. *N Engl J Med*. 1991 Mar;324(12):781-8.
6. Cairns JA, Connolly SJ, Roberts R, Gent M. Randomised trial of outcome after myocardial infarction in patients with frequent or repetitive ventricular premature depolarisations: CAMIAT. Canadian Amiodarone Myocardial Infarction Arrhythmia Trial Investigators. *Lancet (London, England)*. 1997 Mar;349(9053):675-82.
7. Bardy GH, Lee KL, Mark DB, Poole JE, Packer DL, Boineau R, Domanski M, Troutman C, Anderson J, Johnson G, McNulty SE, Clapp-Channing N, Davidson-Ray LD, Fraulo ES, Fishbein DP, Luceri RM, Ip JH. Amiodarone or an implantable cardioverter-defibrillator for congestive heart failure. *N Engl J Med*. 2005 Jan;352(3):225-37.
8. Connolly SJ, Gent M, Roberts RS, Dorian P, Roy D, Sheldon RS, Mitchell LB, Green MS, Klein GJ, O'Brien B. Canadian implantable defibrillator study (CIDS) : a randomized trial of the implantable cardioverter defibrillator against amiodarone. *Circulation*. 2000 Mar;101(11):1297-302.
9. Kuck KH, Cappato R, Siebels J, Rüppel R. Randomized comparison of antiarrhythmic drug therapy with implantable defibrillators in patients resuscitated from cardiac arrest : the Cardiac Arrest Study Hamburg (CASH). *Circulation*. 2000 Aug;102(7):748-54.
10. A comparison of antiarrhythmic-drug therapy with implantable defibrillators in patients resuscitated from near-fatal ventricular arrhythmias. *N Engl J Med*. 1997 Nov;337(22):1576-83.
11. Connolly SJ, Hallstrom AP, Cappato R, Schron EB, Kuck KH, Zipes DP, Greene HL, Boczor S, Domanski M, Follmann D, Gent M, Roberts RS. Meta-analysis of the implantable cardioverter defibrillator secondary prevention trials. AVID, CASH and CIDS studies. Antiarrhythmics vs Implantable Defibrillator study. Cardiac Arrest Study Hamburg. Canadian Implantable Defibrillator Study. *Eur Heart J*. 2000 Dec;21(24):2071-8.
12. Moss AJ, Hall WJ, Cannom DS, Daubert JP, Higgins SL, Klein H, Levine JH, Saksena S, Waldo AL, Wilber D, Brown MW, Heo M. Improved survival with an implanted defibrillator in patients with coronary disease at high risk for ventricular arrhythmia. Multicenter

Automatic Defibrillator Implantation Trial Investigators. *N Engl J Med.* 1996 Dec;335(26):1933-40.

13. Moss AJ, Zareba W, Hall WJ, Klein H, Wilber DJ, Cannom DS, Daubert JP, Higgins SL, Brown MW, Andrews ML. Prophylactic implantation of a defibrillator in patients with myocardial infarction and reduced ejection fraction. *N Engl J Med.* 2002 Mar;346(12):877-83.
14. Køber L, Thune JJ, Nielsen JC, Haarbo J, Videbæk L, Korup E, Jensen G, Hildebrandt P, Steffensen FH, Bruun NE, Eiskjær H, Brandes A, Thøgersen AM, Gustafsson F, Egstrup K, Videbæk R, Hassager C, Svendsen JH, Høfsten DE, Torp-Pedersen C, Pehrson S. Defibrillator Implantation in Patients with Nonischemic Systolic Heart Failure. *N Engl J Med.* 2016 Sep;375(13):1221-30.
15. Jackson SL, Tong X, Yin X, George MG, Ritchey MD. Emergency Department, Hospital Inpatient, and Mortality Burden of Atrial Fibrillation in the United States, 2006 to 2014. *Am J Cardiol.* 2017 Dec;120(11):1966-73.
16. Wyse DG, Waldo AL, DiMarco JP, Domanski MJ, Rosenberg Y, Schron EB, Kellen JC, Greene HL, Mickel MC, Dalquist JE, Corley SD. A comparison of rate control and rhythm control in patients with atrial fibrillation. *N Engl J Med.* 2002 Dec;347(23):1825-33.
17. Van Gelder IC, Hagens VE, Bosker HA, Kingma JH, Kamp O, Kingma T, Said SA, Darmanata JI, Timmermans AJM, Tijssen JGP, Crijns HJGM. A comparison of rate control and rhythm control in patients with recurrent persistent atrial fibrillation. *N Engl J Med.* 2002 Dec;347(23):1834-40.
18. Andrade JG, Wells GA, Deyell MW, Bennett M, Essebag V, Champagne J, Roux J-F, Yung D, Skanes A, Khaykin Y, Morillo C, Jolly U, Novak P, Lockwood E, Amit G, Angaran P, Sapp J, Wardell S, Lauck S, Macle L, Verma A. Cryoablation or Drug Therapy for Initial Treatment of Atrial Fibrillation. *N Engl J Med.* 2021 Jan;384(4):305-15.
19. Marrouche NF, Brachmann J, Andresen D, Siebels J, Boersma L, Jordaens L, Merkely B, Pokushalov E, Sanders P, Proff J, Schunkert H, Christ H, Vogt J, Bänsch D. Catheter Ablation for Atrial Fibrillation with Heart Failure. *N Engl J Med.* 2018 Feb;378(5):417-27.
20. Packer DL, Mark DB, Robb RA, Monahan KH, Bahnson TD, Poole JE, Noseworthy PA, Rosenberg YD, Jeffries N, Mitchell LB, Flaker GC, Pokushalov E, Romanov A, Bunch TJ, Noelker G, Ardashev A, Revishvili A, Wilber DJ, Cappato R, Kuck K-H, Hindricks G, Davies DW, Kowey PR, Naccarelli G V, Reiffel JA, Piccini JP, Silverstein AP, Al-Khalidi HR, Lee KL. Effect of Catheter Ablation vs Antiarrhythmic Drug Therapy on Mortality, Stroke, Bleeding, and Cardiac Arrest Among Patients With Atrial Fibrillation: The CABANA Randomized Clinical Trial. *JAMA.* 2019 Apr;321(13):1261-74.
21. Kirchhof P, Camm AJ, Goette A, Brandes A, Eckardt L, Elvan A, Fetsch T, van Gelder IC, Haase D, Haegeli LM, Hamann F, Heidbüchel H, Hindricks G, Kautzner J, Kuck K-H, Mont L, Ng GA, Rekosz J, Schoen N, Schotten U, Suling A, Taggeselle J, Themistoclakis S, Vettorazzi E, Vardas P, Wegscheider K, Willems S, Crijns HJGM, Breithardt G. Early



- Rhythm-Control Therapy in Patients with Atrial Fibrillation. *N Engl J Med.* 2020 Oct;383(14):1305-16.
22. Tarasoutchi F, Montera MW, Ramos AI de O, Sampaio RO, Rosa VEE, Accorsi TAD, et al. Update of the Brazilian Guidelines for Valvular Heart Disease – 2020. *Arq Bras Cardiol.* 2020 Oct;115(4):720-75.
  23. Wilkins GT, Weyman AE, Abascal VM, Block PC, Palacios IF. Percutaneous balloon dilatation of the mitral valve: an analysis of echocardiographic variables related to outcome and the mechanism of dilatation. *Br Heart J.* 1988 Oct;60(4):299-308.
  24. ROWE JC, BLAND EF, SPRAGUE HB, WHITE PD. The course of mitral stenosis without surgery: ten- and twenty-year perspectives. *Ann Intern Med.* 1960 Apr;52:741-9.
  25. Ellis Lb, Harken De. Closed Valvuloplasty For Mitral Stenosis. A Twelve-Year Follow-Up Study Of 1571 Patients. *N Engl J Med.* 1964 Mar;270:643-50.
  26. Turi ZG, Reyes VP, Raju BS, Raju AR, Kumar DN, Rajagopal P, Sathyanarayana P V, Rao DP, Srinath K, Peters P. Percutaneous balloon versus surgical closed commissurotomy for mitral stenosis. A prospective, randomized trial. *Circulation.* 1991 Apr;83(4):1179-85.
  27. Reyes VP, Raju BS, Wynne J, Stephenson LW, Raju R, Fromm BS, Rajagopal P, Mehta P, Singh S, Rao DP. Percutaneous balloon valvuloplasty compared with open surgical commissurotomy for mitral stenosis. *N Engl J Med.* 1994 Oct;331(15):961-7.
  28. Meneguz-Moreno RA, Costa JRJ, Gomes NL, Braga SLN, Ramos AIO, Meneghelo Z, Maldonado M, Ferreira-Neto AN, Franca JID, Siqueira D, Esteves C, Sousa A, Sousa JE, Abizaid A. Very Long Term Follow-Up After Percutaneous Balloon Mitral Valvuloplasty. *JACC Cardiovasc Interv.* 2018 Oct;11(19):1945-52.
  29. Rosenhek R, Rader F, Klaar U, Gabriel H, Krejc M, Kalbeck D, Schemper M, Maurer G, Baumgartner H. Outcome of watchful waiting in asymptomatic severe mitral regurgitation. *Circulation.* 2006 May;113(18):2238-44.
  30. Kang D-H, Kim JH, Rim JH, Kim M-J, Yun S-C, Song J-M, Song H, Choi K-J, Song J-K, Lee J-W. Comparison of early surgery versus conventional treatment in asymptomatic severe mitral regurgitation. *Circulation.* 2009 Feb;119(6):797-804.
  31. Madesis A, Tsakiridis K, Zarogoulidis P, Katsikogiannis N, Machairiotis N, Kougioumtzi I, Kesisis G, Tsiouda T, Belevelis T, Koletas A, Zarogoulidis K. Review of mitral valve insufficiency: repair or replacement. *J Thorac Dis.* 2014 Mar;6 Suppl 1(Suppl 1):S39-51.
  32. Feldman T, Foster E, Glower DD, Kar S, Rinaldi MJ, Fail PS, Smalling RW, Siegel R, Rose GA, Engeron E, Loghin C, Trento A, Skipper ER, Fudge T, Letsou G V, Massaro JM, Mauri L. Percutaneous repair or surgery for mitral regurgitation. *N Engl J Med.* 2011 Apr;364(15):1395-406.
  33. Takagi H, Ando T, Umemoto T. A review of comparative studies of MitraClip versus surgical repair for mitral regurgitation. *Int J Cardiol.* 2017 Feb;228:289-94.
  34. Stone GW, Lindenfeld J, Abraham WT, Kar S, Lim DS, Mishell JM, Whisenant B, Grayburn PA, Rinaldi M, Kapadia SR, Rajagopal V, Sarembock IJ, Brieke A, Marx SO,

- Cohen DJ, Weissman NJ, Mack MJ. Transcatheter Mitral-Valve Repair in Patients with Heart Failure. *N Engl J Med*. 2018 Dec;379(24):2307-18.
35. Obadia J-F, Messika-Zeitoun D, Leurent G, Iung B, Bonnet G, Piriou N, et al. Percutaneous Repair or Medical Treatment for Secondary Mitral Regurgitation. *N Engl J Med*. 2018 Dec;379(24):2297-306.
  36. Kang D-H, Park S-J, Lee S-A, Lee S, Kim D-H, Kim H-K, et al. Early Surgery or Conservative Care for Asymptomatic Aortic Stenosis. *N Engl J Med*. 2020 Jan;382(2):111-9.
  37. Leon MB, Smith CR, Mack M, Miller DC, Moses JW, Svensson LG, et al. Transcatheter aortic-valve implantation for aortic stenosis in patients who cannot undergo surgery. *N Engl J Med*. 2010 Oct;363(17):1597-607.
  38. Smith CR, Leon MB, Mack MJ, Miller DC, Moses JW, Svensson LG, et al. Transcatheter versus surgical aortic-valve replacement in high-risk patients. *N Engl J Med*. 2011 Jun;364(23):2187-98.
  39. Leon MB, Smith CR, Mack MJ, Makkar RR, Svensson LG, Kodali SK, et al. Transcatheter or Surgical Aortic-Valve Replacement in Intermediate-Risk Patients. *N Engl J Med*. 2016 Apr;374(17):1609-20.
  40. Mack MJ, Leon MB, Thourani VH, Makkar R, Kodali SK, Russo M, et al. Transcatheter Aortic-Valve Replacement with a Balloon-Expandable Valve in Low-Risk Patients. *N Engl J Med*. 2019 May;380(18):1695-705.
  41. Akinseye OA, Pathak A, Ibebuogu UN. Aortic Valve Regurgitation: A Comprehensive Review. *Curr Probl Cardiol*. 2018 Aug;43(8):315-34.
  42. Dujardin KS, Enriquez-Sarano M, Schaff H V, Bailey KR, Seward JB, Tajik AJ. Mortality and morbidity of aortic regurgitation in clinical practice. A long-term follow-up study. *Circulation*. 1999 Apr;99(14):1851-7.
  43. Nishimura RA, Otto CM, Bonow RO, Carabello BA, Erwin JP 3rd, Fleisher LA, et al. 2017 AHA/ACC Focused Update of the 2014 AHA/ACC Guideline for the Management of Patients With Valvular Heart Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation*. 2017 Jun;135(25):e1159-95.
  44. Yusuf S, Pitt B, Davis CE, Hood WBJ, Cohn JN. Effect of enalapril on mortality and the development of heart failure in asymptomatic patients with reduced left ventricular ejection fractions. *N Engl J Med*. 1992 Sep;327(10):685-91.
  45. Pfeffer MA, McMurray JJ V, Velazquez EJ, Rouleau J-L, Køber L, Maggioni AP, et al. Valsartan, captopril, or both in myocardial infarction complicated by heart failure, left ventricular dysfunction, or both. *N Engl J Med*. 2003 Nov;349(20):1893-906.
  46. Dargie HJ. Effect of carvedilol on outcome after myocardial infarction in patients with left-ventricular dysfunction: the CAPRICORN randomised trial. *Lancet (London, England)*. 2001 May;357(9266):1385-90.

47. Packer M, Coats AJ, Fowler MB, Katus HA, Krum H, Mohacsi P, Rouleau JL, Tendera M, Castaigne A, Roecker EB, Schultz MK, DeMets DL. Effect of carvedilol on survival in severe chronic heart failure. *N Engl J Med*. 2001 May;344(22):1651-8.
48. Effect of metoprolol CR/XL in chronic heart failure: Metoprolol CR/XL Randomised Intervention Trial in Congestive Heart Failure (MERIT-HF). *Lancet* (London, England). 1999 Jun;353(9169):2001-7.
49. The Cardiac Insufficiency Bisoprolol Study II (CIBIS-II): a randomised trial. *Lancet* (London, England). 1999 Jan;353(9146):9-13.
50. Pitt B, Zannad F, Remme WJ, Cody R, Castaigne A, Perez A, Palensky J, Wittes J. The effect of spironolactone on morbidity and mortality in patients with severe heart failure. Randomized Aldactone Evaluation Study Investigators. *N Engl J Med*. 1999 Sep;341(10):709-17.
51. McMurray JJ V, Packer M, Desai AS, Gong J, Lefkowitz MP, Rizkala AR, et al. Angiotensin-neprilysin inhibition versus enalapril in heart failure. *N Engl J Med*. 2014 Sep;371(11):993-1004.
52. Wiviott SD, Raz I, Bonaca MP, Mosenzon O, Kato ET, Cahn A, et al. Dapagliflozin and Cardiovascular Outcomes in Type 2 Diabetes. *N Engl J Med*. 2019 Jan;380(4):347-57.
53. Neal B, Perkovic V, Mahaffey KW, de Zeeuw D, Fulcher G, Erondu N, et al. Canagliflozin and Cardiovascular and Renal Events in Type 2 Diabetes. *N Engl J Med*. 2017 Aug;377(7):644-57.
54. Zinman B, Wanner C, Lachin JM, Fitchett D, Bluhmki E, Hantel S, Mattheus M, Devins T, Johansen OE, Woerle HJ, Broedl UC, Inzucchi SE. Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. *N Engl J Med*. 2015 Nov;373(22):2117-28.
55. McMurray JJ V, Solomon SD, Inzucchi SE, Køber L, Kosiborod MN, Martinez FA, Pet al. Dapagliflozin in Patients with Heart Failure and Reduced Ejection Fraction. *N Engl J Med*. 2019 Nov;381(21):1995-2008.
56. Packer M, Anker SD, Butler J, Filippatos G, Pocock SJ, Carson P, et al. Cardiovascular and Renal Outcomes with Empagliflozin in Heart Failure. *N Engl J Med*. 2020 Oct;383(15):1413-24.
57. Heerspink HJL, Stefánsson B V, Correa-Rotter R, Chertow GM, Greene T, Hou F-F, et al. Dapagliflozin in Patients with Chronic Kidney Disease. *N Engl J Med*. 2020 Oct;383(15):1436-46.
58. Cowie MR, Fisher M. SGLT2 inhibitors: mechanisms of cardiovascular benefit beyond glycaemic control. *Nat Rev Cardiol*. 2020 Dec;17(12):761-72.
59. Julian D. Treatment of cardiac arrest in acute myocardial ischaemia and infarction. *Lancet*. 1961;278(7207):840-4.
60. Jones DS. CABG at 50 (or 107?) — The Complex Course of Therapeutic Innovation. *N Engl J Med*. 2017 May;376(19):1809-11.

61. Dallan LAO, Jatene FB. Revascularização miocárdica no século XXI. *Brazilian J Cardiovasc Surg.* 2013;28:137-44.
62. Grüntzig AR, Myler RK, Hanna ES TM. Coronary transluminal angioplasty (abstract). *Circulation.* 1977;55-56(III):84.
63. Grüntzig AR, Senning Å, Siegenthaler WE. Nonoperative Dilatation of Coronary-Artery Stenosis. *N Engl J Med.* 1979 Jul;301(2):61-8.
64. Fischman DL, Leon MB, Baim DS, Schatz RA, Savage MP, Penn I, et al. A Randomized Comparison of Coronary-Stent Placement and Balloon Angioplasty in the Treatment of Coronary Artery Disease. *N Engl J Med.* 1994 Aug;331(8):496-501.
65. Menichelli M, Parma A, Pucci E, Fiorilli R, De Felice F, Nazzaro M, et al. Randomized Trial of Sirolimus-Eluting Stent Versus Bare-Metal Stent in Acute Myocardial Infarction (SESAMI). *J Am Coll Cardiol.* 2007;49(19):1924-30.
66. Spaulding C, Henry P, Teiger E, Beatt K, Bramucci E, Carrié D, et al. Sirolimus-Eluting versus Uncoated Stents in Acute Myocardial Infarction. *N Engl J Med.* 2006 Sep;355(11):1093-104.
67. Hueb WA, Bellotti G, de Oliveira SA, Arie S, de Albuquerque CP, Jatene AD, Pileggi F. The Medicine, Angioplasty or Surgery Study (MASS): A prospective, randomized trial of medical therapy, balloon angioplasty or bypass surgery for single proximal left anterior descending artery stenoses. *J Am Coll Cardiol.* 1995;26(7):1600-5.
68. Hueb W, Soares PR, Gersh BJ, César LAM, Luz PL, Puig LB, Martinez EM, Oliveira SA, Ramires JAF. The medicine, angioplasty, or surgery study (MASS-II): a randomized, controlled clinical trial of three therapeutic strategies for multivessel coronary artery disease: One-year results. *J Am Coll Cardiol.* 2004;43(10):1743-51.
69. Whady H, Neuza L, J. GB, R. SP, E. RE, C. PA, Desiderio F, C. RAS, C. HA, A.F. RJ. Ten-Year Follow-Up Survival of the Medicine, Angioplasty, or Surgery Study (MASS II). *Circulation.* 2010 Sep;122(10):949-57.
70. A Randomized Trial of Therapies for Type 2 Diabetes and Coronary Artery Disease. *N Engl J Med.* 2009 Jun;360(24):2503-15.
71. Comparison of Coronary Bypass Surgery with Angioplasty in Patients with Multivessel Disease. *N Engl J Med.* 1996 Jul;335(4):217-25.
72. Boden WE, O'Rourke RA, Teo KK, Hartigan PM, Maron DJ, Kostuk WJ, et al. Optimal Medical Therapy with or without PCI for Stable Coronary Disease. *N Engl J Med.* 2007 Apr;356(15):1503-16.
73. Velazquez EJ, Lee KL, Deja MA, Jain A, Sopko G, Marchenko A, et al. Coronary-Artery Bypass Surgery in Patients with Left Ventricular Dysfunction. *N Engl J Med.* 2011 Apr;364(17):1607-16.
74. Velazquez EJ, Lee KL, Jones RH, Al-Khalidi HR, Hill JA, Panza JA, et al. Coronary-Artery Bypass Surgery in Patients with Ischemic Cardiomyopathy. *N Engl J Med.* 2016 Apr;374(16):1511-20.

75. Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. *BMJ*. 1999 Dec;319(7223):1492 LP – 1495.
76. Serruys PW, Morice M-C, Kappetein AP, Colombo A, Holmes DR, Mack MJ, et al. Percutaneous Coronary Intervention versus Coronary-Artery Bypass Grafting for Severe Coronary Artery Disease. *N Engl J Med*. 2009 Mar;360(10):961-72.
77. Farkouh ME, Domanski M, Sleeper LA, Siami FS, Dangas G, Mack M, et al. Strategies for Multivessel Revascularization in Patients with Diabetes. *N Engl J Med*. 2012 Nov;367(25):2375-84.
78. Stone GW, Sabik JF, Serruys PW, Simonton CA, Génèreux P, Puskas J, et al. Everolimus-Eluting Stents or Bypass Surgery for Left Main Coronary Artery Disease. *N Engl J Med*. 2016 Oct;375(23):2223-35.
79. Mäkikallio T, Holm NR, Lindsay M, Spence MS, Erglis A, Menown IBA, et al. Percutaneous coronary angioplasty versus coronary artery bypass grafting in treatment of unprotected left main stenosis (NOBLE): a prospective, randomised, open-label, non-inferiority trial. *Lancet*. 2016 Dec;388(10061):2743-52.
80. Al-Lamee R, Thompson D, Dehbi H-M, Sen S, Tang K, Davies J, et al. Percutaneous coronary intervention in stable angina (ORBITA): a double-blind, randomised controlled trial. *Lancet*. 2018 Jan;391(10115):31-40.
81. De Bruyne B, Pijls NHJ, Kalesan B, Barbato E, Tonino PAL, Piroth Z, et al. Fractional Flow Reserve–Guided PCI versus Medical Therapy in Stable Coronary Disease. *N Engl J Med*. 2012 Aug;367(11):991-1001.
82. Maron DJ, Hochman JS, Reynolds HR, Bangalore S, O’Brien SM, Boden WE, et al. Initial Invasive or Conservative Strategy for Stable Coronary Disease. *N Engl J Med*. 2020 Mar;382(15):1395-407.
83. Correia LCL, Rassi Junior A, Correia LCL, Rassi Junior A. Downstream Change of the Primary Endpoint in the ISCHEMIA Trial: the Elephant in the Room. *Arq Bras Cardiol*. 2018;111(2):213-4.
84. Grines CL, Browne KF, Marco J, Rothbaum D, Stone GW, O’Keefe J, et al. A Comparison of Immediate Angioplasty with Thrombolytic Therapy for Acute Myocardial Infarction. *N Engl J Med*. 1993 Mar;328(10):673-9.
85. Randomised trial of intravenous atenolol among 16 027 cases of suspected acute myocardial infarction: ISIS-1. *Lancet*. 1986 Jul;328(8498):57-66.
86. Group. TMTR. Metoprolol in acute myocardial infarction (MIAMI). A randomised placebo-controlled international trial. *Eur Hear J*. 1985;6:199-226.
87. Lewis HD, Davis JW, Archibald DG, Steinke WE, Smitherman TC, Doherty JE, et al. Protective Effects of Aspirin against Acute Myocardial Infarction and Death in Men with Unstable Angina. *N Engl J Med*. 1983 Aug;309(7):396-403.
88. Cairns JA, Gent M, Singer J, Finnie KJ, Froggatt GM, Holder DA, et al. Aspirin, Sulfinpyrazone, or Both in Unstable Angina. *N Engl J Med*. 1985 Nov;313(22):1369-75.

89. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17 187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet*. 1988 Aug;332(8607):349-60.
90. Miocardico Giperlosdn. Medical Science: GISSI-2: A factorial randomised trial of alteplase versus streptokinase and heparin versus no heparin among 12 490 patients with acute myocardial infarction. *Lancet*. 1990 Jul;336(8707):65-71.
91. An International Randomized Trial Comparing Four Thrombolytic Strategies for Acute Myocardial Infarction. *N Engl J Med*. 1993 Sep;329(10):673-82.
92. Van de Werf F, Cannon CP, Luyten A, Houbracken K, McCabe CH, Beroli S, Bluhmki E, Sarelin H, Wang-Clow F, Fox NL, Braunwald E. Safety assessment of single-bolus administration of TNK tissue-plasminogen activator in acute myocardial infarction: The ASSENT-1 trial. *Am Heart J*. 1999;137(5):786-91.
93. Primary versus tenecteplase-facilitated percutaneous coronary intervention in patients with ST-segment elevation acute myocardial infarction (ASSENT-4 PCI): randomised trial. *Lancet*. 2006 Feb;367(9510):569-78.
94. Ibanez B, James S, Agewall S, Antunes MJ, Bucciarelli-Ducci C, Bueno H, et al. 2017 ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: The Task Force for the management of acute myocardial infarction in patients presenting with ST-segment elevation of the European Socie. *Eur Heart J*. 2018 Jan;39(2):119-77.
95. Thérroux P, Ouimet H, McCans J, Latour J-G, Joly P, Lévy G, Pelletier E, Juneau M, Stasiak J, deGuise P, Pelletier GB, Rinzler D, Waters DD. Aspirin, Heparin, or Both to Treat Acute Unstable Angina. *N Engl J Med*. 1988 Oct;319(17):1105-11.
96. Cohen M, Demers C, Gurfinkel EP, Turpie AGG, Fromell GJ, Goodman S, et al. A Comparison of Low-Molecular-Weight Heparin with Unfractionated Heparin for Unstable Coronary Artery Disease. *N Engl J Med*. 1997 Aug;337(7):447-52.
97. Antman EM. TIMI 11B. Enoxaparin versus unfractionated heparin for unstable angina or non-Q-wave myocardial infarction: A double-blind, placebo-controlled, parallel-group, multicenter trial. Rationale, study design, and methods. *Am Heart J*. 1998;135(6, Supplement 2):S353-60.
98. Werner K, Arnd B, E. HS, Scott M, Ginés S, G. TAG, Jan van der M, Eric O, Sven U, Karin L. Comparison of Low-Molecular-Weight Heparin With Unfractionated Heparin Acutely and With Placebo for 6 Weeks in the Management of Unstable Coronary Artery Disease. *Circulation*. 1997 Jul;96(1):61-8.
99. Group TFISS. Comparison of two treatment durations (6 days and 14 days) of a low molecular weight heparin with a 6-day treatment of unfractionated heparin in the initial management of unstable angina or non-Q wave myocardial infarction: FRAX.I.S. (FRAxiparine in Ischa. *Eur Heart J*. 1999 Nov;20(21):1553-62.
100. Effects of Clopidogrel in Addition to Aspirin in Patients with Acute Coronary Syndromes without ST-Segment Elevation. *N Engl J Med*. 2001 Aug;345(7):494-502.

101. Chen Z, Jiang L. Addition of clopidogrel to aspirin in 45 852 patients with acute myocardial infarction: Randomised placebo-controlled trial. *Lancet*. 2005 Nov;366(9497):1607-21.
102. Wiviott SD, Braunwald E, McCabe CH, Montalescot G, Ruzyllo W, Gottlieb S, Neumann F-J, Ardissino D, De Servi S, Murphy SA, Riesmeyer J, Weerakkody G, Gibson CM, Antman EM. Prasugrel versus Clopidogrel in Patients with Acute Coronary Syndromes. *N Engl J Med*. 2007 Nov;357(20):2001-15.
103. Wallentin L, Becker RC, Budaj A, Cannon CP, Emanuelsson H, Held C, Horrow J, Husted S, James S, Katus H, Mahaffey KW, Scirica BM, Skene A, Steg PG, Storey RF, Harrington RA. Ticagrelor versus Clopidogrel in Patients with Acute Coronary Syndromes. *N Engl J Med*. 2009 Sep;361(11):1045-57.
104. Wald DS, Morris JK, Wald NJ, Chase AJ, Edwards RJ, Hughes LO, Berry C, Oldroyd KG. Randomized Trial of Preventive Angioplasty in Myocardial Infarction. *N Engl J Med*. 2013 Sep;369(12):1115-23.
105. Gershlick AH, Khan JN, Kelly DJ, Greenwood JP, Sasikaran T, Curzen N, et al. Randomized Trial of Complete Versus Lesion-Only Revascularization in Patients Undergoing Primary Percutaneous Coronary Intervention for STEMI and Multivessel Disease: The CvLPRIT Trial. *J Am Coll Cardiol*. 2015;65(10):963-72.
106. Engstrøm T, Kelbæk H, Helqvist S, Høfsten DE, Kløvgård L, Holmvang L, et al. Complete revascularisation versus treatment of the culprit lesion only in patients with ST-segment elevation myocardial infarction and multivessel disease (DANAMI-3 PRIMULTI): an open-label, randomised controlled trial. *Lancet*. 2015 Aug;386(9994):665-71.
107. Smits PC, Abdel-Wahab M, Neumann F-J, Boxma-de Klerk BM, Lunde K, Schotborgh CE, et al. Fractional Flow Reserve–Guided Multivessel Angioplasty in Myocardial Infarction. *N Engl J Med*. 2017 Mar;376(13):1234–44.
108. Mehta SR, Wood DA, Storey RF, Mehran R, Bainey KR, Nguyen H, et al. Complete Revascularization with Multivessel PCI for Myocardial Infarction. *N Engl J Med*. 2019 Sep;381(15):1411-21.
109. López-Sendón J, Moreno R. COMPLETE revascularization after STEMI? Sure, go ahead. *REC Interv Cardiol* (English Ed. 2020 Feb);
110. S. HJ, R. RH, Vladimír D, E. BC, Witold R, P. SZ, P. MA, C. CA, M. RJ, D. WH, Suzanne G, A. FS, B. MD, A. LG. Long-Term Effects of Percutaneous Coronary Intervention of the Totally Occluded Infarct-Related Artery in the Subacute Phase After Myocardial Infarction. *Circulation*. 2011 Nov;124(21):2320-8.
111. McNeil JJ, Nelson MR, Woods RL, Lockery JE, Wolfe R, Reid CM, et al. Effect of Aspirin on All-Cause Mortality in the Healthy Elderly. *N Engl J Med*. 2018 Sep;379(16):1519-28.
112. Gaziano JM, Brotans C, Coppolecchia R, Cricelli C, Darius H, Gorelick PB, et al. Use of aspirin to reduce risk of initial vascular events in patients at moderate risk of cardiovascular

- disease (ARRIVE): a randomised, double-blind, placebo-controlled trial. *Lancet*. 2018 Sep;392(10152):1036-46.
113. Effects of Aspirin for Primary Prevention in Persons with Diabetes Mellitus. *N Engl J Med*. 2018 Aug;379(16):1529-39.
  114. Dawber TR, Meadors GF, Moore FE J. Epidemiological approaches to heart disease: the Framingham Study. *Am J Public Heal*. 1951;41(3):279-86.
  115. Shepherd J, Cobbe SM, Ford I, Isles CG, Lorimer AR, Macfarlane PW, McKillop JH, Packard CJ. Prevention of Coronary Heart Disease with Pravastatin in Men with Hypercholesterolemia. *N Engl J Med*. 1995 Nov;333(20):1301-8.
  116. Downs JR, Clearfield M, Weis S, Whitney E, Shapiro DR, Beere PA, Langendorfer A, Stein EA, Kruyer W, Gotto Antonio M. J, Group for the AR. Primary Prevention of Acute Coronary Events With Lovastatin in Men and Women With Average Cholesterol Levels Results of AFCAPS/TexCAPS. *JAMA*. 1998 May;279(20):1615-22.
  117. Ridker PM, Danielson E, Fonseca FAH, Genest J, Gotto AM, Kastelein JJP, et al. Rosuvastatin to Prevent Vascular Events in Men and Women with Elevated C-Reactive Protein. *N Engl J Med*. 2008 Nov;359(21):2195-207.
  118. Group SSSS. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). *Lancet*. 1994 Nov;344(8934):1383-9.
  119. Sacks FM, Pfeffer MA, Moye LA, Rouleau JL, Rutherford JD, Cole TG, Bet al. The Effect of Pravastatin on Coronary Events after Myocardial Infarction in Patients with Average Cholesterol Levels. *N Engl J Med*. 1996 Oct;335(14):1001-9.
  120. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomised placebocontrolled trial. *Lancet*. 2002 Jul;360(9326):7-22.
  121. Sabatine MS. Advances in the treatment of dyslipidemia. *Cleve Clin J Med*. 2016 Mar;83(3):181 LP – 186.
  122. Appel LJ, Moore TJ, Obarzanek E, Vollmer WM, Svetkey LP, Sacks FM, et al. A Clinical Trial of the Effects of Dietary Patterns on Blood Pressure. *N Engl J Med*. 1997 Apr;336(16):1117-24.
  123. Estruch R, Ros E, Salas-Salvadó J, Covas M-I, Corella D, Arós F, et al. Primary Prevention of Cardiovascular Disease with a Mediterranean Diet. *N Engl J Med*. 2013 Feb;368(14):1279-90.
  124. Siontis GCM, Ioannidis JPA. Risk factors and interventions with statistically significant tiny effects. *Int J Epidemiol*. 2011 Oct;40(5):1292-307.
  125. Hansson L, Zanchetti A, Carruthers SG, Dahlöf B, Elmfeldt D, Julius S, et al. Effects of intensive blood-pressure lowering and low-dose aspirin in patients with hypertension: principal results of the Hypertension Optimal Treatment (HOT) randomised trial. *Lancet*. 1998 Jun;351(9118):1755-62.



126. Effects of an Angiotensin-Converting–Enzyme Inhibitor, Ramipril, on Cardiovascular Events in High-Risk Patients. *N Engl J Med.* 2000 Jan;342(3):145-53.
127. Dahlöf B, Devereux RB, Kjeldsen SE, Julius S, Beevers G, de Faire U, et al. Cardiovascular morbidity and mortality in the Losartan Intervention For Endpoint reduction in hypertension study (LIFE): a randomised trial against atenolol. *Lancet.* 2002 Mar;359(9311):995-1003.
128. Julius S, Kjeldsen SE, Weber M, Brunner HR, Ekman S, Hansson L, et al. Outcomes in hypertensive patients at high cardiovascular risk treated with regimens based on valsartan or amlodipine: the VALUE randomised trial. *Lancet.* 2004 Jun;363(9426):2022-31.
129. Group TAO and C for the ACR. Major Outcomes in High-Risk Hypertensive Patients Randomized to Angiotensin-Converting Enzyme Inhibitor or Calcium Channel Blocker vs DiureticThe Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). *JAMA.* 2002 Dec;288(23):2981-97.
130. Telmisartan, Ramipril, or Both in Patients at High Risk for Vascular Events. *N Engl J Med.* 2008 Apr;358(15):1547-59.
131. Jamerson K, Weber MA, Bakris GL, Dahlöf B, Pitt B, Shi V, Hester A, et al. Benazepril plus Amlodipine or Hydrochlorothiazide for Hypertension in High-Risk Patients. *N Engl J Med.* 2008 Dec;359(23):2417-28.
132. A Randomized Trial of Intensive versus Standard Blood-Pressure Control. *N Engl J Med.* 2015 Nov;373(22):2103-16.
133. Effects of Intensive Blood-Pressure Control in Type 2 Diabetes Mellitus. *N Engl J Med.* 2010 Apr;362(17):1575-85.
134. Blood-pressure targets in patients with recent lacunar stroke: the SPS3 randomised trial. *Lancet.* 2013 Aug;382(9891):507-15.
135. Pfeffer MA, McMurray JJ V. Lessons in Uncertainty and Humility — Clinical Trials Involving Hypertension. *N Engl J Med.* 2016 Nov;375(18):1756-66.

# 10. ENDOCRINOLOGIA BASEADA EM EVIDÊNCIAS

## Introdução

A Endocrinologia é a especialidade que estuda e trata doenças de origens hormonais, que podem ser causadas por falta ou excesso de hormônios. A especialidade incorpora também a metabologia, que estuda doenças relacionadas ao metabolismo energético, como obesidade, diabetes e hiperlipidemia, condições muito prevalentes atualmente.

Na endocrinologia clássica, deficiências hormonais absolutas, na grande maioria das vezes, são tratadas com a reposição do hormônio deficitário e, em algumas situações (como a insuficiência adrenal, diabetes tipo 1 e hipotireoidismo franco), o benefício clínico da intervenção é tão grande que prescinde da necessidade de estudos randomizados. O mesmo ocorre em situações de excesso hormonal causado por tumores, como a síndrome de Cushing (excesso de cortisol) e o feocromocitoma (excesso de hormônios catecolaminérgicos), em que pode inclusive haver um risco iminente à vida. Há, porém, situações mais limítrofes, em que estudos mais aprofundados são necessários. Neste grupo que requer estudos específicos para manejo terapêutico estão deficiências hormonais parciais como o hipotireoidismo subclínico (que será abordado nesse capítulo, por ser uma condição comum) e a deficiência de GH em adultos (mais incomum). Cabe lembrar que é qualquer tipo de distúrbio endocrinológico tem espaço para estudos clínicos. Ainda que não seja sobre indicação de tratamento em si, a modalidade terapêutica é um tema clinicamente estudado no tratamento do hipertireoidismo (radioiodoterapia vs. medicação) ou no hiperaldosteronismo primário (cirurgia vs. antagonista mineralocorticoide).

Em relação a neoplasias, o câncer de glândula endocrinológica mais comum é o de tireoide, que, felizmente, tem mortalidade baixa. Assim, discutir, com base em evidências, se há necessidade de rastreamento desse câncer na população geral é fundamental e também será um dos temas abordados.

Doenças crônico-metabólicas, como obesidade e diabetes tipo 2, são associadas a maior risco de diversas complicações de saúde e mortalidade. Assim, é fundamental termos estudos bem conduzidos que nos guiem em

relação a melhor forma de gerar impacto real à longo prazo. Ambas as doenças (assim como o diabetes tipo 1) também serão brevemente abordadas, à luz das melhores evidências atuais.

Nosso objetivo ao longo deste capítulo é discutir a forma como a Medicina baseada em evidências é aplicada em doenças prevalentes dentro da nossa especialidade. Assim, optamos por explorar o hipotireoidismo subclínico, a ultrassonografia como rastreo para câncer de tireoide, a obesidade, o diabetes mellitus e a síndrome metabólica.

## Hipotireoidismo subclínico

O diagnóstico de hipotireoidismo subclínico é bioquímico e definido como aumento de TSH e T4 livre normal por pelo menos 2-3 meses. Apesar de o nome sugerir que sintomas sejam levados em consideração, este não é o caso. O limiar de TSH a partir do qual se deve iniciar o tratamento de hipotireoidismo subclínico é bastante controverso<sup>1</sup> e vamos discutir sobre algumas das principais evidências a seguir. É importante destacar, antes de prosseguir a discussão, que o TSH tende a subir com a idade e isso deve ser levado em consideração ao interpretar exame de pacientes idosos.<sup>2</sup>

### TSH elevado – uma entidade que pode entrar em remissão

Ao identificar um TSH elevado com T4 livre normal, o primeiro passo é repetir o exame após algum tempo, uma vez que o TSH pode estar alterado transitoriamente. Abaixo discutiremos sobre duas grandes análises para ilustrar este ponto.

Em uma grande coorte nos EUA, um subgrupo de 3.594 idosos sem hipotireoidismo conhecido tiveram TSH aferido e 459 (12,8%) tiveram o diagnóstico de hipotireoidismo subclínico. Após 2 anos, 369 dos 459 idosos com TSH inicialmente elevado tiveram TSH recoletados (perdas por morte, falta na consulta ou problema na amostra) e, desses, 56% se mantiveram com hipotireoidismo subclínico, 35% reverteram para eutireoidismo (TSH normal), 7% estavam usando levotiroxina e 2% ficaram hipotireoideos (TSH elevado e T4 livre baixo). Mais de 90% dos pacientes com TSH inicialmente alterados mantiveram-se com hipotireoidismo subclínico ou tornaram-se eutireoideos. É importante lembrar que, apesar de

categorizarmos os diagnósticos em “sim” ou “não”, o TSH é uma variável contínua. Aqueles que tinham TSH mais elevado tiveram maior chance de progredir para hipotireoidismo.<sup>3</sup>

Em uma análise ainda maior, envolvendo um banco de dados de 2,3 milhões de adultos em Israel, 422.242 indivíduos sem doença tireoidiana conhecida e sem usar levotiroxina tinham o TSH aferido. Desses pacientes, 95% tinham TSH normal e 3,7% tinham TSH elevado. Dos com TSH elevado, 93% tinham T4 livre normal. Dos pacientes com TSH elevado na primeira medida 62,1% tiveram TSH normal em uma segunda medida.<sup>4</sup>

## Hipotireoidismo subclínico confirmado: vale a pena tratar?

Uma vez detectada que a alteração de TSH não é transitória e que se trata de fato de hipotireoidismo subclínico, a dúvida do(a) médico(a) é se o tratamento (levotiroxina) deve ser iniciado. Discorreremos abaixo sobre este tema, mas adiantamos que os motivos para indicar tratamento são 1) melhora de qualidade de vida e/ou 2) diminuição do risco cardiovascular. Começaremos discutindo estudos referentes a risco cardiovascular.

Um interessante estudo de coorte conectando hipotireoidismo subclínico a doença cardiovascular envolve dados de clínicas privadas em Taiwan em 1998 e 1999 com seguimento até 2008. Nesses 10 anos de seguimento, 3.669 dos 115.476 pacientes morreram, sendo 680 por doença cardiovascular. Houve associação de hipotireoidismo subclínico com mortalidade para todas as causas e mortalidade cardiovascular. As análises foram corrigidas para sexo, idade, tabagismo, uso de álcool, atividade física, renda, nível educacional, hipertensão arterial sistêmica, diabetes mellitus, dislipidemia e até noz de Areca, um alimento local cujo consumo está associado a doença cardiovascular e, portanto, pode ser um confundidor nessa análise.<sup>5</sup>

Interessante colocar que uma meta-análise de artigos publicados entre 1950 e 2010, envolvendo um total de 55.287 pacientes e publicada antes do estudo descrito no parágrafo anterior, já havia estabelecido associação de doença cardíaca coronariana com TSH mais alto, especialmente TSH  $\geq 10$  mUI/L (valor de referência aproximado de 0,5 a 4,5 mUI/L).<sup>6</sup>

Uma vez que múltiplos estudos de coorte demonstraram relação entre hipotireoidismo subclínico e doença cardiovascular, algumas novas dúvidas

aparecem. Essa relação pode ser explicada por algum confundidor, pode ser consequência de alterações metabólicas do hipotireoidismo subclínico (dislipidemia, hipertensão) ou pode ser um efeito direto do hipotireoidismo subclínico no sistema cardiovascular. Essas 3 hipóteses não são excludentes e podem acontecer simultaneamente.

Em 2017 uma meta-análise mostrou que o tratamento com levotiroxina reduz a espessura da da camada íntima-média da carótida em pacientes com hipotireoidismo subclínico<sup>7</sup>. Apesar da intuição inicial ser creditar a redução da espessura da carótida à levotiroxina, é importante considerar diversos fatores ao ler este tipo de estudo, sendo os mais importantes 1) os indivíduos são comparados com eles mesmos antes do tratamento, sem um grupo controle que não tenha tratamento e 2) a diminuição da espessura da carótida pode acontecer por melhora de parâmetros metabólicos, como colesterol total, colesterol LDL, pressão arterial com uso de levotiroxina e não por efeito direto do hormônio tireoidiano nos vasos sanguíneos. Finalmente, é importante lembrar que a diminuição da espessura da camada íntima-média da carótida é um marcador substituto para doença cardiovascular e isso não significa automaticamente que uma espessura se menor se traduza em risco cardiovascular menor.

Um outro estudo que corrobora para a possibilidade de redução de risco cardiovascular ao se usar levotiroxina em pacientes com hipotireoidismo subclínico é uma análise observacional de pessoas entre 40-70 anos. Nesta análise, aqueles tratados tiveram menos doença coronariana isquêmica na análise ajustada para sexo, idade e TSH basal (HR: 0,61, IC 95%: 0,39-0,95). Para os idosos com mais de 70 anos, não foi observado esse efeito protetor da levotiroxina para doença coronariana isquêmica quando ajustado para sexo, idade e TSH basal (HR: 0,99, IC 95% 0,59-1,33).<sup>8</sup>

Existem, portanto, alguns indícios de que o tratamento com levotiroxina pode reduzir o risco cardiovascular em pacientes com hipotireoidismo subclínico, especialmente os mais jovens. É importante lembrar que é pouco provável que um ensaio clínico randomizado para avaliar o impacto do tratamento nesses pacientes mais jovens seja desenhado, já que a baixa incidência de eventos cardiovasculares nessa população torna esse método de estudo longo e caro. Por sua vez, em populações selecionadas, os ensaios clínicos são factíveis e discutiremos a seguir dois deles que avaliam questões cardiovasculares.

Um ensaio clínico randomizou pacientes com TSH persistentemente  $> 4$  mUI/L logo após infarto agudo do miocárdio para usar levotiroxina ou placebo. O desfecho primário foi fração de ejeção (FE) de ventrículo esquerdo após 52 semanas, aferida através de ressonância magnética. A randomização foi estratificada por tipo de infarto (com ou sem supradesnivelamento de segmento ST) e o objetivo do tratamento foi manter TSH entre 0,4 e 2,5 mU/L no grupo tratado com a droga ativa. Os médicos eram cegos e a dose foi ajustada por médicos independentes que recomendavam qual caixa de remédio deveria ser fornecida aos pacientes. Esse ensaio clínico foi desenhado com 90% de poder de captar uma diferença absoluta de FE de 3%, que é considerada clinicamente relevante. A diferença ajustada de FE foi de 0,76% (IC 95% : -0,93 a 2,43), o que, portanto, não permitiu descartar a hipótese nula – tratamento com levotiroxina não melhorou a FE<sup>9</sup>.

Outro estudo avaliou função cardíaca randomizando 185 pacientes idosos ( $\geq 65$  anos) com hipotireoidismo subclínico para levotiroxina ou placebo tendo como desfecho primário função cardíaca. A mediana de seguimento foi 18,5 meses e também não houve diferença significativa de FE: 0,4% (IC 95% -1,8% a 2,5%)<sup>10</sup>.

Feita uma ilustração da dúvida referente a tratamento de hipotireoidismo subclínico e questões cardiovasculares, outro ponto a se explorar é o efeito do tratamento na qualidade de vida dos pacientes. Uma vez que sintomas de hipotireoidismo são inespecíficos e pouco comuns em pacientes com hipotireoidismo subclínico, medir possíveis benefícios em qualidade de vida com tratamento com levotiroxina não é tarefa simples. Abaixo discorreremos sobre dois ensaios clínicos randomizados que se propuseram a fazer essa avaliação.

Um pequeno estudo randomizado placebo controlado e duplo cego (n=60) teve como desfecho primário a melhora de uma escala de sintomas depressivos (escala de Beck) em pacientes com hipotireoidismo subclínico. Os pacientes recrutados eram jovens, em torno dos 35 anos e, após tratamento com levotiroxina por 12 semanas houve uma melhora na pontuação da escala de Beck, de uma média de 16,79 para 12,37 no grupo levotiroxina (p:0,04). No grupo controle a média inicial era de 13,77 e após uso de placebo caiu para 10,71 (p:0,16). Apesar dos autores concluírem que houve uma melhora de sintomas depressivos com tratamento ativo, não concordamos. Para se ter essa conclusão, seria necessário estabelecer que a

diferença de pontuação entre o basal e após tratamento é significativamente maior no grupo levotiroxina. Importante, essa diferença tem que ser não só estatisticamente significativa, mas atingir um valor mínimo que sabidamente traga benefícios clínicos.<sup>11</sup>

O estudo de maior alcance (publicado no NEJM) em relação à qualidade de vida no hipotireoidismo é o TRUST, um ensaio clínico com 737 idosos ( $\geq 65$  anos) com hipotireoidismo subclínico persistente que foram randomizados para placebo ou levotiroxina com a meta de manter TSH dentro do valor de referência (0,4 a 4,59 mUI/L) por 12 meses. A dose foi ajustada por um sistema computadorizado, tanto para levotiroxina quanto para placebo, sem que médico ou paciente soubessem o que era ajustado de fato. O ensaio clínico foi inicialmente desenhado para ter 2 desfechos primários – eventos cardiovasculares e escala de qualidade de vida especificamente desenhada para hipotireoidismo. Ao longo do estudo os autores mudaram o protocolo ao perceber que não havia poder para detectar diferença em eventos cardiovasculares. O TSH médio inicial era de 6,4 mUI/L e após 12 meses era 5,48 mUI/L no grupo placebo e 3,63 mUI/L no grupo tratado com levotiroxina. As escalas apresentaram valores praticamente idênticos em ambos os grupos ( $p: 0,99$ ). Também não houve aumento de efeitos adversos relacionados a excesso de hormônio tireoidiano<sup>12</sup>. Importante notar neste estudo que o TSH médio inicial era 6,4 mUI/L e, portanto, pouco elevado. Também relevante que a população alvo foram idosos, um grupo cujo TSH tende mesmo a ser mais alto.

A interpretação de estudos envolvendo hipotireoidismo subclínico requer uma série de cuidados. Não à toa existe controvérsia em relação ao tratamento desta entidade.

Enumeramos abaixo alguns fatores importantes a se levar em consideração nos estudos acima mencionados:

1. Dados observacionais são sujeitos a muito confundidores. Apesar do ajuste para confundidores ser factível, é importante lembrar que, possivelmente, há confundidores desconhecidos.
2. Para aumentar a validade interna, os ensaios clínicos selecionam populações restritas, e dessa forma a generalização dos dados fica comprometida.
3. Há mais evidência da ausência de benefício do tratamento em idosos. Faltam ensaios clínicos randomizados em jovens. Um dos

motivos para haver menos estudos em jovens é a dificuldade em se ter eventos, especialmente cardiovasculares, o que torna estudos clínicos randomizados longos e caros.

## Ultrassonografia como rastreamento de câncer de tireoide

A ultrassonografia de tireoide em adultos assintomáticos é frequentemente realizada na prática clínica. Em sua última edição, publicada em 2017, a *US Preventive Task Force* contraindica a triagem de câncer de tireoide na população geral e discute que os riscos dessa prática superam os benefícios.<sup>13</sup> É importante lembrar que o tripé da decisão terapêutica vem de um balanço entre benefício, malefício e custo. Não existe procedimento, exame, medicação ou qualquer tipo de ato médico que escape a essa ponderação.

Essa decisão da *US Preventive Task Force* não tem como base ensaios clínicos randomizados, mas sim um conjunto de dezenas de trabalhos clínicos com outros métodos. A seguir discutiremos dois desses trabalhos, com o intuito de ilustrar como é feita a interpretação e conseqüentemente tomada de decisões de acordo com os dados disponíveis. Importante lembrar que, baseado em estudos de coortes e análises de autópsia, é sabido de longa data que a grande maioria dos cânceres de tireoide são do tipo papilífero, cujo crescimento é lento e indolente com sobrevida livre de doença muito alta, acima de 90%.

A incidência de câncer de tireoide triplicou da década de 1970 para os anos 2000. O aumento mais expressivo aconteceu na década de 1990, justamente quando a ultrassonografia e a punção aspirativa por agulha fina tornaram-se mais populares. Assim, é muito provável que vivamos uma epidemia de diagnóstico de câncer de tireoide e não uma epidemia de câncer de tireoide em si. A evolução das últimas décadas mostra que a quase totalidade de incremento diagnóstico se deve justamente ao diagnóstico de carcinomas papilíferos, os mais comuns e mais indolentes que acometem a glândula. Nessa mesma análise observacional também chama a atenção que a mortalidade pela doença vem se mantendo estável e em torno de 1/200.000 pessoas/ano ao longo das últimas décadas. É esperado que, com aumento do diagnóstico de doença que pode levar à morte, o conseqüente aumento de tratamento reduza a mortalidade, mas a mortalidade global do câncer de tireoide não mudou.<sup>14</sup> Ainda que os



resultados aqui descritos até agora corroborem para argumentar que o diagnóstico de câncer de tireoide não tenha benefício na mortalidade, é mais uma vez importante lembrar que essa análise é observacional e sujeita a diversos vieses. Nesse caso a análise é populacional e não individualizada, trata-se, portanto, de um estudo ecológico. Apenas para ilustrar um viés possível, existe a possibilidade hipotética de os cânceres papilíferos de tireoide, por qualquer motivo, terem se tornado mais agressivos ao longo do tempo, e, assim o maior alcance diagnóstico e consequente tratamento destes cânceres evitaram um aumento de mortalidade que ocorreria, o que justificaria portanto a estabilidade de mortalidade sem tirar a importância do rastreio.

Outra análise didática é a de um centro no Japão, que em 1993, começou a oferecer a pacientes com diagnóstico de carcinoma papilífero de tireoide  $\leq$  1 cm sem metástase regional, a possibilidade de cirurgia ou observação. Aqueles que optaram pela observação foram submetidos a ultrassonografia de tireoide periódica. Um total de 340 pacientes optaram pela observação enquanto 1.055 optaram pela cirurgia. Durante o período de observação (mediana de 74 meses), apenas 31 (9,1%) pacientes do grupo observação apresentaram aumento do câncer significativo ( $\geq$  3 mm) e apenas 7 (2,1%) apresentaram metástase linfonodal. Dos 340 pacientes, apenas 109 (32%) passaram por tireoidectomia (diversos motivos, a depender da decisão do paciente e do médico que estava tratando). Ao se comparar desfechos pós-operatórios daqueles pacientes que foram submetidos a cirurgia ao diagnóstico com aqueles submetidos a cirurgia após observação, a taxa de metástase linfonodal pós-cirúrgica ao longo do tempo foi a mesma.<sup>15</sup>

Uma vez que há estudos, além dos dois aqui expostos obviamente, sugerindo não haver benefício no diagnóstico precoce de câncer de tireoide (especialmente o carcinoma papilífero), a pergunta que se segue é sobre possíveis malefícios. A investigação de câncer de tireoide por si só leva a custos e estresse emocional, como acontece em qualquer investigação de neoplasia. Há, porém, outras complicações do rastreio e consequente tratamento de câncer de tireoide. O hipotireoidismo é uma consequência óbvia da tireoidectomia total, mas as complicações mais temidas são lesão permanente do nervo laríngeo recorrente (2%) e hipoparatiroidismo permanente (5%), que levam respectivamente a alteração da patência da voz e hipocalcemia.<sup>13</sup> O hipoparatiroidismo requer tratamento contínuo, com

diversas doses de medicações diárias e monitoramento laboratorial frequente.

## Obesidade

A obesidade é uma doença crônica associada, em estudos observacionais, a maior morbidade, incapacidade e mortalidade. Epidemiologicamente, define-se obesidade como um Índice de Massa Corporal (IMC, peso/altura ao quadrado) acima de 30 kg/m<sup>2</sup>. Embora muitas críticas existam ao IMC em nível individual (pois alterações em composição corporal e distribuição de gordura façam com que indivíduos com IMCs semelhantes tenham riscos de saúde muito diferentes), populacionalmente essa é considerada uma ferramenta útil.<sup>16</sup>

Estudos com seguimento longo tendem a demonstrar aumento de riscos com IMCs acima de 25 kg/m<sup>2</sup> (sobrepeso), e forte associação com diversas doenças crônicas (mais de 200), como diabetes tipo 2, doenças cardiovasculares e uma série de cânceres.<sup>16-17</sup>

Como todo dado epidemiológico, essa relação da obesidade com mortalidade é sujeita a vieses e eventualmente, estudos questionam essa relação.<sup>18</sup> Entre os dois maiores vieses, estão:

1. Doenças crônicas podem levar a perda de peso e assim “mascarar” o efeito do IMC. Para reduzir esse viés, o ideal é que estudos que avaliam essa associação tenham seguimento suficientemente longo, o que nem sempre é o caso.<sup>19</sup>
2. Tabagismo, um fator de risco importante para uma série de doenças, está associado a IMC mais baixo e quanto maior o grau de tabagismo, menor o IMC. Assim, alguns estudos fazem ajustes para essa condição.

Quando esses vieses são levados em consideração, em geral, vê-se uma relação clara entre obesidade e risco de mortalidade e outras doenças.<sup>17, 19-20</sup>

No entanto, a evidência de que a perda de peso reduza doenças, risco cardiovascular e mortalidade é menos clara na literatura, como veremos a seguir.

## Perda de peso e redução de riscos de saúde

A evidência de que a perda de peso reduza risco cardiovascular é escassa na literatura e talvez a principal razão para isso seja: **perder peso com estratégias comportamentais e manter é muito difícil.**<sup>21</sup>

A literatura é vasta em estudos randomizados avaliando resposta ponderal a diversas intervenções dietéticas e baseadas em exercício físico. Os estudos são muito variados e heterogêneos, mas meta-análises sugerem que a média de perda de peso com intervenções de mudança de estilo de vida é de aproximadamente 3 kg.<sup>22</sup> Até hoje, os estudos com tempos de duração acima de 6 meses também demonstram que não existe superioridade de uma dieta específica sobre outra, e a adesão ao tratamento é fundamental.<sup>23-24</sup>

A perda de peso com exercício físico, sem dieta, tende a ser ainda menor, de menos de 1 kg.<sup>25</sup> Apesar disso, é sabido que a variabilidade de resposta clínica é muito grande. Há, sim, um número de indivíduos capazes de perder quantidades grandes de peso com mudanças de estilo de vida (conhecidos como respondedores), mas, em estudos clínicos, o número de pacientes que atingem 10% do peso perdido em geral é ao redor de 10% (dependendo da intervenção e do tempo de seguimento).<sup>26</sup> O objetivo de um tratamento clínico é gerar perdas de peso nessa faixa, pois perdas maiores são raras e difíceis de serem sustentadas. Além disso, 5%-10% de perda do peso já é capaz de reduzir o risco de diversas doenças crônicas e marcadores intermediários.<sup>27-28</sup>

O melhor estudo clínico que avaliou o impacto de mudança de estilo de vida focando perda de peso para avaliar redução de risco cardiovascular chama-se LOOK AHEAD e foi realizado em pacientes com diabetes mellitus do tipo 2.<sup>29</sup> Após 9 anos, o estudo foi interrompido por futilidade, isto é, não houve redução do risco cardiovascular com essa estratégia. A perda de peso entre o grupo-intervenção (em que houve contatos frequentes com médicos, profissionais de educação física e nutricionistas, além da possibilidade de uso de substitutos de refeição e orlistate) e o grupo-controle, que apenas recebeu orientações vagas, foi de menos de 3% no final dos 9 anos, embora a perda de peso inicial tenha sido grande.<sup>29</sup>

De forma interessante, uma análise *post-hoc* desse estudo, que avaliou apenas os pacientes que perderam mais do que 10% do peso no primeiro ano, demonstrou redução de cerca de 20% do risco cardiovascular.<sup>30</sup> Como uma análise *post-hoc* é uma evidência muito mais baixa, esse dado não confirma o benefício da perda de peso em redução de riscos, porém levanta

uma hipótese de alta plausibilidade. Infelizmente, porém, é muito difícil fazer um estudo dedicado que confirme essa teoria pois, como a perda de peso entre indivíduos é muito variada, é difícil garantir que um grupo perca e o outro não perca peso. Mesmo avaliando intervenções mais eficientes, como medicações ou cirurgia, é em geral muito difícil separar se o efeito da intervenção é ou não é peso-dependente. Até o momento, no entanto, não há evidências que nenhuma medicação para obesidade reduza risco cardiovascular. A liraglutida demonstrou redução de risco cardiovascular em população com diabetes e alto risco, mas não foi feito um estudo dedicado a essa avaliação em obesidade.<sup>31</sup>

Em relação à cirurgia bariátrica, no entanto, temos evidências de estudos observacionais prospectivos que há redução de mortalidade, aumento da expectativa de vida e redução de doenças associadas.<sup>32-33</sup> Dados de estudos randomizados em cirurgia são mais escassos, pois não é possível cegamento e o número de pacientes estudados em geral é pequeno para uma avaliação de desfechos duros no longo prazo. Assim o estudo com mais tempo de duração, o SOS, foi um estudo observacional pois não era considerado ético, à época, randomizar pacientes para a cirurgia, que era considerada um procedimento de alto risco. Apesar do grupo cirúrgico ter obesidade mais grave e mais fatores de risco, houve redução da mortalidade total e aumento de expectativa de vida, com análises de até 24 anos de seguimento.<sup>33</sup>

Há evidências também de redução de complicações crônicas do diabetes com a cirurgia e inclusive de mortalidade de câncer em mulheres.<sup>32</sup> Dessa forma, a cirurgia bariátrica é considerada um procedimento eficaz e seguro quando os critérios de elegibilidade são preenchidos (IMC acima de 40 kg/m<sup>2</sup> ou IMC acima de 35kg/m<sup>2</sup> com comorbidades associadas à obesidade).

## Diabetes mellitus

O diabetes mellitus é uma doença definida pelos altos níveis de glicose no sangue. Esse aumento pode ter causas variadas, sendo as mais comuns o diabetes tipo 2, caracterizado pela resistência à insulina, e o diabetes tipo 1, caracterizado pela não produção de insulina pelo pâncreas.<sup>34</sup> Assim, ambos têm apresentações clínicas, tratamento e prognóstico muito diferentes entre

si. Cabe aqui uma breve introdução geral e depois algumas considerações específicas para cada uma dessas duas condições.

O grande marco histórico no tratamento do diabetes foi a descoberta da insulina, em 1921.<sup>35</sup> Até então, pouco se sabia sobre a fisiopatologia da doença, mas sabia-se que crianças ou adultos jovens que desenvolviam um tipo grave de diabetes morriam em poucas semanas de inanição e desidratação. O uso da insulina em uma criança com essa condição permitiu uma melhora clínica rápida posterior e muito se acreditou que havia se descoberto a cura da doença. De fato, o prognóstico do que hoje conhecemos como diabetes tipo 1, por deficiência de insulina, mudou de forma abrupta e radical, de uma doença fatal em semanas a uma doença controlável a longo prazo. Nesse caso, devido à magnitude do efeito da insulina, não foi necessário (e nem seria ético), a necessidade de qualquer estudo de intervenção clínica randomizado.

Porém, logo se percebeu que esse tratamento:

1. Tinha menos impacto em pessoas com diabetes mais tardio, associado ao ganho de peso.
2. O uso de insulina evitava a inanição e desidratação, mas eram necessárias múltiplas doses para manter glicemias próximas ao normal e havia risco de hipoglicemia no caso de doses excessivas.
3. Embora o prognóstico clínico desses pacientes tivesse melhorado substancialmente, viu-se que, após alguns anos ou décadas de tratamento, os pacientes passaram a apresentar outras complicações da doença, principalmente doenças da retina, rins, neuropatias e complicações cardiovasculares.

Logo dúvidas emergiram: será que o controle estrito da glicose reduziria o risco de complicações do diabetes ou essa seria a história natural da doença, relacionada à genética ou características intrínsecas da doença?

Apenas na década de 1980 se iniciaram *trials* para responder a essas questões, quando já se sabia separar claramente as diferenças dos dois tipos de diabetes.

No caso do diabetes tipo 1, em 1982 se iniciou um ensaio clínico randomizado chamado DCCT, de 10 anos de duração (e que ainda seguiu por mais tempo com dados observacionais posteriores) que se propôs a avaliar se o controle estrito da glicemia era capaz de reduzir complicações

do diabetes, comparando um esquema com bomba de insulina ou mínimo de três aplicações diárias versus o tratamento “convencional” com uma ou duas doses de insulina.<sup>36</sup>

O DCCT provou que tratar intensivamente o diabetes tipo 1 reduziu, após 10 anos, o risco de retinopatia em 76%, de albuminúria em 54% (um desfecho substituto) e de neuropatia em 60%. Dados posteriores confirmaram que a redução de albuminúria foi também associada à redução de risco de desfechos renais nessa população e a presença da “memória metabólica”: isto é, após o final do estudo, mesmo com ambos os grupos recebendo o tratamento intensivo (que passou a ser considerado o “padrão-ouro”), ainda assim o grupo que anos antes recebeu o tratamento convencional teve mais novas complicações.<sup>36-37</sup> A partir desse ponto, o tratamento do DM1 evoluiu com tentativas de se criar insulinas mais fisiológicas, que reduzam o risco de hipoglicemias, melhorem o controle glicêmico e a qualidade de vida.<sup>38</sup> Nenhuma medicação oral até o momento mostrou um benefício claro, embora estudos recentes sugeriram que inibidores de SGLT2 possam ajudar na redução da HbGlicada, no peso e nas doses insulina.<sup>39</sup> O grande problema dessa classe de drogas é o risco elevado de cetoacidose, que é uma complicação grave do DM1 e, por isso, agências regulatórias mundiais se dividem sobre a aprovação ou não dessa opção terapêutica.

Também na década de 1980, as mesmas dúvidas que existiam sobre o tratamento do DM1 existiam em relação ao diabetes tipo 2. Nessa doença, a fisiopatologia não é de deficiência absoluta de insulina, mas relativa, ou seja, uma produção de insulina insuficiente para contrabalançar o aumento da resistência à insulina causado, principalmente, pelo acúmulo de gordura em região visceral. O tratamento com insulina, embora reduza glicemia, é menos fisiológico.<sup>40</sup> Esses pacientes em geral têm um risco cardiovascular bastante elevado (sendo a principal causa de morte nessa população) além dos riscos de doenças microvasculares.<sup>41</sup> Assim, medicações orais que reduzem a resistência (como a metformina) ou aumentam a secreção de insulina (sulfonilureias) eram utilizadas para tratá-la, mas pouco se sabia sobre os efeitos a longo prazo.

Nesse cenário foi iniciado o famoso estudo UKPDS, que assim como o DCCT, tinha o objetivo de avaliar se o tratamento intensivo do DM2 reduzia complicações e avaliou basicamente pessoas com diagnóstico muito recente de DM2.<sup>42</sup> O desenho do estudo é relativamente complicado, mas

ele avaliava se a terapia com metformina, insulina ou sulfoniluréia (escolhidos de forma não randomizada) com objetivo de manter glicemia de jejum abaixo de 108 mg/dL era superior a um tratamento “convencional” em que a meta glicêmica, à base de dieta, era uma glicemia abaixo de 270mg/dL.<sup>43</sup> Caso essas metas não fossem atingidas, era possível escalonar o tratamento. Esse estudo nos ensinou (e ainda nos ensina pois o seguimento de longo prazo continua) muito sobre a história natural do diabetes.

No UKPDS houve redução de doença microvascular geral naqueles no grupo com medicação e objetivo de glicemia mais baixa em 25%, que atingiu o limiar de significância estatística. No entanto, não houve significância na redução de eventos cardiovasculares. Em uma análise mais exploratória, ao se observar apenas os dados com a metformina (e não com insulina e sulfoniluréias), houve uma redução significativa, de 32% em desfechos gerais relacionados ao diabetes; 42% para mortalidade por diabetes e 36% para mortalidade total. Vale ressaltar que o n estudado foi pequeno (apenas 342 pacientes receberam metformina) e que, como descrito anteriormente, a alocação para diferentes opções medicamentosas não foi randomizada.<sup>44</sup>

Quando analisados os dados após 10 anos do fim do *trial*, houve uma leve mudança nos resultados: o grupo sulfoniluréia/insulina manteve a redução de doença microvascular (24%), mas houve também diferença estatisticamente significativa na incidência de infarto do miocárdio (15%) e morte por todas as causas (13%). No grupo metformina, que já havia atingido significância originalmente, os dados se mantiveram, com um a redução de IAM em 33% e morte por todas as causas de 27%.<sup>45</sup> O interessante é notar que, ao fim do *trial*, os grupos não foram mais randomizados e houve uma tendência a uma equivalência nos níveis de HbGlicada. Assim, surgiu o conceito de **legado metabólico**, isto é: o controle glicêmico intensivo no início do diagnóstico do diabetes leva a menor complicações crônicas a longo prazo, mesmo que, após um período de tempo, o controle glicêmico esteja semelhante ao de um grupo que não controlou tão bem no início. Um outro achado do UKPDS foi que os grupos de tratamento clínico tiveram piora progressiva da glicemia com o passar dos anos, necessitando cada vez mais novas medicações e eventualmente insulina,<sup>44</sup> sugerindo que o diabetes seria uma doença progressiva.

Outros estudos posteriores também avaliaram se o controle glicêmico intensivo reduzia as complicações do diabetes tipo 2. Alguns dos icônicos estudos nesta linha são o ADVANCE, ACCORD e VADT,<sup>46-48</sup> que se propuseram a avaliar pacientes com diabetes de longa data, muitos anos após o diagnóstico. Eles novamente mostraram que o controle glicêmico reduz complicações microvasculares, mas não houve redução de doença cardiovascular ou mortalidade. Mais surpreendentemente foi o dado de que no estudo ACCORD, o controle glicêmico intensivo (HbA1c<6,0%) aumentou a mortalidade.<sup>47</sup> Esse dado é ainda muito debatido na literatura, mas é possível que tenha relação com o risco de hipoglicemia causado por um controle glicêmico agressivo em uma população mais idosa e frágil.

Um outro dado que surpreendeu a comunidade médica foi uma metanálise que sugeriu que a rosiglitazona, uma medicação antidiabética cujo o mecanismo de ação é reduzir a resistência à insulina, poderia aumentar o risco cardiovascular.<sup>49</sup> Após muitas controvérsias, a rosiglitazona acabou proibida em vários países, e esse dado, com o ACCORD, levou à uma mudança da regulação de novas medicações pelas agências regulatórias, que passaram a demandar estudos de segurança cardiovascular de qualquer medicação nova que fosse aprovada.<sup>50</sup>

Assim, vê-se que, embora os pacientes com diabetes tenham como maior causa de mortalidade doenças cardiovasculares, simplesmente reduzir a glicemia tem um impacto limitado em reduzir esses eventos cardiovasculares. No entanto, controlar, com a glicemia, a pressão arterial e o colesterol (estatinas) tem um impacto muito maior: o estudo STENO 2, com uma intervenção multifatorial, mostrou redução de 46% na mortalidade, o que demonstra que o tratamento do DM2 não deve ser **glicocêntrico** (ou seja, preocupado apenas com a glicemia) e sim centrado na redução de diversos fatores de risco.<sup>51</sup>

No entanto, os novos estudos cardiovasculares demandados pelas agências regulatórias acabaram levando a uma nova era no tratamento do diabetes tipo 2, ao demonstrar que algumas classes de medicações para diabetes eram capazes de reduzir desfechos de forma **independente do controle glicêmico**, ao menos em pacientes com doença cardiovascular estabelecida.

As duas classes de medicações que demonstraram esse benefício foram os inibidores de SGLT2 e os análogos de GLP-1, embora haja algumas diferenças entre diferentes moléculas da mesma classe, o que torna difícil



saber o que é o efeito de uma medicação específica e o que é um efeito de classe.<sup>52</sup>

De forma geral, dois inibidores de SGLT2 (a empagliflozina e a canagliflozina) demonstraram, em estudos e populações diferentes, mas com maioria da população com doença cardiovascular estabelecida, redução de desfecho primário cardiovascular, ao redor de 30%. Houve redução de hospitalização por insuficiência cardíaca, redução de desfechos renais e, no estudo com a empagliflozina, houve redução de mortalidade por todas as causas, um desfecho difícil de ser atingido em estudos clínicos, ainda mais em uma população que já recebe diversos outros tratamentos com proteção estabelecida (como estatinas, anti-hipertensivos e antiagregantes).<sup>53-54</sup> Já a dapagliflozina estudou uma população basicamente de prevenção primária e demonstrou, assim como as duas outras, redução de hospitalização por insuficiência cardíaca e desfechos renais, sem atingir significância para o desfecho cardiovascular primário ou morte.<sup>55</sup> Os bons resultados desses estudos levaram inclusive à realização de estudos dedicados à insuficiência cardíaca e insuficiência renal (mesmo em indivíduos sem diabetes), com resultados também muito bons, mas que não são o foco da discussão deste capítulo.

A outra classe de medicação que demonstrou redução de eventos cardiovasculares foi a dos análogos de GLP-1. Aqui, com mais medicações estudadas, houve maior heterogeneidade dos resultados, mas, a grosso modo, a liraglutida demonstrou redução de eventos cardiovasculares em população de prevenção secundária.<sup>31</sup> Um dado semelhante foi visto com a semaglutida, porém com um estudo menor, não pré-especificado para superioridade.<sup>56</sup> A dulaglutida demonstrou também redução de eventos em população majoritariamente de prevenção primária, mas alto risco cardiovascular.<sup>57</sup> A redução de risco gira em torno de 12%-15% e acredita-se que a maior redução se dê por um efeito aterosclerótico e talvez por redução de AVC.

De toda forma, a robustez de estudos recentes levou a uma importante mudança nas diretrizes de tratamento de diabetes tipo 2, que anteriormente consideravam várias classes de medicações como equivalentes desde que fossem capazes de reduzir glicemia. Atualmente é recomendado que indivíduos com história de doença cardiovascular recebam ou um inibidor de SGLT2 ou um análogo de GLP-1. Caso o paciente tenha doença renal ou insuficiência cardíaca, um inibidor de SGLT2 é recomendado.<sup>58-59</sup>

Em conclusão, portanto, o tratamento do diabetes mudou consideravelmente nos 100 anos desde o descobrimento da insulina, muito por base de estudos randomizados. No DM1, a insulina permanece como o tratamento obrigatório, mas o tratamento intensivo com múltiplas doses de insulina, que busca formas mais semelhantes à secreção fisiológica de insulina é associado a redução de riscos (tecnologias mais modernas, como bombas de insulina tem também suas indicações para reduzir hipoglicemia, melhorar o controle glicêmico e a qualidade de vida, mas ainda são de difícil acesso para a maior parte da população). No DM2, sabemos que o controle glicêmico deve ser acompanhado de controle de outros fatores de risco e que, em pacientes de alto risco ou condições específicas, há medicamentos que levam a redução de desfechos duros independentemente de seu papel na glicemia.

## Pré-diabetes e síndrome metabólica

Existe muita discussão sobre o conceito de “pré-diabetes” e de síndrome metabólica. O que eles significam? Ambos os conceitos tem pontos em comum, mas também há pontos diferentes.

A ideia do pré-diabetes é reconhecer que existem pessoas cujo risco de desenvolver diabetes é maior que outros (aqueles com glicemia entre 100-125 mg/dL, com hemoglobina glicada entre 5,7%-6,4% ou uma glicemia duas horas após teste de tolerância oral à glicose entre 140-199).<sup>60</sup> Mas devemos tratar pré-diabetes?

Bons estudos randomizados, como o Diabetes Prevention Program (DPP), demonstraram que a modificação de estilo de vida (MEV), com enfoque em dieta hipocalórica e exercícios para perda de cerca de 7% do peso foi capaz de reduzir o risco de diabetes em 58% comparado ao placebo num período de seguimento de 2,8 anos.<sup>61</sup> No mesmo estudo, a metformina, uma medicação conhecida para tratamento do diabetes propriamente dito, reduziu esses riscos em 31%. As incidências de diabetes foram de 11/100 pessoas-ano no grupo placebo; 7,8/100 pessoas-ano no grupo metformina e 4,8 no grupo MEV. No entanto, não há evidências de longo prazo (após o final do estudo ativo) do estudo DPP no intuito de demonstrar que qualquer uma das intervenções reduza o risco de futuras possíveis complicações do diabetes, o que faz com a necessidade ou não de intervenção seja controversa.<sup>62</sup> No entanto, entender que cada quilo de peso perdido reduz o

risco de diabetes em 17%, e que mais de 60% do efeito benéfico da metformina sobre a redução desse risco também foi devido ao peso, pode permitir ao médico e paciente decidirem sobre a necessidade ou não de intervenção e qual seja esta.<sup>63</sup> A metformina, que é muito utilizada no tratamento do pré-diabetes na prática clínica, mostrou menor benefício em indivíduos idosos e maior em pessoas jovens, com obesidade e também mulheres com história de diabetes gestacional.<sup>63</sup> Outras medicações para perda de peso já demonstraram também redução do risco do diabetes, provavelmente por efeito peso-dependente, mas é importante entender que, caso interrompida, há risco de ganho de peso e retorno aos risco mais alto de desenvolver diabetes.<sup>64</sup>

A síndrome metabólica foi definida como uma doença associada ao acúmulo de gordura visceral e resistência à insulina, que aumenta o risco não só de diabetes, mas de doenças cardiovasculares.<sup>65</sup> Vários critérios já foram propostos, levando em conta circunferência de abdômen, glicemia de jejum ou outro parâmetro glicêmico, triglicérides, HDL-colesterol e pressão arterial. Embora pessoas que tenham alteração em algum desses parâmetros tenha risco maior de alterações nos outros, não há evidências de que diagnosticar alguém com síndrome metabólica (uma análise dicotômica) tenha mais utilidade do que avaliar cada um dos fatores que a compõe separadamente. Cada um desses parâmetros pode ser avaliado separadamente e alocado em tabelas de risco cardiovascular (uma análise contínua).<sup>65</sup> Assim, embora a ideia de que a gordura abdominal e a resistência à insulina levem a outras alterações que aumentem riscos cardiovasculares seja verdadeira, e que a perda de peso pode ajudar na melhora desse quadro, não há utilidade clínica claramente demonstrada de se fazer o diagnóstico de síndrome metabólica dentro de um consultório. Avaliar todos os fatores de risco e usar calculadoras de risco é uma estratégia que tende a ser mais eficiente pelo uso de variáveis contínuas, ao invés de dicotômicas.

## Conclusões

Apesar da grande quantidade de conhecimento acumulado ao longo das últimas décadas, o limiar para o qual se indica tratamento do hipotireoidismo subclínico continua controverso, já que o benefício dessa

intervenção ainda não é claro. O rastreio do câncer de tireoide em população geral, por sua vez, é cada vez mais desencorajado.

Embora a obesidade e o diabetes sejam claramente associados a maiores riscos de complicações de saúde e mortalidade, é fundamental que haja estudos bem conduzidos para testar a hipótese de que o reverso também é verdadeiro: que redução de peso e de glicemia tenham impacto direto na redução dos riscos de complicações de saúde. Em relação ao diabetes, já temos boas evidências de que a redução glicêmica impacta em longo prazo ao reduzir complicações microvasculares e, em menor impacto, ao reduzir complicações macrovasculares (estas dependem também do controle pressórico e lipídico). Mais recentemente, alguns fármacos usados para tratar diabetes demonstraram reduzir risco cardiovascular independentemente do efeito direto na glicose. Em relação à obesidade, a melhor evidência de redução de riscos e mortalidade vem com a cirurgia bariátrica, provavelmente também devido a sua eficácia de perda e manutenção da perda de peso substancialmente maior que qualquer tratamento clínico disponível.

Importante, estudos clínicos estão sendo realizados a todo momento e é possível, na verdade, esperado, que as recomendações futuras para as morbidades aqui discutidas sejam diferentes das atuais.

## Referências bibliográficas

1. Peeters RP. Subclinical Hypothyroidism. *N Engl J Med*. 2017 Jun;376(26):2556-65.
2. Biondi B, Cappola AR, Cooper DS. Subclinical Hypothyroidism: A Review. *JAMA*. 2019 Jul;322(2):153-60.
3. Somwaru LL, Rariy CM, Arnold AM, Cappola AR. The natural history of subclinical hypothyroidism in the elderly: the cardiovascular health study. *J Clin Endocrinol Metab*. 2012 Jun;97(6):1962-9.
4. Meyerovitch J, Rotman-Pikielny P, Sherf M, Battat E, Levy Y, Surks MI. Serum thyrotropin measurements in the community: five-year follow-up in a large network of primary care physicians. *Arch Intern Med*. 2007 Jul;167(14):1533-8.
5. Tseng F-Y, Lin W-Y, Lin C-C, Lee L-T, Li T-C, Sung P-K, et al. Subclinical hypothyroidism is associated with increased risk for all-cause and cardiovascular mortality in adults. *J Am Coll Cardiol*. 2012 Aug;60(8):730-7.
6. Rodondi N, den Elzen WPJ, Bauer DC, Cappola AR, Razvi S, Walsh JP, et al. Subclinical hypothyroidism and the risk of coronary heart disease and mortality. *JAMA*. 2010 Sep;304(12):1365-74.

7. Aziz M, Kandimalla Y, Machavarapu A, Saxena A, Das S, Younus A, et al. Effect of Thyroxin Treatment on Carotid Intima-Media Thickness (CIMT) Reduction in Patients with Subclinical Hypothyroidism (SCH): a Meta-Analysis of Clinical Trials. *J Atheroscler Thromb*. 2017 Jul;24(7):643-59.
8. Razvi S, Weaver JU, Butler TJ, Pearce SHS. Levothyroxine treatment of subclinical hypothyroidism, fatal and nonfatal cardiovascular events, and mortality. *Arch Intern Med*. 2012 May;172(10):811-7.
9. Jabbar A, Ingoe L, Junejo S, Carey P, Addison C, Thomas H, et al. Effect of Levothyroxine on Left Ventricular Ejection Fraction in Patients With Subclinical Hypothyroidism and Acute Myocardial Infarction: A Randomized Clinical Trial. *JAMA*. 2020 Jul;324(3):249-58.
10. Gencer B, Moutzouri E, Blum MR, Feller M, Collet T-H, Delgiovane C, et al. The Impact of Levothyroxine on Cardiac Function in Older Adults With Mild Subclinical Hypothyroidism: A Randomized Clinical Trial. *Am J Med*. 2020 Jul;133(7):848-856.e5.
11. Najafi L, Malek M, Hadian A, Ebrahim Valojerdi A, Khamseh ME, Aghili R. Depressive symptoms in patients with subclinical hypothyroidism--the effect of treatment with levothyroxine: a double-blind randomized clinical trial. *Endocr Res*. 2015;40(3):121-6.
12. Stott DJ, Rodondi N, Kearney PM, Ford I, Westendorp RGJ, Mooijaart SP, et al. Thyroid Hormone Therapy for Older Adults with Subclinical Hypothyroidism. *N Engl J Med*. 2017 Jun;376(26):2534-44.
13. Lin JS, Bowles EJA, Williams SB, Morrison CC. Screening for Thyroid Cancer: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA*. 2017 May;317(18):1888-903.
14. Davies L, Welch HG. Current thyroid cancer trends in the United States. *JAMA Otolaryngol Head Neck Surg*. 2014 Apr;140(4):317-22.
15. Ito Y, Miyauchi A, Inoue H, Fukushima M, Kihara M, Higashiyama T, et al. An observational trial for papillary thyroid microcarcinoma in Japanese patients. *World J Surg*. 2010 Jan;34(1):28-35.
16. Wharton S, Lau DCW, Vallis M, Sharma AM, Biertho L, Campbell-Scherer D, et al. Obesity in adults: a clinical practice guideline. *C Can Med Assoc J = J l'Association medicale Can*. 2020 Aug;192(31):E875-91.
17. Whitlock G, Lewington S, Sherliker P, Clarke R, Emberson J, Halsey J, et al. Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. *Lancet (London, England)*. 2009 Mar;373(9669):1083-96.
18. Flegal KM, Kit BK, Orpana H, Graubard BI. Association of all-cause mortality with overweight and obesity using standard body mass index categories: a systematic review and meta-analysis. *JAMA*. 2013 Jan;309(1):71-82.

19. Xu H, Cupples LA, Stokes A, Liu C-T. Association of Obesity With Mortality Over 24 Years of Weight History: Findings From the Framingham Heart Study. *JAMA Netw open*. 2018 Nov;1(7):e184587.
20. Khan SS, Ning H, Wilkins JT, Allen N, Carnethon M, Berry JD, et al. Association of Body Mass Index With Lifetime Risk of Cardiovascular Disease and Compression of Morbidity. *JAMA Cardiol*. 2018 Apr;3(4):280-7.
21. Fildes A, Charlton J, Rudisill C, Littlejohns P, Prevost AT, Gulliford MC. Probability of an Obese Person Attaining Normal Body Weight: Cohort Study Using Electronic Health Records. *Am J Public Health*. 2015 Sep;105(9):e54-9.
22. Leblanc ES, O'Connor E, Whitlock EP, Patnode CD, Kapka T. Effectiveness of primary care-relevant treatments for obesity in adults: a systematic evidence review for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2011 Oct;155(7):434-47.
23. Gardner CD, Trepanowski JF, Del Gobbo LC, Hauser ME, Rigdon J, Ioannidis JPA, et al. Effect of Low-Fat vs Low-Carbohydrate Diet on 12-Month Weight Loss in Overweight Adults and the Association With Genotype Pattern or Insulin Secretion: The DIETFITS Randomized Clinical Trial. *JAMA*. 2018 Feb;319(7):667-79.
24. Johnston BC, Kanters S, Bandayrel K, Wu P, Naji F, Siemieniuk RA, et al. Comparison of weight loss among named diet programs in overweight and obese adults: a meta-analysis. *JAMA*. 2014 Sep;312(9):923-33.
25. Shaw K, Gennat H, O'Rourke P, Del Mar C. Exercise for overweight or obesity. *Cochrane database Syst Rev*. 2006 Oct;(4):CD003817.
26. Wadden TA, Volger S, Sarwer DB, Vetter ML, Tsai AG, Berkowitz RI, et al. A two-year randomized trial of obesity treatment in primary care practice. *N Engl J Med*. 2011 Nov;365(21):1969-79.
27. Magkos F, Fraterrigo G, Yoshino J, Luecking C, Kirbach K, Kelly SC, et al. Effects of Moderate and Subsequent Progressive Weight Loss on Metabolic Function and Adipose Tissue Biology in Humans with Obesity. *Cell Metab*. 2016 Apr;23(4):591-601.
28. Ryan DH, Yockey SR. Weight Loss and Improvement in Comorbidity: Differences at 5%, 10%, 15%, and Over. *Curr Obes Rep*. 2017 Jun;6(2):187-94.
29. Wing RR, Bolin P, Brancati FL, Bray GA, Clark JM, Coday M, et al. Cardiovascular effects of intensive lifestyle intervention in type 2 diabetes. *N Engl J Med*. 2013 Jul;369(2):145-54.
30. Gregg E, Jakicic J, Blackburn G, Bloomquist P, Bray G, Clark J, et al. Association of the magnitude of weight loss and changes in physical fitness with long-term cardiovascular disease outcomes in overweight or obese people with type 2 diabetes: a post-hoc analysis of the Look AHEAD randomised clinical trial. *lancet Diabetes Endocrinol*. 2016 Nov;4(11):913-21.

31. Marso SP, Daniels GH, Brown-Frandsen K, Kristensen P, Mann JFE, Nauck MA, et al. Liraglutide and Cardiovascular Outcomes in Type 2 Diabetes. *N Engl J Med*. 2016 Jul;375(4):311-22.
32. Sjöström L. Review of the key results from the Swedish Obese Subjects (SOS) trial – a prospective controlled intervention study of bariatric surgery. *J Intern Med*. 2013 Mar;273(3):219-34.
33. Carlsson LMS, Sjöholm K, Jacobson P, Andersson-Assarsson JC, Svensson P-A, Taube M, et al. Life Expectancy after Bariatric Surgery in the Swedish Obese Subjects Study. *N Engl J Med*. 2020 Oct;383(16):1535-43.
34. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2020. *Diabetes Care*. 2020 Jan;43(Suppl 1):S14-31.
35. Vecchio I, Tornali C, Bragazzi NL, Martini M. The Discovery of Insulin: An Important Milestone in the History of Medicine. *Front Endocrinol (Lausanne)*. 2018;9:613.
36. Nathan DM, Genuth S, Lachin J, Cleary P, Crofford O, Davis M, et al. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med*. 1993 Sep;329(14):977-86.
37. Nathan DM. The diabetes control and complications trial/epidemiology of diabetes interventions and complications study at 30 years: overview. *Diabetes Care*. 2014;37(1):9-16.
38. Sabet S, Condren ME, Boston AF, Doak LC, Chalmers LJ. Evolving Pharmacotherapeutic Strategies for Type 1 Diabetes Mellitus. *J Pediatr Pharmacol Ther JPPT Off J PPAG*. 2018;23(5):351-61.
39. Taylor SI, Blau JE, Rother KI, Beitelshes AL. SGLT2 inhibitors as adjunctive therapy for type 1 diabetes: balancing benefits and risks. *lancet Diabetes Endocrinol*. 2019 Dec;7(12):949-58.
40. Lebovitz HE. Type 2 diabetes: an overview. *Clin Chem*. 1999 Aug;45(8 Pt 2):1339-45.
41. Gregg EW, Cheng YJ, Srinivasan M, Lin J, Geiss LS, Albright AL, et al. Trends in cause-specific mortality among adults with and without diagnosed diabetes in the USA: an epidemiological analysis of linked national survey and vital statistics data. *Lancet (London, England)*. 2018 Jun;391(10138):2430-40.
42. Turner RC, Holman RR. Lessons from UK prospective diabetes study. *Diabetes Res Clin Pract*. 1995 Aug;28 Suppl:S151-7.
43. Levy JC. Review: UKPDS odyssey – 2001. *Br J Diabetes Vasc Dis [Internet]*. 2001;1(1):14-21 [[accesso](#) em: 31 mar. 2021].
44. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS

- 33). UK Prospective Diabetes Study (UKPDS) Group. *Lancet* (London, England). 1998 Sep;352(9131):837-53.
45. Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HAW. 10-year follow-up of intensive glucose control in type 2 diabetes. *N Engl J Med*. 2008 Oct;359(15):1577-89.
46. Patel A, MacMahon S, Chalmers J, Neal B, Billot L, Woodward M, et al. Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. *N Engl J Med*. 2008 Jun;358(24):2560-72.
47. Gerstein HC, Miller ME, Byington RP, Goff DCJ, Bigger JT, Buse JB, et al. Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med*. 2008 Jun;358(24):2545-59.
48. Duckworth W, Abraira C, Moritz T, Reda D, Emanuele N, Reaven PD, et al. Glucose control and vascular complications in veterans with type 2 diabetes. *N Engl J Med*. 2009 Jan;360(2):129-39.
49. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med*. 2007 Jun;356(24):2457-71.
50. Kieffer CM, Robertson AS. Impact of FDA-Required Cardiovascular Outcome Trials on Type 2 Diabetes Clinical Study Initiation From 2008 to 2017. *Ther Innov Regul Sci*. 2019 Jul;2168479019860122.
51. Gaede P, Lund-Andersen H, Parving H-H, Pedersen O. Effect of a multifactorial intervention on mortality in type 2 diabetes. *N Engl J Med*. 2008 Feb;358(6):580-91.
52. Zelniker TA, Wiviott SD, Raz I, Im K, Goodrich EL, Furtado RHM, et al. Comparison of the Effects of Glucagon-Like Peptide Receptor Agonists and Sodium-Glucose Cotransporter 2 Inhibitors for Prevention of Major Adverse Cardiovascular and Renal Outcomes in Type 2 Diabetes Mellitus. *Circulation*. 2019 Apr;139(17):2022-31.
53. Zinman B, Wanner C, Lachin JM, Fitchett D, Bluhmki E, Hantel S, et al. Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. *N Engl J Med*. 2015 Nov;373(22):2117-28.
54. Neal B, Perkovic V, Mahaffey KW, de Zeeuw D, Fulcher G, Erondu N, et al. Canagliflozin and Cardiovascular and Renal Events in Type 2 Diabetes. *N Engl J Med*. 2017 Aug;377(7):644-57.
55. Wiviott SD, Raz I, Bonaca MP, Mosenzon O, Kato ET, Cahn A, et al. Dapagliflozin and Cardiovascular Outcomes in Type 2 Diabetes. *N Engl J Med*. 2019 Jan;380(4):347-57.
56. Marso SP, Bain SC, Consoli A, Eliaschewitz FG, Jódar E, Leiter LA, et al. Semaglutide and Cardiovascular Outcomes in Patients with Type 2 Diabetes. *N Engl J Med*. 2016 Nov;375(19):1834-44.
57. Gerstein HC, Colhoun HM, Dagenais GR, Diaz R, Lakshmanan M, Pais P, et al. Dulaglutide and cardiovascular outcomes in type 2 diabetes (REWIND): a double-blind, randomised placebo-controlled trial. *Lancet* (London, England). 2019 Jul;394(10193):121-30.



58. Davies MJ, D'Alessio DA, Fradkin J, Kernan WN, Mathieu C, Mingrone G, et al. Management of hyperglycaemia in type 2 diabetes, 2018. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetologia*. 2018 Dec;61(12):2461-98.
59. Cosentino F, Grant PJ, Aboyans V, Bailey CJ, Ceriello A, Delgado V, et al. 2019 ESC Guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD. *Eur Heart J*. 2020 Jan;41(2):255-323.
60. Buyschaert M, Bergman M. Definition of prediabetes. *Med Clin North Am*. 2011 Mar;95(2):289-97, vii.
61. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med*. 2002 Feb;346(6):393-403.
62. Knowler WC, Fowler SE, Hamman RF, Christophi CA, Hoffman HJ, Brenneman AT, et al. 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. *Lancet (London, England)*. 2009 Nov;374(9702):1677-86.
63. Lachin JM, Christophi CA, Edelstein SL, Ehrmann DA, Hamman RF, Kahn SE, et al. Factors associated with diabetes onset during metformin versus placebo therapy in the diabetes prevention program. *Diabetes*. 2007 Apr;56(4):1153-9.
64. Oh TJ. The Role of Anti-Obesity Medication in Prevention of Diabetes and Its Complications. *J Obes Metab Syndr*. 2019 Sep;28(3):158-66.
65. Reaven GM. The metabolic syndrome: time to get off the merry-go-round? *J Intern Med*. 2011 Feb;269(2):127-36.

# 11. ONCOLOGIA BASEADA EM EVIDÊNCIAS

## Introdução

Do ponto de vista de metodologia científica e da chamada Medicina Baseada em Evidências, a oncologia oferece desafios e perspectivas formidáveis. Perspectivas porque é uma das especialidades médicas com o maior número de estudos clínicos de alta qualidade. Desafios porque existem tumores raros e situações clínicas incomuns, onde é impossível se obter informações derivadas de evidências de alta qualidade.<sup>1,2</sup>

A decisão sobre intervenções médicas em oncologia necessita de um balanço muito fino em termos de risco/benefício, pois, por características próprias desses tratamentos, o potencial benefício pode estar associado a eventos adversos e toxicidades dos medicamentos. Uma análise criteriosa da qualidade dos estudos clínicos, dos desfechos usados, da magnitude do efeito é necessária, assim como o julgamento clínico das preferências dos pacientes.<sup>1,3</sup> Um aspecto importante a se considerar no tratamento do câncer é que a proximidade iminente de morte pela doença faz muitas vezes que nesta equação de risco-benefício o paciente esteja disposto a tomar riscos maiores em relação ao tratamento que em outras especialidades. Essa característica tem implicações éticas e econômicas que vão além do objetivo deste capítulo, mas que precisa ser sempre lembrada.<sup>3</sup>

Um outro complicador do processo de tomada de decisão clínica em oncologia é que mesmo em estudos bem desenhados, muitas vezes é impossível avaliar desfechos finais como sobrevida e o uso de *end points* intermediários (surrogates) abunda na oncologia e é largamente aceito pela comunidade de profissionais de saúde, bem como por agências regulatórias.<sup>1,2</sup>

Outro aspecto importante da decisão clínica em oncologia é que o objetivo de tratamento de um mesmo tumor varia de acordo com o estágio da doença. Muitas vezes a cura, objetivo principal dos tratamentos, não é possível. Passa-se então a ter-se como objetivo a palição de sintomas, a extensão do tempo de vida ou do tempo para a recorrência dos tumores, como explicado a seguir.<sup>1,2</sup>

## Principais objetivos de tratamento na prática oncológica

- **Curativo:** o uso da medicação antineoplásica é feito com intuito primário de curar o paciente. É o caso do tratamento de leucemias, alguns linfomas, tumores de testículo metastáticos, entre outros. Apesar de ser obviamente o objetivo de tratamento desejado, obter cura em oncologia é muito difícil, exceto para tumores precoces ou para alguns poucos tumores avançados, como os citados acima.
- **Adjuvante:** o tratamento é feito após um tratamento principal (cirurgia ou radioterapia), que foi feito com intuito curativo. O uso de medicação antineoplásica tem uma função coadjuvante de retardar ou impedir a recorrência do tumor e também de estender a vida dos pacientes. É o caso de quimioterapia para tumores de mama ou cólon localizados, operados, sem metástase a distância.
- **Neoadjuvante:** o uso de medicação antineoplásica é feito antes de um tratamento principal (cirurgia ou radioterapia), com o intuito de diminuir o tamanho do tumor para facilitar o tratamento principal e/ou aumentar as chances de cura. É o caso, por exemplo, dos tumores de mama e osteossarcoma.
- **Aumento de sobrevida em longo prazo:** o uso da medicação antineoplásica é feito com o intuito de oferecer aos pacientes um aumento de sobrevida por um prazo longo. É o caso de alguns tipos de linfoma, em que mesmo não havendo a possibilidade de cura, os pacientes podem viver por bastante tempo, com a doença relativamente controlada.
- **Paliativo:** o uso da medicação antineoplásica tem como objetivo principal aumentar a sobrevida dos pacientes e/ou oferecer conforto nos casos em não há possibilidade de cura. É o caso de

grande parte dos tratamentos oncológicos, como tumor de pulmão avançado, pâncreas etc.

Diferentes desfechos clínicos são usados em oncologia e a escolha do desfecho mais adequado varia de acordo com o tipo de tumor, a presença ou não de “*surrogates*” validados, do objetivo do tratamento oferecido no estudo e da magnitude esperada do efeito.<sup>1-2,4</sup>

O uso de desfechos intermediários (*surrogate*, em inglês), é comum em oncologia. Preferencialmente, *surrogates* devem passar por um processo de validação em estudos clínicos, onde se analisa a capacidade estatística do *surrogate* prever o desfecho final. O processo de validação de *surrogates* é complexo, depende frequentemente de estudos clínicos feitos com o propósito específico de validação e por isso são raros os desfechos intermediários realmente validados.<sup>5</sup> Há, porém alguns *surrogates* validados (sobrevida livre de progressão, resposta tumoral, tempo para progressão da doença ) que são capazes de prever ganhos de sobrevida para alguns tipos de tumores, como ovário, cólon e cabeça e pescoço.<sup>6-7</sup> A aceitabilidade do resultado de um estudo clínico que *surrogates*, situação comum na oncologia, depende do tipo de tumor e da magnitude do resultado. Não há uma regra clara e a maioria das entidades regulatórias recomenda pesar os resultados do estudo com a experiência clínica anterior para julgar a adequação dos resultados.<sup>2,7</sup>

## Os principais desfechos em oncologia são:

### Sobrevida global

Em oncologia, estender o tempo de vida do paciente é o desfecho preferido. Porém, há situações em que não se consegue analisar de forma adequada a sobrevida global, seja por questões práticas (doenças com evolução muito longa), éticas/ técnicas (cruzamento de pacientes para o braço com a terapia inovadora em estudos randomizados ou uso de múltiplas linhas de tratamento).

A sobrevida global é formalmente definida como o tempo da randomização até morte por qualquer causa. Sua análise deve ser feita preferencialmente em estudos randomizados, pois desfechos de sobrevida usualmente não podem ser avaliados de forma confiável em estudos com controles históricos.<sup>8</sup> A exceção à regra fica por conta daqueles raros

estudos que mostram uma magnitude de efeito extrema, em que estudos com desenho antes/depois podem ter um grau adequado de confiabilidade.<sup>9</sup>

Outro ponto importante a ser considerado é que sobrevida é um desfecho dependente do tempo e sua medida deve sempre ser reportada como “time to event data” (análise de sobrevivência) em curvas de sobrevida. Análises pontuais de sobrevida em tempos específicos ou a sobrevida mediana não são desfechos válidos por si só, são medidas auxiliares, que devem ser consideradas no contexto total da análise das curvas de sobrevida.<sup>10</sup> As análises de “time to event data” são reportadas como razão de risco (em inglês, *hazard ratio*) e representam as chances relativas de se estar vivo ao longo do tempo do estudo. Importante entender que as análises de sobrevivência se referem a qualquer evento relacionado ao tempo de evolução da doença e não somente à sobrevida global – tempo para progressão da doença, sobrevida livre de progressão sobrevida livre de doença, entre outros, são tipos de análises de sobrevivência usados em oncologia, como veremos mais adiante e já foi visto antes neste livro (Capítulo 6).

## Sobrevida livre de progressão

A sobrevida livre de progressão (SLP) é definida como o tempo entre a randomização e a progressão da doença ou morte do paciente,<sup>9-11, 13</sup> sendo um dos *end points* mais relatados em oncologia. Há uma percepção tanto de médicos quanto de pacientes que se um câncer está respondendo ao tratamento (i.e., não cresceu, não teve aumento de marcadores tumorais), o tratamento está sendo benéfico para o paciente. Porém, como relatamos acima, nem sempre esta correlação existe ou é conhecida, pela falta de validação adequada do *surrogate*.<sup>2, 6-7</sup>

## Tempo para progressão da doença

O tempo para a progressão da doença é definido como o tempo entre a randomização e a progressão da doença. A diferença entre TTP da doença e a SLP é que o primeiro não inclui as mortes.<sup>9, 11-12</sup>

## Sobrevida livre de doença

A sobrevida livre de doença é definida como tempo entre a randomização e a recorrência do tumor ou morte, em pacientes que foram

submetidos a um tratamento com intuito curativo (cirurgia ou radioterapia) e não têm doença detectável – é muito semelhante à SLP, porém, nesse caso, como o paciente está sem evidência de doença, por não ter doença detectável (por métodos de imagem ou outros), usa-se o termo sobrevida livre de doença (e não de progressão). A SLD é o desfecho usado em tratamentos adjuvante ou que têm alto percentual de remissão (desaparecimento) completo do tumor, caso de algumas neoplasias hematológicas.<sup>9, 11-12</sup>

## Resposta tumoral

As avaliações de resposta são as proporções de pacientes que tiveram redução do tumor ao longo do tratamento (resposta). São consideradas respostas as parciais (RP) e as completas (RC) ou estabilização da doença. Há diversos critérios para avaliação de resposta, entre os mais usados está o *Response Evaluation Criteria in Solid Tumours* (RECIST).<sup>13-14</sup> De forma genérica, considera-se que uma redução de 50% nas dimensões do tumor indica resposta parcial e o desaparecimento do mesmo nos exames de imagem indicam resposta completa. Se os tumores têm entre 0 e 50% de redução, considera-se estabilização da doença.

## Avaliação clínica de tumores hematológicos

Tumores hematológicos que não desenvolvem massas, como é o caso das leucemias, utilizam critérios diferentes (usualmente mais complexos) de avaliação de resposta como *endpoint*.<sup>15</sup> Estes incluem contagem de células sanguíneas, avaliação de medula óssea, e morfologia celular. Em alguns casos, a análise genética (citogenética) é usada na avaliação de resposta em tumores hematológicos.<sup>1</sup>

## Desfechos relatados pelos pacientes (qualidade de vida)

A análise de desfechos relatados pelos pacientes (*patient reported outcomes PROs*) é um campo de interesse crescente em oncologia, pois traz a percepção do paciente sobre seu próprio tratamento, seus sintomas e sua qualidade de vida.<sup>16</sup> Em tese, traduzem a percepção que o paciente tem de seu tratamento. Porém, os PROs têm como limitação a dificuldade de

transformar a percepção subjetiva do paciente em mensuração objetiva, bem como a extrema variabilidade de valorização de sintomas entre as pessoas. Essas análises de qualidade de vida, apesar de muito comentadas e desejadas, têm uma aplicabilidade ainda limitada na avaliação de estudos clínicos, por conta destas limitações.<sup>16-17</sup>

## Tumores raros

Um capítulo à parte em oncologia é a análise de tratamentos para situações infrequentes ou para tumores raros,<sup>18</sup> onde não é factível a realização de um estudo com grande tamanho amostral. Nesses casos pode ser aceitável evidência derivada de estudos pequenos, randomizados ou não, com avaliação de desfechos intermediários.<sup>2, 11, 17</sup>

A definição do que é um tumor raro varia. A definição oficial é um tumor com incidência de menos que 65 novos casos por 100.000 pessoas por ano.<sup>18-19</sup> O capítulo “tumores raros” tende a crescer muito na oncologia nos próximos anos, pois com o desenvolvimento do conhecimento molecular, uma boa parte dos tumores está se tornando, por definição, uma doença rara, devido à identificação de diversos subtipos de câncer de uma mesma denominação comum.<sup>2, 17</sup>

## Questões adicionais a serem feitas nos estudos clínicos

Ao se avaliar um estudo clínico em oncologia, algumas questões adicionais devem ser feitas, além daquelas tradicionais:

Uma delas é sobre o *endpoint* utilizado, que abordamos acima. Outras duas são:

1. O estudo foi desenhado para mostrar superioridade ou não inferioridade?
2. A população estudada representa a população geral em que a intervenção será usada?

## Superioridade X não inferioridade

A maior parte dos estudos clínicos (randomizados) em oncologia (e nas outras disciplinas). são desenhados para mostrar que o medicamento/

intervenção é superior a outro. As análises de qualidade metodológicas destes estudos de superioridade em oncologia são, de forma geral, as mesmas usadas em para todos os estudos, considerando-se as particularidades da especialidade.<sup>2, 20</sup>

Nos últimos anos, em oncologia, outro tipo de estudo randomizado tem ganhado espaço. Em certas situações, o que se deseja medir não é a superioridade de um medicamento sobre o outro, mas a falta de diferença entre eles em termos de eficácia, os chamados estudos de não inferioridade. Esses estudos existem para avaliar medicamentos/ intervenções que ofereçam resultados muito parecidos em termos da eficácia primária, mas um deles traz alguma vantagem, como, por exemplo, melhor conforto para o paciente, menos eventos adversos, menor custo etc.<sup>2, 20</sup>

Do ponto de vista metodológico, estudos de não inferioridade são de desenho e interpretação mais complexos, pois requerem um controle muito estrito do erro tipo II (i.e., declarar que as diferenças não existem, quando elas na realidade existem). Ao se desenhar um estudo de não inferioridade o pesquisador declara que aceita uma diferença para o desfecho primário de até x%, e que se as duas drogas testadas ficarem com resultados dentro desta margem de não inferioridade, podem ser consideradas parecidas o suficiente para que uma substitua a outra em termos de eficácia.<sup>2, 8, 20-22</sup> É importante salientar que estes estudos têm sido chamados erroneamente de “estudos de equivalência”. Por razões estatísticas mais complexas, há uma diferença importante entre equivalência e não inferioridade – provar equivalência requer tamanho de amostra imensos, frequentemente na casa de dezenas de milhares de pacientes em cada estudo. No caso dos estudos de não inferioridade, se aceita que PODE haver uma diferença entre as intervenções, mas que ela é tão pequena que pode ser desprezada. Nesse caso, o tamanho amostral requerido é bem menor.<sup>2, 20</sup>

Um movimento recente, intitulado “*valued based practice*” tem estimulado a realização de estudos de não inferioridade, comparando medicamentos de eficácia parecida, mas com custos diferentes a fim de promover o uso de intervenções com melhor perfil de custo efetividade. Provavelmente, veremos um crescimento nas publicações usando desenhos de não inferioridade nos próximos anos.<sup>20</sup>

## População estudada



É extremamente importante que durante o processo de avaliação a população estudada seja claramente especificada, principalmente em relação aos critérios de inclusão do estudo – intenção e linha do tratamento, gravidade da doença (estágio) etc.

Infelizmente, os estudos clínicos em oncologia têm utilizado critérios de inclusão cada vez mais rígidos e por isso muitos questionamentos têm sido feitos sobre se a população estudada realmente representa a população na qual a intervenção será utilizada. Alguns dados mostram que pacientes com a mesma doença, tratados na comunidade, submetidos ao mesmo tratamento que pacientes em estudos clínicos têm resultados piores que esses. Há um debate importante sobre a validade externa dos resultados dos estudos clínicos que deverá ganhar tração nos próximos anos.<sup>23</sup>

## Conclusão

A avaliação de estudos clínicos em oncologia segue em parte os mesmos princípios das avaliações de estudos clínicos em geral, mas há especificidades que precisam ser levadas em conta.

Um debate ético importante a se considerar é o desejo do paciente – que em oncologia pode estar disposto a aceitar um certo risco, quando vislumbra um possível desfecho fatal de sua doença.

## Referências bibliográficas

1. Janiaud P, Serghiou S, Ioannidis JPA. New clinical trial designs in the era of precision medicine: An overview of definitions, strengths, weaknesses, and current use in oncology. *Cancer Treat Rev.* 2019;73:20-30.
2. Kilickap S, Demirci U, Karadurmus N, Dogan M, Akinci B, Sendur MAN. Endpoints in oncology clinical trials. *J buon.* 2018;23(7):1-6.
3. Kuchuk I, Bouganim N, Beusterien K, Grinspan J, Vandermeer L, Gertler S, et al. Preference weights for chemotherapy side effects from the perspective of women with breast cancer. *Breast Cancer Res Treat.* 2013;142(1):101-7.
4. Hudis CA, Barlow WE, Costantino JP, Gray RJ, Pritchard KI, Chapman JA, et al. Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: the STEEP system. *J Clin Oncol.* 2007;25(15):2127-32.
5. Williams G. Colorectal Cancer Endpoints Workshop Summary. In: Food and Drugs Administration, editor. *Regulations and Endpoints: FDA;* 2003.

6. Sherrill B, Kaye JA, Sandin R, Cappelleri JC, Chen C. Review of meta-analyses evaluating surrogate endpoints for overall survival in oncology. *Onco Targets Ther.* 2012;5:287-96.
7. Prasad V, Kim C, Burotto M, Vandross A. The Strength of Association Between Surrogate End Points and Survival in Oncology: A Systematic Review of Trial-Level Meta-analyses. *JAMA Intern Med.* 2015;175(8):1389-98.
8. Brody T. *Clinical Trials: Study Design, Endpoints and Biomarkers, Drug Safety, and FDA and ICH Guidelines.* Amsterdam: Elsevier; 2012.
9. Food and Drugs Administration. Guidance for Industry—clinical trial endpoints for the approval of cancer drugs and biologics 2007 [[acesso em: 8 abr. 2021](#)].
10. Bland JM, Altman DG. Survival probabilities (the Kaplan-Meier method). *BMJ.* 1998;317(7172):1572.
11. European Medicines Agency EMA. Guideline on the evaluation of anticancer medicinal products in man – EMA/CHMP/205/95/Rev.4 2012 [[acesso em: 8 abr. 2021](#)].
12. European Medicines Agency EMA. Appendix 1 to the guideline on the evaluation of anticancer medicinal products in man – EMA/CHMP/27994/2008/Rev.1 2012 [[acesso em: 8 abr. 2021](#)].
13. Bellera CA, Pulido M, Gourgou S, Collette L, Doussau A, Kramar A, et al. Protocol of the Definition for the Assessment of Time-to-event Endpoints in CANcer trials (DATECAN) project: formal consensus method for the development of guidelines for standardised time-to-event endpoints' definitions in cancer clinical trials. *Eur J Cancer.* 2013;49(4):769-81.
14. Fiteni F, Westeel V, Pivot X, Borg C, Vernerey D, Bonnetain F. Endpoints in cancer clinical trials. *J Visc Surg.* 2014.
15. Appelbaum FR, Rosenblum D, Arceci RJ, Carroll WL, Breitfeld PP, Forman SJ, et al. End points to establish the efficacy of new agents in the treatment of acute leukemia. *Blood.* 2007;109(5):1810-6.
16. Secord AA, Coleman RL, Havrilesky LJ, Abernethy AP, Samsa GP, Cella D. Patient-reported outcomes as end points and outcome indicators in solid tumours. *Nat Rev Clin Oncol.* 2015;12(6):358-70.
17. Zettler M, Basch E, Nabhan C. Surrogate End Points and Patient-Reported Outcomes for Novel Oncology Drugs Approved Between 2011 and 2017. *JAMA Oncol.* 2019;5(9):1358-9.
18. Gaddipati H, Liu K, Pariser A, Pazdur R. Rare cancer trial design: lessons from FDA approvals. *Clin Cancer Res.* 2012;18(19):5172-8.
19. Ministério da Saúde. Portaria Nº 199, de 30 de janeiro de 2014 – Institui a Política Nacional de Atenção Integral às Pessoas com Doenças Raras, aprova as Diretrizes para Atenção Integral às Pessoas com Doenças Raras no âmbito do Sistema Único de Saúde (SUS) e institui incentivos financeiros de custeio. 2014 [[acesso em: 8 abr. 2021](#)].

20. Stebbing J, Mainwaring PN, Curigliano G, Pegram M, Latymer M, Bair AH, et al. Understanding the Role of Comparative Clinical Studies in the Development of Oncology Biosimilars. *J Clin Oncol*. 2020;38(10):1070-80.
21. Pocock SJ. *Clinical Trials: A Practical Approach*. New York: John Wiley and Sons; 1983.
22. Haynes RB, Sackett DL, Guyat GH, Tugwell P. *Clinical Epidemiology: How to Do Clinical Practice Research 3ed*: Lippincott, Williams and Wilkins; 2006.
23. Green AK, Curry M, Trivedi N, Bach PB, Mailankody S. Assessment of Outcomes Associated With the Use of Newly Approved Oncology Drugs in Medicare Beneficiaries. *JAMA Netw Open*. 2021;4(2):e210030.

# 12. RASTREAMENTO DE CÂNCER BASEADO EM EVIDÊNCIAS

## Introdução

Simplificadamente, rastreamento é a procura por uma doença (ou fator de risco de uma doença) em pessoas *que não têm sinais e sintomas dessa doença*. Neste capítulo, após uma introdução geral, focaremos questões metodológicas do rastreamento do câncer. Porém, é importante pontuar que o rastreamento de diversas outras doenças é também largamente utilizado na prática clínica de forma geral, como, por exemplo, o rastreamento de doenças cardiovasculares.

O que hoje conhecemos por rastreamento provavelmente teve origem com o exame médico periódico em pessoas saudáveis, conhecido popularmente como *check-up*. O exame periódico tem origem desconhecida, mas muitos autores consideram “o surgimento intelectual” em meados de 1860, com um médico inglês chamado Horace Dobell, um especialista em tuberculose e outras doenças pulmonares. Dobell acreditava que as doenças eram precedidas por “estágios fisiológicos preexistentes” de “baixa saúde” e que a intervenção terapêutica era mais eficaz nesses estágios precoces ou iniciais das doenças. Com isso, Dobell propôs o exame médico periódico como uma maneira de identificar “os estágios iniciais de defeito no estado fisiológico e adotar medidas para remediá-los”. Ele defendia a coleta de um extensivo histórico de saúde do paciente, um exame físico meticuloso e uso de exames laboratoriais. Dessa ideia, surge a noção de que até uma pequena anormalidade fisiológica tem relevância patogênica e sua detecção confere maior poder terapêutico.<sup>1</sup>

George Gould também teve papel importante na defesa do exame periódico, mas por motivos diferentes: adquirir conhecimento científico sobre o histórico natural da doença.

No início dos anos 1900, o exame periódico de crianças foi defendido como maneira de detectar anormalidades físicas e doenças contagiosas. A ideia do exame físico anual provavelmente começou a ter popularidade quando a National Tuberculosis Association dedicou em 1915 uma semana para que todas as pessoas fossem submetidas a um exame físico. Já em 1918,

mesmo com pouco conhecimento de câncer, vários médicos recomendavam exames anuais específicos para detecção de câncer – com a esperança de que o diagnóstico precoce aumentaria a chance de cura.<sup>1</sup>

A indústria de seguro de vida também teve o seu papel. Empresas de seguro de vida usavam exames médicos para determinar o risco de morte dos aplicantes ao seguro de vida. Essas companhias criaram departamentos médicos e contrataram médicos para realização de exames físicos. Em pouco tempo, as empresas passaram de um único exame físico para realização periódica.<sup>1</sup> Além disso, os dados coletados pelas empresas mostravam aqueles que persistiam nos exames físicos tinham mortalidade menor do que esperado. Hoje sabemos que essa análise tinha um *viés de seleção*: aqueles com maior expectativa de vida seguiam mais as recomendações de exames físicos periódicos. No entanto, na época, foi fácil assumir que os exames periódicos eram *a causa* de menor mortalidade.<sup>2</sup> Como os objetivos das companhias eram econômicos, qualquer anormalidade em algum exame era considerada um risco até ser provado o contrário. O interessante é que quase todas as pessoas saudáveis tinham alguma anormalidade, já que os exames eram exaustivos. Isso deveria levar ao questionamento se essas anormalidades não deveriam ser consideradas dentro do intervalo da normalidade. A detecção dessas anormalidades, no entanto, apenas reforçou a ideia de que os exames periódicos eram realmente necessários.<sup>2</sup>

Em 1922, a American Medical Association não só recomendou o exame periódico como participou em grandes campanhas de informação pública com objetivo de aumentar o número aderentes ao exame periódico. A principal razão foi interesse, já que na época os programas públicos de saúde estavam tirando a relação exclusiva de médicos particulares e seus pacientes. O exame periódico se tornou uma arma na mão dos médicos privados que eram contra o seguro de saúde compulsório, programas de saúde em escolas e programas de saúde municipais. Por volta de 1950, o exame periódico já era prática padrão.<sup>2</sup>

Embora alguns tipos de rastreamento sejam de fato eficazes, a comprovação científica desse benefício é muito mais complexa do que parece e um grande desafio para a compreensão não apenas para a população em geral e para mídia leiga, mas também para pesquisadores, gestores, médicos e outros profissionais de saúde.

## O que é rastreamento?

O rastreamento é a realização de testes de rotina em indivíduos assintomáticos, para detecção precoce de determinada alteração ou doença. O pressuposto básico é que ao antecipar a referida detecção, os tratamentos serão mais eficazes e mortes serão evitadas. Em alguns poucos casos é possível realizar a prevenção primária do câncer, com a detecção e tratamento de lesões pré-neoplásicas, como pode ocorrer nos rastreamentos dos cânceres de colo do útero e colorretal. É importante entender que, ao utilizarmos o termo rastreamento, não nos referimos apenas a um único teste ou exame. Rastreamento envolve toda a sequência de eventos necessária para atingir a desejada redução de risco, seja o risco de morrer em virtude de um tipo de câncer, seja – em alguns casos – o de desenvolver um determinado câncer. Em outras palavras, rastreamento é um processo complexo e não um único teste. A Figura 1 ilustra um fluxo básico de um programa de rastreamento. Além disso, para que o resultado seja mais benéfico do que danoso, esse processo precisa ser bem-organizado.

**Figura 1.**

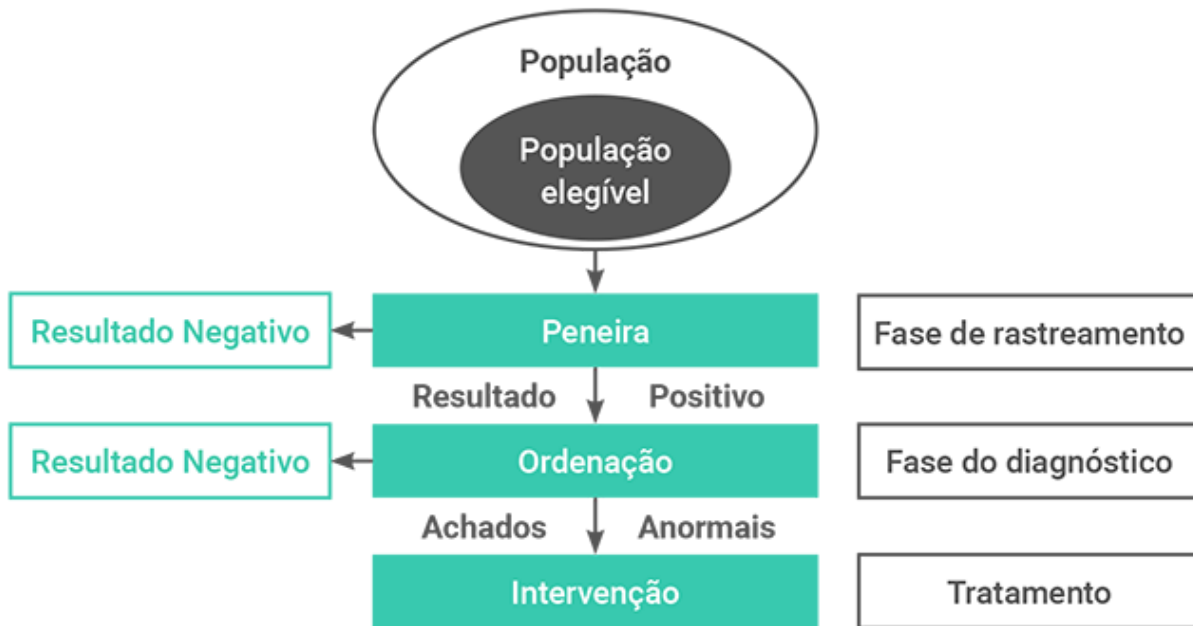


Diagrama de fluxo básico de rastreamento. Por ser um diagrama de fluxo básico, não inclui, por exemplo, resultados inconclusivos. Adaptada.<sup>2</sup>

Nesse sentido, é fundamental estabelecer a diferença entre rastreamento *organizado* e rastreamento *oportunistico*. O primeiro pressupõe a existência de programas de rastreamento de base populacional, com diretrizes clínicas

baseadas em evidências bem definidas, convocação ativa da população-alvo, monitoramento da qualidade, seguimento dos casos alterados no rastreamento, monitoramento da confirmação diagnóstica e do encaminhamento para o tratamento. Já o rastreamento oportunístico tipicamente ocorre em função da oportunidade de uma visita ao serviço de saúde por outras causas, sendo o teste de rastreamento oferecido ou prescrito ao usuário/paciente.

## Os efeitos do rastreamento

Normalmente, os estudos de rastreamento possuem duas fases. A primeira onde os pacientes do grupo intervenção são submetidos ao rastreamento. Posteriormente, os pacientes de ambos os grupos são apenas acompanhados. Isso ocorre porque o objetivo do rastreamento é a detecção precoce. Com isso, é normal que um número maior de cânceres seja detectado no grupo rastreamento em comparação com o grupo controle durante os primeiros anos, o que é esperado por refletir uma antecipação do diagnóstico. Casos que seriam diagnosticados clinicamente mais tarde e em uma fase da doença mais avançada são diagnosticados pelo rastreamento precocemente. Com isso, o rastreamento busca fazer uma redução do estágio da doença no momento do diagnóstico. Se esses casos identificados realmente corresponderem à detecção precoce, o número de casos no grupo controle aumentará com o passar dos anos até ficar igual aos número de casos reportado no grupo dos rastreados. Dessa maneira, é necessário um extenso acompanhamento nos estudos randomizados de rastreamento. Contudo, à medida que os grandes estudos de rastreamento foram sendo realizados, começou-se a perceber que a realidade era mais complexa do que o esperado.

Talvez, a melhor forma de entendermos os efeitos do rastreamento é começarmos pelo câncer de próstata, tanto por sua grande popularidade quanto pelos problemas advindos de sua prática.

### Rastreamento do câncer de próstata

Há dois grandes estudos que avaliam o rastreamento do câncer de próstata: *Prostate, Lung, Colorectal, Ovarian Cancer Screening Trial* (PLCO) e *European Randomized Study of Screening for Prostate Cancer*

(ERSPC). Ambos utilizaram o antígeno prostático específico (PSA) com ou sem exame de toque digital.

No estudo PLCO, aproximadamente 77 mil homens com idade entre 55 e 74 anos foram randomizados entre rastreamento com PSA anual por seis anos ou tratamento padrão. Provavelmente, a maior limitação do PLCO foi a contaminação. A cada ano da fase de rastreamento, por volta de 46% dos homens do grupo controle foram rastreados, em comparação com 85% dos homens do grupo rastreamento. Por isso, alguns autores consideram que esse estudo comparou a eficácia de um rastreamento organizado com rastreamento oportunístico. Mesmo com extenso acompanhamento de 14 anos, mais cânceres foram detectados no grupo rastreamento. No entanto, o rastreamento não se mostrou eficaz; a mortalidade do câncer de próstata foi idêntica em ambos os grupos.<sup>3</sup>

O estudo ERSPC começou em 1993 na Holanda e Bélgica, com outros centros europeus começando mais tarde (até 2003). Apesar de uma faixa etária maior e mais pessoas foram recrutadas, o estudo reporta os resultados de 163 mil homens entre 55 a 69 anos. O rastreamento com PSA foi feito a cada quatro anos (com exceção da Suécia, que foi feito a cada dois anos). O estudo reportou uma redução relativa de 21% na mortalidade do câncer de próstata. Isso traduz em uma redução absoluta de 1,3 mortes a cada 1.000 homens. O número necessário para convidar para o rastreamento (análogo ao NNT apresentado no capítulo 6) foi de 781, com 27 homens sendo diagnosticados e tratados, para que 1 morte de câncer de próstata seja evitada.<sup>3</sup> Apesar de terem feito rastreamento, 5 homens ainda morrem de câncer de próstata.

É muito importante entender o que aconteceu com esses 27 homens, porque representam o dano mais sério do rastreamento: sobrediagnóstico e, sua consequência, sobretratamento. Esses 27 homens foram diagnosticados e tratados *desnecessariamente* com uma combinação de prostatectomia, quimioterapia, radioterapia e hormonioterapia.

Sobrediagnóstico não é um erro de diagnóstico, não devendo ser confundido com um resultado falso-positivo. Sobrediagnóstico ocorre quando uma condição é diagnosticada corretamente, mas que não causaria sintomas nem a morte. Falando especificamente do câncer, a lesão detectada preenche corretamente os critérios histopatológicos da doença. No entanto, o câncer regrediria ou evoluiria tão lentamente que o paciente morreria de outras causas. Se não fosse pelo rastreamento (ou outros exames – porque



pode ocorrer mesmo na ausência de rastreamento), o câncer não seria diagnosticado. Outra definição de sobrediagnóstico é a simples detecção de uma pseudoenferma.<sup>4</sup> Portanto, o sobrediagnóstico é um problema essencialmente de desconhecimento do prognóstico das “doenças” diagnosticadas, o que em geral, resulta também em excesso de tratamento (sobretreamento).

Se após um extenso acompanhamento a incidência da doença for maior no grupo controle do que no grupo rastreamento, o aumento da incidência no grupo controle indica sobrediagnóstico. No PLCO, 20,7% dos cânceres de próstata detectados pelo rastreamento (ou 16,4% do total) foram sobrediagnosticados. No ERSPC, 50,4% dos cânceres de próstata detectados pelo rastreamento (ou 33,2% de todos os casos) foram sobrediagnosticados.<sup>3</sup>

O dilema do rastreamento do câncer é que, no momento da detecção, não sabemos quais casos vão evoluir ou não. Por essa maneira, a maioria das pessoas são tratadas. Homens sobrediagnosticados com câncer de próstata, por definição, não podem se beneficiar; apenas sofrem os danos do tratamento. Em torno de 17% dos homens submetidos a prostatectomia radical sofrerão dificuldades urinárias e 36% daqueles que recebem radiação sofrem com disfunção erétil.<sup>3</sup>

Outros danos do rastreamento incluem resultados falso positivos, procedimentos desnecessários, que pode ter complicações. No PLCO, aproximadamente 10% recebeu pelo menos um resultado PSA falso-positivo. Em comparação, no ERSPC foi 17%. Entre homens submetidos à biópsia, 60% (PLCO) e 75% (ERSPC) das biópsias foram negativas para câncer. Além disso, no PLCO ocorreram 20 complicações (infecções em sua maioria) a cada 1.000 biópsias.

## O rastreamento do câncer de mama

O dano por resultados falsos-positivos pode ser melhor compreendido analisando dados do rastreamento mamográfico para o câncer de mama. Nos Estados Unidos, em torno de 50% das mulheres rastreadas anualmente por dez anos recebem um pelo menos um resultado falso positivo e 19% são biopsiadas desnecessariamente.<sup>5</sup> Além disso, mulheres que recebem resultados falsos-positivos têm consequências psicológicas, podendo permanecer três anos após o resultado.<sup>6</sup> Assim como vimos para o rastreamento do câncer de próstata, o dano mais sério é o sobrediagnóstico e sobretreamento.

De acordo com ensaios clínicos, para uma morte por câncer de mama ser evitada, 2.000 mulheres entre 50 e 60 anos precisam ser rastreadas regularmente por 10 anos.<sup>7</sup> Além disso, aproximadamente 10 mulheres foram sobrediagnosticadas e sobretratadas com uma combinação de mastectomia, quimioterapia, radioterapia. Devido ao sobretratamento, o pressuposto de que o rastreamento mamográfico ofereceria um tratamento menos agressivo não foi confirmado. O rastreamento levou a um aumento de 30% de mastectomias realizadas (observe que uma mulher pode ser operada mais de uma vez).

É fundamental destacarmos que os grandes ensaios randomizados do rastreamento mamográfico foram realizados entre 1963 e 1990, acarretando duas implicações na interpretação dos seus resultados. A primeira é que os mamógrafos atuais têm maior sensibilidade. Com isso, é possível que os danos por resultados falsos-positivos, biópsias desnecessárias e sobrediagnósticos sejam maiores do que os reportados nos ensaios clínicos. Segundo, desde a realização dos ensaios clínicos, o tratamento do câncer de mama de melhorou, reduzindo a mortalidade do câncer de mama.<sup>8</sup> Para o rastreamento ser eficaz, a detecção precoce apenas não é suficiente; o tratamento precoce precisa ser mais eficaz do que o tratamento mais tarde. Por conta disso, o benefício atual do rastreamento do câncer de mama pode ser ainda menor do que o reportado nos ensaios clínicos.

A Tabela 1 resume os efeitos do rastreamento dos cânceres de próstata e de mama.

**Tabela 1.** Efeitos do rastreamento do câncer de próstata e mama

	Rastreamento do câncer de próstata (dados do estudo ERSPC3)	Rastreamento do câncer de mama (quando não indicado, dados da revisão da Cochrane7)
Faixa etária	50 – 69 anos	50 – 60 anos
Tempo de acompanhamento	13 anos	10 anos
Intervalo entre exames de rastreamento	4 anos	1 ano
Número de pessoas rastreadas	781	2000
Redução da mortalidade geral	Nenhuma	nenhuma

	Rastreamento do câncer de próstata (dados do estudo ERSPC3)	Rastreamento do câncer de mama (quando não indicado, dados da revisão da Cochrane7)
Redução da mortalidade específica do câncer	1	1
Número de pacientes sobrediagnosticados e sobretratados	27	10
Resultados falsos-positivos acumulados no período de acompanhamento	17% dos homens rastreados	50% das mulheres rastreadas <sup>5</sup>
Mortes pelo câncer rastreado em pessoas rastreadas	5	5

## Vieses em avaliações de rastreamento

Para entendermos vieses do rastreamento, vamos recorrer a uma comparação que bastante citada na literatura científica. Rudy Giuliani, ex-prefeito de Nova York, foi diagnosticado com câncer de próstata. Quando concorreu à presidência, fez um discurso, onde mostrou-se contente por morar nos Estados Unidos, já que a chance de ele sobreviver à essa doença no país era de 82%, enquanto na Inglaterra era 44%. A intenção do ex-prefeito era fazer um argumento político, argumentando que o sistema de saúde americano era melhor que a medicina “socializada” da Inglaterra.<sup>9</sup> Esse tipo de comparação não é válido para comprovar a tese do ex-prefeito, conforme veremos.

O primeiro erro é comparar pacientes rastreados com pacientes não rastreados: mais homens são rastreados para o câncer de próstata nos Estados Unidos do que na Inglaterra, onde é mais comum o diagnóstico desse câncer pelos sintomas. Essa comparação é um erro porque as pessoas que se submetem ao rastreamento de geralmente possuem uma saúde melhor do que aqueles que não fazem rastreamento.

Isso é conhecido como viés ou *efeito do rastreado saudável*.<sup>2</sup> A população rastreada tende a ter nível socioeconômico mais alto, praticar mais exercícios, se alimentar melhor, não fumar do que os não rastreados. Além de os não rastreados tenderem a ter mais problemas de saúde e sociais,

eles podem não ter recursos ou tempo para faltar ao trabalho para se submeter à realização de um exame. Eles também acreditam que a má saúde deles é algo inevitável e não querem se submeter a outro exame que pode lhes dar um novo problema de saúde.<sup>2</sup> O importante é notar que, sem um grupo controle adequado, é impossível saber se os efeitos foram causados pelo rastreamento ou se ocorreriam de qualquer de forma, já que as pessoas que fazem rastreamento formam um seletivo grupo de pessoas mais saudáveis.

Esse viés também pode ocorrer no sentido inverso: pacientes em alto risco para uma doença podem procurar mais o rastreamento do que os pacientes em risco normal. Por exemplo, pode haver uma tendência de pessoas com o medo de um forte histórico familiar buscarem o rastreamento. Como elas têm mais chances de desenvolver e morrer desse câncer, a comparação entre rastreados e não rastreados se torna enviesada.

A forma de resolver esse viés, como o leitor já deve ter imaginado, é através da randomização. A randomização serve justamente garantir que os grupos sejam comparáveis com exceção da intervenção sendo testada.

Porém, estudos clínicos randomizados podem ter problemas de randomização, o que pode levar a viés de seleção. Por exemplo, no ensaio clínico de Edimburgo do rastreamento com mamografia para o câncer de mama, 53% do grupo rastreado pertencia ao nível socioeconômico mais alto, comparado com 26% no grupo controle.<sup>7</sup> Além disso, o estudo reportou uma redução de 26% da mortalidade cardiovascular no grupo do rastreamento em relação ao controle. Como isso não é um efeito esperado do rastreamento com mamografia, é um forte indicativo do impacto do viés de seleção nos resultados do estudo.

Outras questões metodológicas relevantes são o sigilo de alocação, de forma a impedir o conhecimento da alocação dos sujeitos de pesquisa nos grupos de intervenção do ensaio clínico; e o cegamento dos investigadores responsáveis por aferir os desfechos do ensaio clínico sobre essa alocação dos participantes.

Giuliani, no seu discurso, se referiu a dados do ano 2.000, quando 49 de 100 mil ingleses foram diagnosticados com câncer de próstata, dos quais 44% (ou 28 a cada 100 mil) morriam 5 anos após o diagnóstico. Esse número é a sobrevivência em 5 anos. Já para os Estados Unidos, Giuliani citou que a sobrevivência em 5 anos era de 82%. Tais números claramente sugerem que a chance de sobreviver à doença nos Estados Unidos era o dobro das chances na Inglaterra.

A sobrevida em 5 anos – a proporção de pacientes que estão vivos 5 anos após o diagnóstico – provavelmente é a estatística mais utilizada para reportar o progresso no tratamento do câncer.<sup>10</sup> A sobrevida média – média de anos que os pacientes viveram após o diagnóstico – também é utilizada com frequência. Geralmente, maiores taxas de sobrevida refletem que os pacientes estão vivendo por mais tempo. Para calcularmos a sobrevida em 5 anos para o câncer de próstata, precisamos dividir o número de pacientes vivos cinco após o diagnóstico dessa doença pelo número total de pacientes diagnosticados com a doença.

Mesmo se tais taxas de sobrevida fossem obtidas em um ensaio clínico com perfeita randomização, garantindo que os grupos fossem iguais a não ser pelo rastreamento, elas não demonstram a eficácia do rastreamento. Voltando ao discurso de Giuliani, podemos comparar a mortalidade do câncer de próstata em ambos os países no mesmo período. Eis a surpresa: a mortalidade era idêntica: 26 mortes por câncer de próstata por 100 mil americanos e 27 mortes a cada 100 mil ingleses.<sup>9</sup>

Como é possível observamos um país com o dobro da taxa de sobrevivência de outro, mas com taxas iguais de mortalidade? Isso ocorre por três motivos.

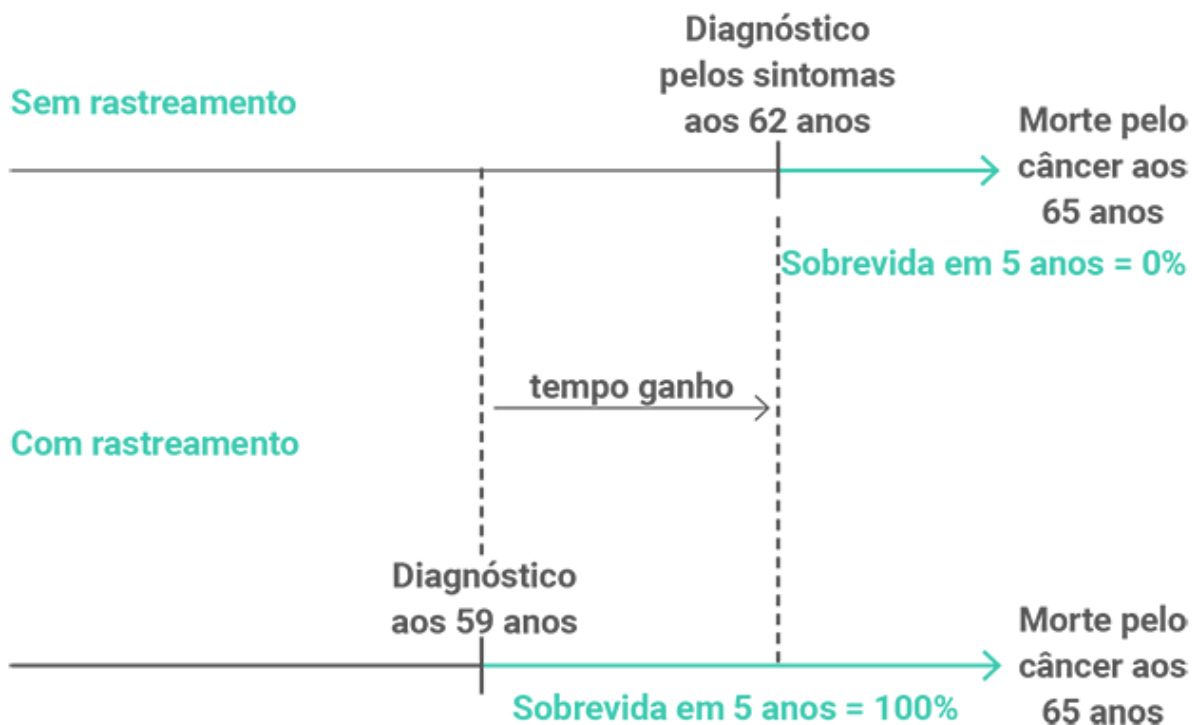
## Viés do tempo ganho (ou viés de tempo de antecipação)

O objetivo do rastreamento é antecipar o tempo do diagnóstico. Isso significa que, após a introdução de um rastreamento em uma população, é esperada uma redução na média de idade das pessoas diagnosticadas. A diferença de tempo entre o diagnóstico do câncer feito pelo rastreamento e feito pelos sintomas é o tempo ganho. Para ser eficaz, o rastreamento precisa introduzir um tempo ganho, mas o próprio tempo ganho envia a estatística de sobrevida.

Um exemplo hipotético é ilustrado na Figura 2. Suponha que, sem rastreamento, um determinado câncer é diagnosticado clinicamente aos 61 anos de idade levando os pacientes a óbito aos 65. Ou seja, os pacientes viveram quatro anos após o diagnóstico, conferindo uma sobrevida em 5 anos de 0%. Como o rastreamento busca a detecção precoce; suponha também que esse câncer é detectado pelo rastreamento aos 59 anos, mas os pacientes morrem também aos 65 anos. Como consequência, os pacientes rastreados morreram 6 anos após o diagnóstico, conferindo uma sobrevida

em 5 anos de 100%. Apesar de um aumento drástico no tempo de sobrevivência em 5 anos, nenhuma vida foi prolongada; os pacientes em ambos os casos morreram na mesma idade.

**Figura 2.**



Viés do tempo ganho (ou da antecipação do diagnóstico).

## Viés do tempo de duração (ou viés de duração)

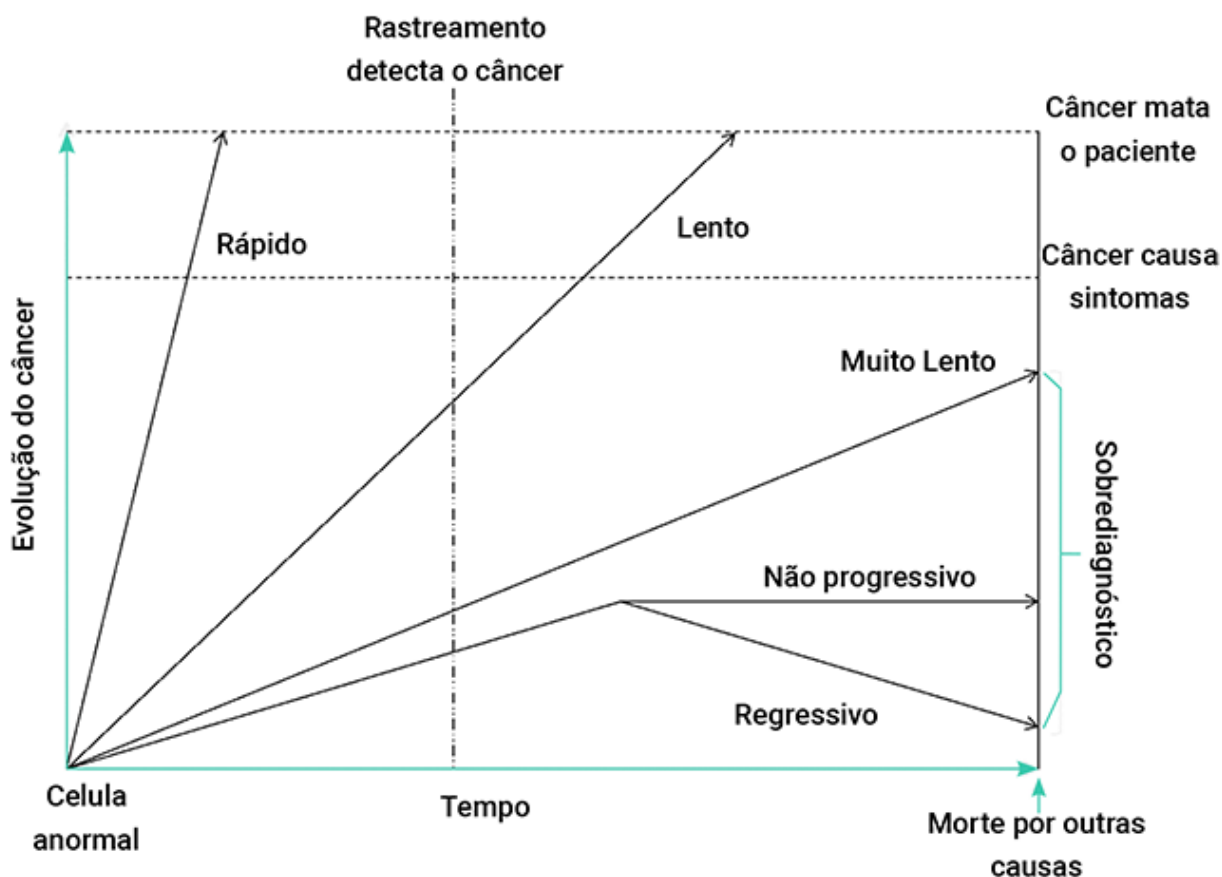
O segundo viés está relacionado com o fato de que o rastreamento é feito de forma periódica – uma vez ao ano, ou uma vez a cada cinco anos – e o câncer é uma doença heterogênea, com diferentes taxas de progressão.

Os tumores mais agressivos e mais letais normalmente não são detectados pelo rastreamento. Como crescem rápido, eles possuem uma fase assintomática curta, sendo mais provável causarem sintomas entre dois exames de rastreamento, sendo chamados de cânceres de intervalo. Similarmente, o rastreamento tende a detectar cânceres menos agressivos, com melhor prognóstico, já que eles progridem mais devagar e têm uma fase assintomática maior.<sup>11</sup>

Observe na Figura 3 que, quanto mais tempo um tumor existe antes de causar sintomas, maiores as chances de ser detectado pelo rastreamento.

Mesmo quando o rastreamento não tem benefício nenhum, casos detectados pelo rastreamento têm melhor prognóstico que casos diagnosticados clinicamente. Como consequência, taxas de sobrevida são mais altas para quem teve a doença detectada pelo rastreamento.

**Figura 3.**



A progressão heterogênea do câncer – adaptada.<sup>4,11</sup>

Qual seria a forma mais extrema do viés do tempo de duração? O rastreamento detectar um tumor com prognóstico tão favorável que ele não causaria sintomas nem a morte do paciente. Esse fenômeno é denominado sobrediagnóstico.

## Sobrediagnóstico

*Sobrediagnóstico* é o diagnóstico de uma anormalidade que, apesar de preencher critérios histopatológicos do câncer, não causaria sintomas nem a morte do paciente. Ocorre quando o tumor detectado não progrediria ou até

regrediria de forma espontânea. Ocorre também quando o rastreamento detecta tumores de progressão lenta o suficiente que o paciente morre de outras causas antes do câncer causar sintomas. Com isso, o sobrediagnóstico é mais comum em idosos, já que eles têm mais chances de morrer por outras causas antes de o tumor causar complicações.

Pode ser surpreendente falarmos de tumores que não progridem ou que regredem. Mas já sabemos alguns dos processos que levam a esse fenômeno, como o correto reconhecimento e contenção do tumor pelo sistema imunológico do paciente. Em outros casos, o tumor não recebe a quantidade que necessita de sangue para continuar crescendo.<sup>4</sup>

Esses casos não teriam sido diagnosticados na ausência do rastreamento. O dilema do rastreamento é que, como no momento do diagnóstico não há como saber quais casos progridem ou não, a maioria deles é tratado. Quando há sobrediagnóstico, tanto o numerador quanto o denominador da fórmula da sobrevida aumentam.

A Figura 4 ilustra essa situação. Imagine que numa população 1.000 pacientes foram diagnosticados clinicamente com câncer. Após cinco anos, 100 deles continuam vivos, o que corresponde a uma sobrevivência em cinco anos de 10%. Suponha também que, o rastreamento detecte corretamente os 1.000 pacientes que ficariam clinicamente doentes. Mas, além disso, considere que o rastreamento detectou adicionais 4.000 pacientes com câncer não progressivo, que representam sobrediagnóstico. Após cinco anos, temos 4.100 pacientes vivos do total de 5.000 pacientes, fazendo aumentar a sobrevivência em cinco anos para 82%. Portanto, mesmo sem aumento da sobrevida real, o sobrediagnóstico resulta em aumento espúrio do tempo de sobrevida. Em função disso, não se recomenda o uso de sobrevida como desfecho de estudos de rastreamento, embora esse equívoco seja muito comum na mídia, em campanhas de rastreamento e em até no ambiente acadêmico.<sup>12</sup>

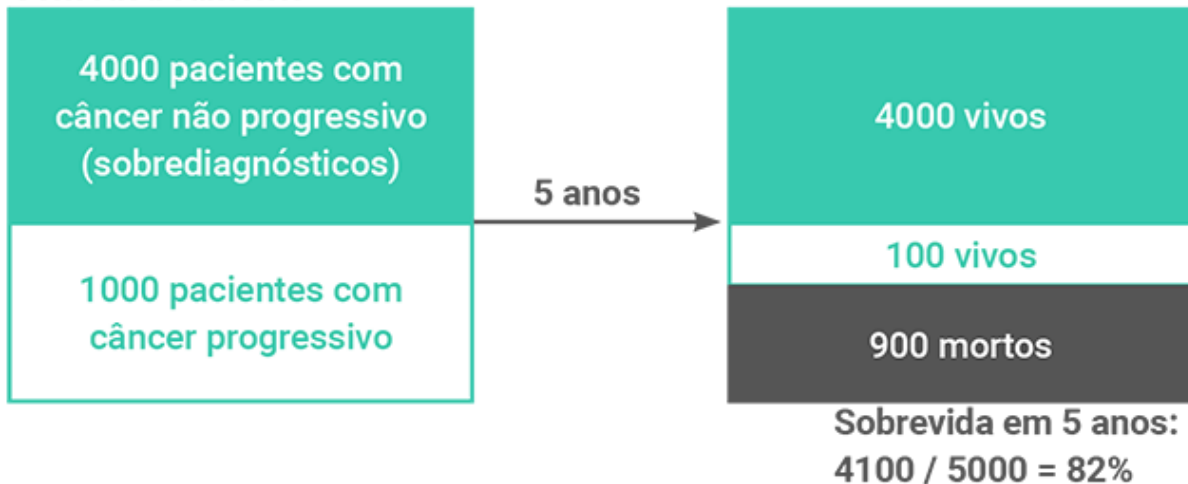
#### **Figura 4.**



### Sem rastreamento



### Com rastreamento



Viés no tempo de sobrevida induzido pelo sobrediagnóstico.

Note que o discurso de Giuliani, na realidade, demonstra o oposto do que ele desejava: diagnósticos e tratamentos de câncer próstata são muito mais frequentes nos Estados Unidos do que na Inglaterra, mas sem que isso tenha impacto na mortalidade da doença.

Perceba que o enviesamento da sobrevida pode ocorrer mesmo quando a mortalidade é maior no grupo de pessoas rastreadas. Em um ensaio clínico de rastreamento com radiografia de tórax para câncer de pulmão em fumantes, a sobrevida em 5 anos foi 35% para os rastreados e 19% para os não rastreados. Mesmo assim, a mortalidade, em números absolutos, foi levemente mais alta no grupo dos rastreados.<sup>13</sup> A não correlação entre aumento de sobrevida e taxas mortalidade foi confirmada em outro estudo para vinte tipos de câncer diferentes.<sup>14</sup>

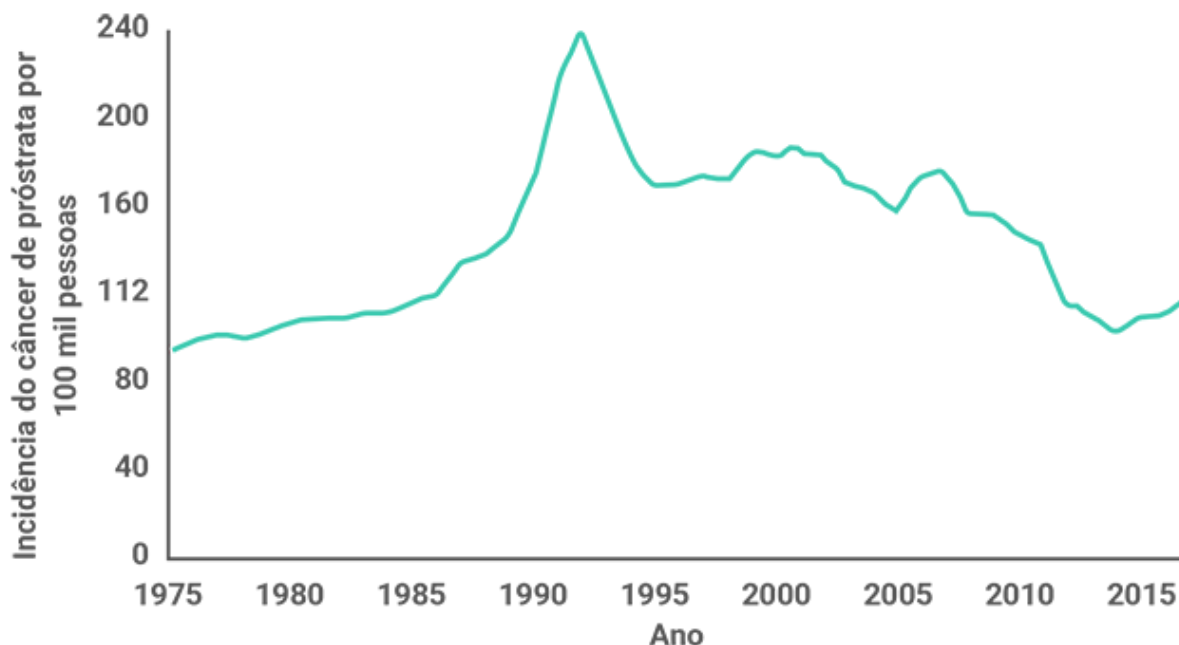
Comparar as estatísticas de sobrevida de um determinado câncer em épocas ou lugares diferentes é também problemático, já que os pacientes podem ter sido diagnosticados de diferentes maneiras, incluindo o uso rastreamento. Curiosamente, a mudança mais drástica nas taxas de sobrevida

em 5 anos, durante 1950-1995 nos Estados Unidos, foi no câncer de próstata – um período de crescente uso do rastreamento com PSA.

Embora os diversos vieses da sobrevida são amplamente documentados na literatura científica, os médicos não estão informados. Em uma pesquisa, 76% dos médicos participantes consideraram aumento na sobrevida como evidência de benefício do rastreamento.<sup>15</sup> Inclusive, o uso da sobrevida para supostamente reportar um benefício já foi encontrado em artigos publicados em revistas de prestígio, como *New England Journal of Medicine*.<sup>12</sup>

Em função do sobrediagnóstico, a incidência de diversos tipos de câncer aumentou consideravelmente devido aos esforços dos serviços de saúde em procurar por formas precoces da doença. Observe na Figura 5 a variação na incidência do câncer de próstata nos EUA entre 1975 e 2017. Nenhum processo carcinogênico ou biologia tumoral pode explicar tanto o rápido aumento quanto o declínio da incidência do câncer de próstata. Essa variação da incidência se deve à mudança nas práticas médicas: com o grande aumento acontecendo após a promoção do exame PSA e a queda após as recomendações contra o rastreamento.<sup>16</sup>

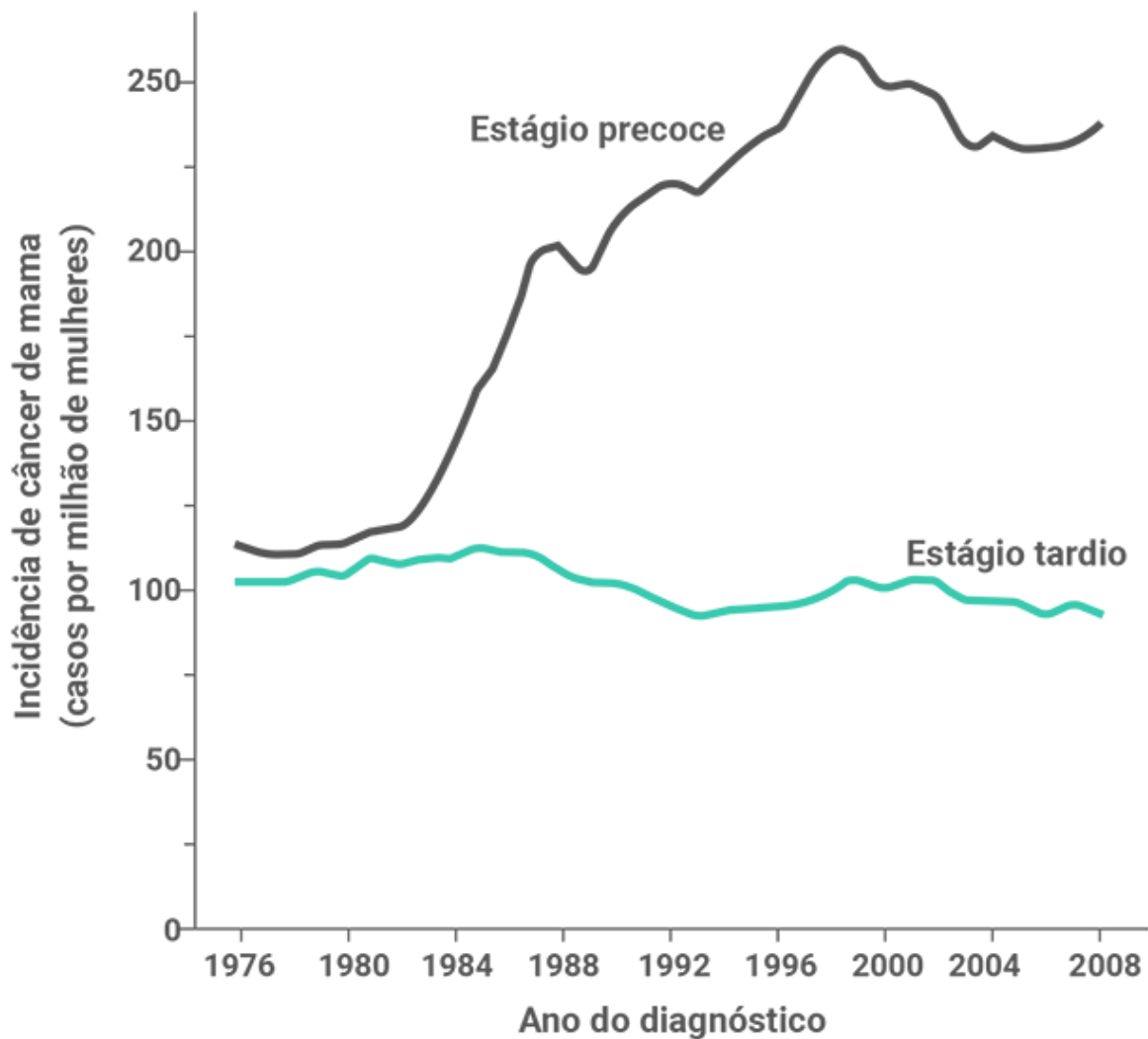
**Figura 5.**



Varição na incidência do câncer de próstata nos Estados Unidos entre 1975 e 2017 (Dados: *Surveillance, Epidemiology, and End Results Program*).

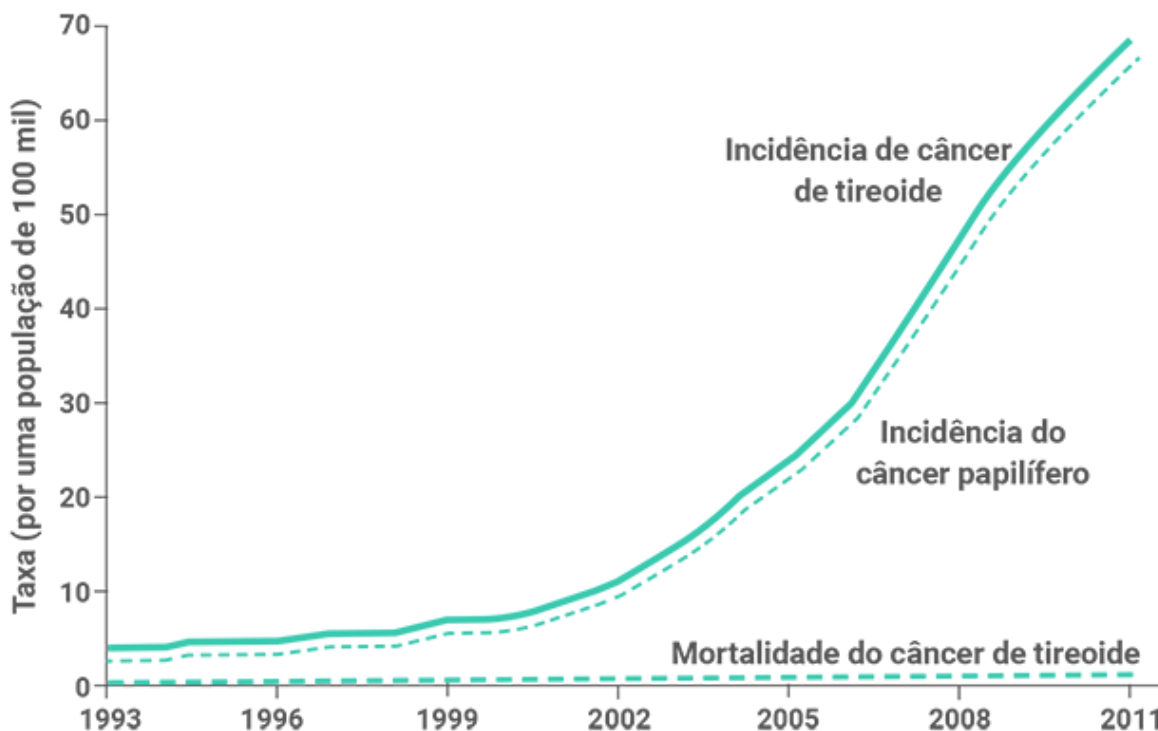
Outros exemplos conhecidos são o do rastreamento mamográfico do câncer de mama nos EUA e do rastreamento do câncer de tireoide na Coreia do Sul. A Figura 6 mostra que, apesar de um considerável aumento da incidência do câncer de mama diagnosticado no estágio inicial, não ocorreu a esperada diminuição de casos avançados, sugerindo sobrediagnóstico da doença na fase inicial.<sup>8</sup> Já a Figura 7 mostra que a incidência do câncer de tireoide na Coreia do Sul aumentou 15 vezes durante 11 anos. Esse grande aumento da incidência com a mortalidade permanecendo estável é um grande indicativo de sobrediagnóstico.<sup>17</sup>

**Figura 6.**



Varição da incidência do câncer de mama nos EUA de acordo com o estágio. As linhas cinza e laranja representam as incidências dos estágios precoce e tardio, respectivamente.<sup>8</sup>

**Figura 7.**

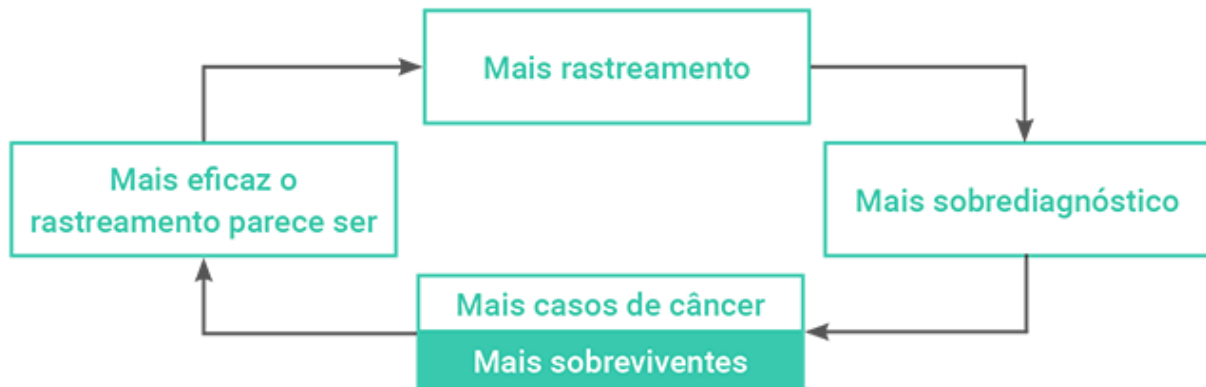


Aumento de incidência do câncer de tireoide em 11 vezes na Coreia do Sul, porém com mortalidade estável no período: sugestivo de sobrediagnóstico.<sup>17</sup>

Uma das consequências do aumento dos números de casos de câncer é que percepção do risco da população pode aumentar. Isso, por sua vez, pode fazer com que mais pessoas busquem o rastreamento ou aquelas pessoas que já são rastreadas podem fazer com mais frequência. Histórias de sobreviventes, cada vez mais comuns devido ao sobrediagnóstico, apoiadas por infladas taxas de sobrevida passam a impressão de que o rastreamento é eficaz mesmo em situações nas quais há ausência de benefício real, ou, ao menos, superestimam sua eficácia.

Isso pode formar um ciclo (Figura 8): mais rastreamento causa mais sobrediagnósticos e sobretratamento – mais casos de câncer na população e mais sobreviventes – aumentando ainda mais os esforços da própria população para ser rastreada, levando a mais sobrediagnósticos e sobretratamento. Os pesquisadores Angela Raffle e Muir Gray chamaram isso de **paradoxo da popularidade**: “Quanto maior o dano por sobrediagnóstico e sobretratamento causados pelo rastreamento, mais pessoas acreditarão que devem sua saúde, ou até suas vidas, ao programa”.<sup>2</sup>

**Figura 8.**



O paradoxo da popularidade do rastreamento. Quanto mais sobrediagnóstico um rastreamento causa, mais popular ele é.

Fora de um programa de rastreamento organizado, esse fenômeno tende a ser ainda maior, pois há uma tendência a sobrerastreamento fora da periodicidade e faixa etária onde o rastreamento é mais eficaz, aumentando a detecção de cânceres ou, em alguns casos, lesões precursoras destinadas que não evoluiriam clinicamente em vida. É sabido, por exemplo, que o rastreamento oportunístico do câncer de próstata é ainda mais ineficiente do que o realizado nos ensaios clínicos, com menor eficácia, quase dobrando o número necessário de homens a rastrear para gerar benefício, além de gerar mais sobrediagnóstico.<sup>18</sup> Outro problema importante do rastreamento oportunístico é a maior dificuldade de adesão às diretrizes de população-alvo e periodicidade, essenciais para garantir um balanço favorável entre riscos e benefícios de um rastreamento.<sup>19</sup> Por exemplo, o rastreamento oportunístico do câncer do colo do útero induz o sobrerastreio de mulheres jovens e detecção de lesões precursoras que iriam regredir espontaneamente.<sup>20</sup>

## Vieses da mortalidade específica do câncer

A mortalidade específica da doença é o desfecho mais utilizado em ensaios clínicos randomizados de rastreamento. O rastreamento busca reduzir a mortalidade específica pela doença. O uso desse desfecho, tem algumas limitações, pois assume que a causa da morte pode ser determinada corretamente. Há dois vieses importantes podem ocorrer no desfecho da mortalidade específica da doença em um estudo de rastreamento.

O primeiro viés (*slippery linkage to screening*) ocorre porque a mortalidade referente ao tratamento do câncer em pacientes rastreados

frequentemente não é levada em consideração no desfecho da mortalidade específica do câncer. De forma mais simples, o tratamento aumenta mortalidade por outras causas. Como exemplo, considere uma mulher rastreada para câncer de mama que morreu de infarto como consequência tardia da radioterapia. Idealmente, em estudos randomizados de rastreamento, a mortalidade por câncer de mama deveria contabilizar mortes como essa. No entanto, não é isso que ocorre. O problema ocorre até mesmo em casos em que a ligação entre rastreamento e óbito são mais simples do que o exemplo anterior. Esse tipo de viés fica ainda mais difícil de evitar quando estão envolvidos indivíduos muito idosos nos quais são muitas as causas de mortalidade competitivas, como tipicamente ocorre no câncer de próstata, tornando ainda mais difícil a correlação entre os procedimentos gerados pelo rastreamento e o desfecho adverso observado.<sup>21</sup> Nesses casos o rastreamento é favorecido pelo viés.

O segundo viés (*sticky diagnosis from screening*) ocorre quando mortes por outras causas no grupo rastreamento são falsamente atribuídas à doença. Quando esse viés ocorre em um estudo randomizado, a mortalidade específica da doença tende a ser maior no grupo rastreamento do que no grupo controle, enviesando o resultado contra o rastreamento.

A mortalidade geral não é afetada por esses vieses. Dessa maneira, é considerada por muitos como o melhor desfecho para avaliar a eficácia de um rastreamento. No entanto, os estudos de rastreamento, como envolvem muitas pessoas saudáveis e de baixo risco, geralmente precisariam de milhões de pessoas para que tenha poder suficiente para detectar possíveis diferenças pequenas na mortalidade geral.

## Conclusão

A indicação de rastreamento, de doenças em geral e de câncer em particular, deve ser baseada nas melhores evidências disponíveis de forma que todos os vieses aqui apresentados sejam minimizados ao máximo. Para evitá-los, as melhores diretrizes de rastreamento disponíveis utilizam revisões sistemáticas de ensaios clínicos randomizados, com avaliação da qualidade das evidências, utilização dos desfechos de mortalidade e consideram o balanço entre riscos e benefícios do rastreamento em suas recomendações. Recomendações de rastreamento devem expressar indicações precisas de método, população-alvo e periodicidade e devem estar inseridas em programas organizados que assegurem sua eficácia e a

continuidade e a integralidade do cuidado com acesso a todos os processos subsequentes ao rastreamento.

## Referências bibliográficas

1. Han PK. Historical changes in the objectives of the periodic health examination. *Ann Intern Med.* 1997 Nov;127(10):910-7.
2. Raffle AE, Gray J. M. *Screening: Evidence and Practice.* Oxford: Oxford University Press; 2007.
3. Fenton JJ, Weyrich MS, Durbin S, Liu Y, Bang H, Melnikow J. Prostate-specific antigen-based screening for prostate cancer evidence report and systematic review for the us preventive services task force. *JAMA – J Am Med Assoc.* 2018;319(18):1914-31.
4. Welch HG, Black WC. Overdiagnosis in cancer. *J Natl Cancer Inst.* 2010;102(9):605-13.
5. Elmore JG, Barton MB, Mocerri VM, Polk S, Arena PJ, Fletcher SW. Ten-year risk of false positive screening mammograms and clinical breast examinations. *N Engl J Med.* 1998 Apr;338(16):1089-96.
6. Brodersen J, Siersma VD. Long-term psychosocial consequences of false-positive screening mammography. *Ann Fam Med.* 2013;11(2):106-15.
7. Gøtzsche PC, Jørgensen KJ. Screening for breast cancer with mammography. *Cochrane Database Syst Rev.* 2013;2013(6).
8. Bleyer A, Welch HG. Effect of Three Decades of Screening Mammography on Breast-Cancer Incidence. *N Engl J Med [Internet].* 2012 Nov 22 [cited 2021 Jan 26];367(21):1998–2005 [[acesso em: 6 abr. 2021](#)].
9. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients make sense of health statistics. *Psychol Sci Public Interes Suppl.* 2007;8(2):53-96.
10. Wegwarth O, Gaissmaier W, Gigerenzer G. Deceiving numbers: Survival rates and their impact on doctors' risk communication. *Med Decis Mak.* 2011;31(3):386-94.
11. Carter SM, Barratt A. What is overdiagnosis and why should we take it seriously in cancer screening? *Public Heal Res Pract.* 2017;27(3):3-7.
12. Gigerenzer G, Wegwarth O. Five year survival rates can mislead. *BMJ.* 2013;346(7893):1-2.
13. Woloshin S, Schwartz LM, Welch HG. *Know Your Chances: Understanding Health Statistics [Internet].* Berkeley(CA): University of California Press; 2008 [[acesso em: 6 abr. 2021](#)].
14. Welch HG, Schwartz LM, Woloshin S. Are increasing 5-year survival rates evidence of success against cancer? *J Am Med Assoc.* 2000;283(22):2975-8.

15. Wegwarth O, Schwartz LM, Woloshin S, Gaissmaier W, Gigerenzer G. Do physicians understand cancer screening statistics? A national survey of primary care physicians in the United States. *Ann Intern Med.* 2012;156(5):340-9.
16. Welch HG, Brawley OW. Scrutiny-dependent cancer and self-fulfilling risk factors. *Ann Intern Med.* 2018;168(2):143-4.
17. Ahn HS, Kim HJ, Welch HG. Korea's Thyroid-Cancer "Epidemic" — Screening and Overdiagnosis. *N Engl J Med [Internet].* 2014;371(19):1765-7 [[acesso](#) em: 6 abr. 2021].
18. Arnsrud Godtman R, Holmberg E, Lilja H, Stranne J, Hugosson J. Opportunistic testing versus organized prostate-specific antigen screening: outcome after 18 years in the Göteborg randomized population-based prostate cancer screening trial. *Eur Urol.* 2015 Sep;68(3):354-60.
19. Migowski A, Dias MBK, Nadanovsky P, Silva GA e, Sant'Ana DR, Stein AT. Diretrizes para detecção precoce do câncer de mama no Brasil. III – Desafios à implementação. *Cad Saude Publica [Internet].* 2018;34(6) [[acesso](#) em: 6 abr. 2021].
20. Raffle AE, Alden B, Quinn M, Babb PJ, Brett MT. Outcomes of screening to prevent cancer: analysis of cumulative incidence of cervical abnormality and modelling of cases and deaths prevented. *BMJ.* 2003 Apr;326(7395):901.
21. Juffs HG, Tannock IF. Screening trials are even more difficult than we thought they were. *J Natl Cancer Inst.* 2002;94(3):156-7.



# 13. CIRURGIA BASEADA EM EVIDÊNCIAS

## A cirurgia baseada em dogma

Daniel J. Boortstin, historiador americano no início do século XX, certa vez disse: “O maior obstáculo para a descoberta não é a ignorância; é a ilusão do conhecimento”.<sup>1</sup> Historicamente, o conhecimento médico na área da cirurgia foi construído devido ao positivismo, ou seja, o conhecimento “genuíno” cirúrgico é derivado exclusivamente da experiência do cirurgião, de suas ações com suas propriedades e relações. Assim, as informações derivadas da experiência sensorial do cirurgião, conforme interpretadas por meio de suas ações e da lógica, constituem a fonte genérica de todo conhecimento tido como certo. O positivismo na cirurgia, portanto, consiste na assumpção que todo conhecimento “genuíno” nessa área é arquitetado a *posteriori*.

Essa forma de pensar parecia assertiva, uma vez que a atuação do cirurgião era solicitada em pacientes com doenças graves, agudas e muitas vezes traumáticas, onde a ação rápida e precisa de uma intervenção poderia mudar um desfecho fatal. Situações como estas ocorreram por séculos desde o Egito antigo até o século XX, especialmente nas duas grandes guerras mundiais, onde a arte cirúrgica se resumia à fixação de fraturas expostas, amputação de membros, traqueostomia em vias aéreas obstruídas, compressões hemostáticas, laparotomias por perfurações de vísceras e cauterização de lesões.<sup>2-3</sup> Essas técnicas forneciam resultados extraordinários, entretanto, constituem um princípio chamado de plausibilidade extrema (princípio do “paraquedas”),<sup>4</sup> que consiste na experimentação do óbvio, não sendo necessário experimentos controlados para avaliar sua eficácia.

Por sua vez, esse paradigma positivista, envolvendo a ação lógica do cirurgião, também foi usado no intuito de tratar inúmeras doenças crônicas, cirúrgicas ou não, que motivaram intervenções fúteis que pareciam dar certo. Um exemplo é a sangria, procedimento realizado por séculos (desde Hipócrates) para tratar diversos tipos de doenças, consistindo em uma flebotomia para retirar sangue dos pacientes, devido à uma teoria lógica

vigente (teoria do equilíbrio dos quatro humores). Uma vítima deste procedimento foi o primeiro presidente americano, George Washington em dezembro de 1799, que, devido ao que parecia ser uma infecção de garganta, foi submetido a uma sangria, vindo a falecer dias depois.<sup>5-6</sup> Claro que tal procedimento soa absurdo nos dias de hoje, mas na época, era algo considerado razoável.

Evidentemente o exposto é um simples epítome de como surgiu o conhecimento na história da cirurgia, perpetuando uma tradição de treinamento vinculada a um ecossistema hierárquico, onde o professor determinava através de sua experiência lógica como deveria ser feito o diagnóstico, tratamento e o tipo de técnica cirúrgica empregada. Tal forma de pensar está no cerne mental dos cirurgiões, sendo ensinada e perpetuada por gerações, com poucos questionamentos, constituindo axiomas imutáveis até o século XX.

## O problema do dogmatismo

A manutenção dessa tradição associada ao comportamento humano (repleto de vieses cognitivos e uso de intuição)<sup>7</sup> gera alguns problemas. Um deles é a utilização de atalhos mentais (heurísticas)<sup>8-11</sup> para solucionar problemas de difícil solução, quando buscamos tratar um paciente que constitui um ser complexo biologicamente, imerso em um ambiente multifatorial, com inúmeras variáveis que podem influenciar em um resultado. Por exemplo: um cirurgião em formação, ingênuo (não conhecedor das evidências) e dogmático, que ao realizar uma cirurgia de hérnia não utiliza antibiótico, e por acaso o paciente apresenta uma infecção na ferida operatória, irá pensar que o próximo paciente com hérnia ao ser operado apresentará alta probabilidade de evoluir também com infecção (quando na realidade o risco é de 2%).<sup>12-16</sup> Esse é um exemplo da heurística de disponibilidade, que ocorre quando o “bom senso” trai nossa consciência, em um processo de julgamento onde estimamos a frequência ou a probabilidade de um evento, pela imediata facilidade com que situações similares vêm à mente.

Portanto, essa forma de pensar nos confunde, nos causa a ilusão de que achamos que sabemos de algo (quando na realidade não sabemos), fazemos falsas associações e nos convencemos disso através de explicações mecanicistas após a ocorrência de eventos (*a posteriori*). Isso nos remete à importância do amadurecimento da arte de pensar, que graças à revolução

cognitiva, fez abrir nossos olhos, permitindo reavaliar nossas ilusões intuitivas ou dogmatismo (no sentido não religioso, aceitando axiomas sem questionamentos). Esse processo permitiu o surgimento da curiosidade científica, que se tornou a maneira mais sofisticada de buscar a verdade objetiva, ou seja, como de fato as coisas em nosso mundo funcionam.

## A mudança do paradigma tradicional para o paradigma científico na cirurgia

Um novo paradigma para a prática médica na cirurgia emerge (ou deveria emergir) com a compreensão de três importantes figuras:

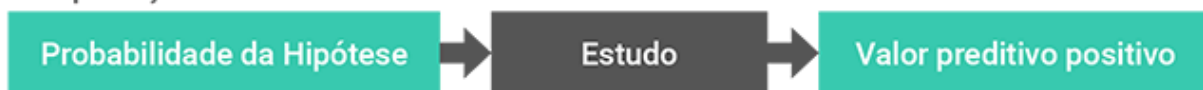
- Thomas Bayes:<sup>17-18</sup> reverendo que no século XVIII descreveu o conceito de probabilidade condicional. A interpretação do raciocínio clínico mediado por testes diagnósticos (anamnese, exame físico, exames complementares) deve ser condicionada à probabilidade pré-teste (ou pela prevalência da doença naquela população). Assim também como a interpretação do resultado de um estudo deve ser interpretado através da análise Bayesiana (Figura 1; ver capítulo 5).

**Figura 1.**

Raciocínio diagnóstico:



Interpretação de estudos:



Fluxograma do raciocínio clínico Bayesiano e da interpretação Bayesiana de estudos científicos.

- Sir Karl Raimund Popper: um dos filósofos cientistas mais importantes do século XX, descreveu a importância do racionalismo crítico, rejeitando o método científico tradicional baseado na observação indutiva, com conclusões geradas a posteriori. Ele criou o método científico contemporâneo, baseado

na dedução. Segundo Popper, toda hipótese gerada *a priori* deve ser testada por experimentos científicos controlados, no intuito de buscar sua refutação (falseabilidade).<sup>19-20</sup>

- David Lawrence Sackett: médico americano-canadense, fundador do conceito de Medicina Baseada em Evidências (professor das Universidades de McMaster e Oxford), que em 1996 a definiu como: “o uso consciente, explícito e judicioso da melhor evidência disponível na tomada de decisões sobre o cuidado individual do paciente”.<sup>21</sup>

A falta do entendimento destes conceitos fez com que inúmeros cirurgiões aprendessem práticas médicas ditas “inquestionáveis” baseadas na indução vinculadas à tradição (Tabela 1). É através destes conceitos que mudamos o paradigma da cirurgia baseada em dogma ou tradição, e passamos a aplicar o conhecimento científico testável, aproximando a ciência da clínica cirúrgica assistencial. Surge uma ferramenta científica para o cirurgião: a medicina baseada em probabilidade.

**Tabela 1.**

Práticas médicas tradicionais
Preparo de cólon reduz a morbidade de cirurgias colorretais
Antibióticos profiláticos orais reduzem infecção em toda cirurgia
Sonda nasogástrica descomprime o trato digestivo e reduz a incidência de fístulas
Atrasar uma apendicectomia em 12 horas aumenta o risco de perfuração ou abscesso
Analgésico antes do exame físico abdominal mascara o diagnóstico de abdome agudo
Introdução de dieta no pós-operatório de cirurgias abdominais deve ser gradativa
Cateter de Swan-Ganz reduz mortalidade em pacientes cirúrgicos graves
Nutrição parenteral total dias antes da cirurgia previne complicações
Uso rotineiro de drenos em cirurgias abdominais reduz a incidência de infecções
Aumentar a temperatura da sala reduz a incidência de infecções
Não interromper o uso de AAS antes de cirurgias aumenta o risco de sangramentos

## Práticas médicas tradicionais

Usar rotineiramente antibióticos na pancreatite aguda

Alguns exemplos de práticas médicas tradicionais em cirurgia que foram testadas e refutadas por experimentos científicos de boa qualidade.

## O raciocínio clínico probabilístico na cirurgia

Durante a graduação de Medicina no Brasil, aprendemos que uma anamnese bem feita é aquela produzida cautelosamente, com extensão de um calhamaço, inundada de detalhes e inúmeras perguntas sobre: sinais e sintomas, interrogatório de sistemas, antecedentes de todos os tipos e claro, um exame físico detalhado para triunfamos com um diagnóstico sindrômico que gera mais dúvida do que a obra *Ulysses*, de James Joyce. Depois, aprendemos que os exames complementares são solicitados diante dessa dúvida, e de forma determinística, transformará a dúvida em um diagnóstico incontestavelmente certo. Agora nos resta apenas tratar nosso paciente, utilizando conceitos fisiopatológicos tradicionais (como ciclo de Krebs, farmacocinética e técnicas operatórias), criando assim, explicações mecanicistas para nortear nossas intervenções, posteriormente “confirmadas” por evidências de péssima qualidade (viés de confirmação).

Essa perspectiva é inadequada, pois o raciocínio clínico e eventuais intervenções devem ser norteados por probabilidade. O cirurgião deve, por meio da queixa principal do paciente, estimar a probabilidade pré-teste (prevalência) e posteriormente, através de uma anamnese e exame físico acurados, moldar essa probabilidade à medida que o paciente responde sim (ou não) para as suas indagações, ou algum sinal clínico é encontrado no exame físico. É utilizando o conceito de probabilidade condicional, que chegamos a uma probabilidade pós-teste, que nada mais é que o valor preditivo positivo (VPP) clínico de uma determina doença.

Tudo isso parece confuso, entretanto em sua brilhante frase no século passado, William Osler disse: “a medicina é a ciência da incerteza e a arte da probabilidade”.<sup>22-23</sup>

O objetivo deste capítulo não é criar “*guidelines*” ou estratégias que devem ser utilizados como regras absolutas em todos os casos clínicos, mas sim, despertar a forma mais adequada de chegar a um diagnóstico mais

preciso do paciente. Vamos exemplificar: imagine um paciente do sexo masculino, de 25 anos, que dá entrada no pronto-socorro de um hospital se queixando de dor abdominal. Imagino que este paciente pode ser portador de uma apendicite aguda. E agora?

## Qual a prevalência?

A probabilidade pré-teste ou prevalência estimada é facilmente obtida pela literatura, encontrada em estudos epidemiológicos. Em média, 10% dos pacientes procuram o serviço de emergência com dor abdominal, e desses 5%-7% são devido à apendicite aguda.<sup>24-30</sup>

## Quais testes devem ser aplicados?

A anamnese e o exame físico nada mais são do que testes diagnósticos que também apresentam sensibilidade e especificidade, por isso, é importante fazer perguntas precisas e dirigidas para a queixa do paciente. Neste caso, a caracterização da dor é imprescindível: localização, tipo, intensidade, irradiação, fatores desencadeantes e alívio. Sintomas como anorexia, náuseas e vômitos também apresentam relativa precisão neste caso, e todas essas informações quando reunidas e presentes apresentam boa acurácia clínica. Um modelo probabilístico diagnóstico para este caso, foi publicado por Alfredo Alvarado<sup>31</sup> e revalidado por outros autores<sup>32</sup> (Figura 2).

**Figura 2.**

Sintomas	Pontuação	Sensibilidade (%)	Especificidade (%)
Naúseas	1	74	36
Anorexia	1	61	72
Dor migratória	1	69	84
Sinal			
Desconforto em QID	2	100	64
Dor em defesa no QID	1	63	69
Temperatura > 37,3°C	1	67	79
Laboratório			
Leucocitose	2	93	38
Desvio à esquerda	1	79	40

**Escore Alvarado 0-4**  
Baixo risco

**Escore Alvarado 5-6**  
Risco intermediário

**Escore Alvarado 7-10**  
Alto risco

Escore de Alvarado. QID = Quadrante inferior direito.

Imaginemos dois cenários no exemplo dado em um paciente com provável apendicite aguda:

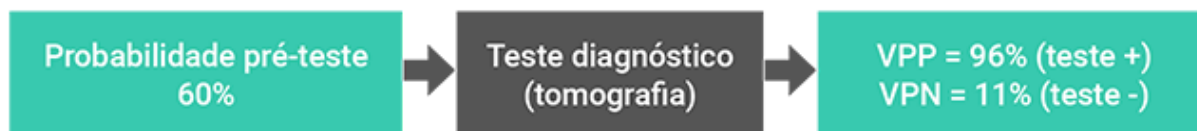
- a. A probabilidade pré-teste do paciente é 5%. O paciente apresenta dor migratória iniciada há 2 dias, para o quadrante inferior direito, anorexia e náuseas. Ao examinar o paciente percebemos dor intensa à palpação neste quadrante com sinal de descompressão brusca (sinal de Blumberg). Temos, portanto, 6 pontos no escore de Alvarado, com risco intermediário de apendicite aguda e uma probabilidade pós-teste (VPP) de aproximadamente 60%.
- b. A probabilidade pré-teste do paciente é 5%. O paciente apresenta o mesmo quadro clínico, porém solicitamos um hemograma e é identificado 15.000 leucócitos/mm<sup>3</sup> associado à 15% de bastões. Agora temos 9 pontos no escore de Alvarado, onde o paciente apresenta alto risco de apendicite aguda, e um VPP de aproximadamente 90%.

Perceba, nesse caso, que não é necessário realizar indagações fúteis, como, por exemplo: “Você tem prurido?”; “Houve queda do seu cabelo?”; “Você tem dor articular?”; “Você já tomou banho de rio na infância?”; “Pode me falar sobre o seu calendário vacinal ou qual o tipo de parto que você nasceu?”. São perguntas inúteis que não apresentam acurácia na resolução do problema clínico em questão.

Assim seguimos em frente. No cenário (b) ficamos extremamente confortáveis em indicar um tratamento para este paciente, pois o quadro clínico apresentado, associado a uma anamnese acurada utilizando a probabilidade condicional Bayesiana, moldou tão bem a probabilidade pré-teste, que chegamos a um alto VPP. Por sua vez, no cenário (a) ficamos desconfortáveis para tratar o paciente, uma vez que há imprecisão no diagnóstico sendo, portanto, fundamental aumentar o valor preditivo. Nesse caso, se torna importante a aplicação de um exame complementar acurado para aumentar a precisão do diagnóstico, como por exemplo a utilização de uma tomografia computadorizada de abdome (sensibilidade de 92% e especificidade de 95%).<sup>24, 29, 33, 34</sup> Com a aplicação deste teste, os 60% de VPP se tornam a probabilidade pré-teste da tomografia de abdome, que deve moldar esta probabilidade, aumentando ou reduzindo o risco de doença a depender do resultado do exame (Figura 3):

**Figura 3.**

Raciocínio diagnóstico:



Fluxograma do raciocínio clínico Bayesiano e da interpretação Bayesiana dependendo do resultado do teste diagnóstico.

Perceba que neste cenário, que o teste diagnóstico acurado adequadamente aplicado aumenta ou diminui consideravelmente a probabilidade pós-teste, nos deixando confortáveis para indicar ou não um tratamento.

Interessante, não? Agora imaginemos o mesmo caso clínico, do mesmo paciente do sexo masculino de 25 anos, com queixa de dor abdominal há 2 dias que dá entrada no pronto-socorro. Porém quem o avalia é um médico ingênuo, mecanicista e preso ao pensamento determinístico equivocado da

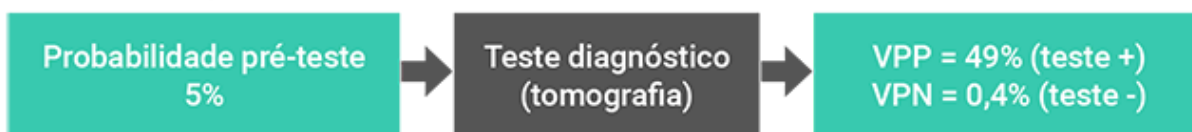


medicina. Ao fazer uma anamnese e exame físico inadequados, sem o exercício probabilístico, geramos dúvida diagnóstica e ficamos na esperança que o exame complementar definirá o diagnóstico certo. Essa forma de pensar é equivocada, vejamos porquê.

Mesmo ao fazer as mesmas indagações na anamnese e examinar o paciente, entretanto sem compreender o conceito de probabilidade condicional, ficamos na dúvida qual seria a probabilidade pré-teste ou pós-teste do paciente. Se utilizamos a prevalência de apendicite aguda estimada, com uma história clínica pouco acurada, veja o que ocorre (Figura 4).

**Figura 4.**

Raciocínio diagnóstico determinístico:



Fluxograma do raciocínio clínico determinístico associado à interpretação Bayesiana dependendo do resultado do teste diagnóstico.

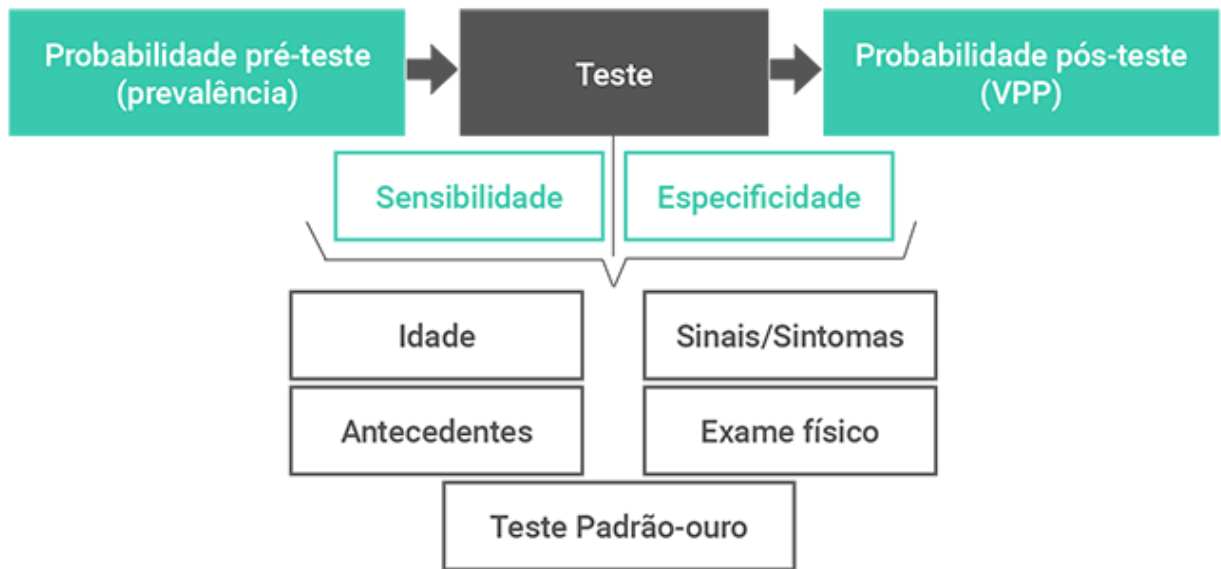
Vemos este cenário rotineiramente na prática assistencial, e é interessante como a maioria dos médicos tende a superestimar a efetividade de suas intervenções,<sup>35</sup> incluindo o resultado dos testes diagnósticos. Neste caso, provavelmente um número razoável de médicos ingenuamente acharia que uma tomografia com resultado positivo, definiria o diagnóstico com uma certeza platônica (90%-100% de VPP) considerando apenas o resultado do exame, independentemente do exercício probabilístico.

## Fluxograma do raciocínio clínico probabilístico

O exemplo do caso clínico em questão facilita o processo de aprendizagem para o entendimento do conceito de probabilidade na área da cirurgia, dispondo de um modelo preditor diagnóstico pragmático (escore de Alvarado). Entretanto, como devemos proceder nas situações clínicas da prática cirúrgica onde não foram desenvolvidos modelos probabilísticos? A resposta é simples: exatamente da mesma forma, mesmo em problemas clínicos complexos.

Uma ideia do que discutimos pode ser observada na Figura 5.

**Figura 5.**



Fluxograma do raciocínio clínico probabilístico.

## Cirurgia baseada em evidências

Enfatizo que o propósito deste capítulo não é criticar o ensino médico tradicional, ele foi importante para aprimorarmos e exercemos uma prática cirúrgica que ajudou a tratar muitos pacientes durante nossa história. No entanto, o amadurecimento da arte de pensar, entendendo melhor o que caracteriza uma evidência científica de qualidade e como aplicá-la na atividade assistencial cirúrgica, constitui a melhor maneira de sermos mais efetivos e humanizados, criando uma verdadeira reforma iluminista na cirurgia. Um cuidado precisa ser tomado: devemos reconhecer que a experiência clínica é fundamental para que a prática cirúrgica não corra o risco de se tornar tiranizada por evidências, já que uma evidência científica de ótima qualidade pode ser inaplicável ou inadequada para um paciente individual. Por exemplo: um paciente muito idoso, com muitas comorbidades, portador de insuficiência arterial coronariana, deveria ser submetido a uma cirurgia cardíaca convencional para revascularização do miocárdio? Tal procedimento pode ser inadequado, pois pode trazer um dano maior que a sua doença.

Também não é nosso intuito trazer burocracia ou reduzir a arte de pensar a um “*checklist*” simplista. Assim, filosoficamente, a cirurgia baseada em evidências poderia ser definida como a interação de evidências científicas de qualidade (pesquisas científicas relevantes para diagnóstico, prognóstico e tratamento) associada à expertise clínica do cirurgião (experiência clínica,

habilidade cirúrgica e cognição sofisticada) aplicados ao paciente individual (valores e preferências do mesmo).

Definições ou conceitos facilitam o processo de aprendizado, entretanto, não deveria existir o termo “baseado em evidências”, afinal, só existe um tipo de especialidade cirúrgica: a norteada por evidências científicas de qualidade.

Precisamos tomar cuidado com expressões encontradas na literatura como “a melhor evidência disponível”, por uma única razão: a melhor evidência disponível pode ser uma evidência científica de péssima qualidade (alto risco de erro aleatório ou presença de erros metodológicos – vieses).

Espero que o exposto tenha despertado a curiosidade do leitor em aprofundar a busca pelo conhecimento, porém imagino que uma pergunta surja: como devemos aplicar o conceito científico a problemas clínicos práticos na cirurgia? Existem modelos práticos que ajudam no processo, com os seguintes passos:<sup>3, 36-40</sup>

- a. Defina uma pergunta para o caso clínico em questão.
- b. Procure por evidências científicas de qualidade.
- c. Avalie criticamente a(s) evidência(s) científica(s) encontrada(s).
- d. Aplique o resultado encontrado no seu paciente individual, avaliando seus valores e preferências.
- e. Observe o resultado.

Imagine o seguinte caso clínico: uma paciente do sexo feminino, de 50 anos de idade, deu entrada no hospital com um quadro de pancreatite aguda grave, necrotizante, com intensa resposta inflamatória sistêmica. Ao ser tratada na UTI com medidas clínicas respaldadas por evidências científicas (hidratação, jejum e analgesia), surge uma dúvida recorrente na prática clínica: devemos ou não utilizar antibióticos neste caso?

A pergunta já está estabelecida: terapia com antibióticos na pancreatite aguda grave previne infecção? Terapia com antibióticos na pancreatite aguda grave previne morte?

Há um racional nesse tipo de conduta, uma vez que estamos diante de um grave processo inflamatório retroperitoneal, com necrose do tecido glandular pancreático, podendo evoluir com complicações infecciosas graves e eventual morte dessa paciente. Estudos observacionais mostram que parece haver um benefício no uso de antibióticos na pancreatite aguda grave, porém, sabemos que este tipo de evidência é de baixa qualidade para avaliar

eficácia, pois apresenta fatores de confusão que incorrem em falsas associações.

Com essas perguntas objetivas, faremos uma busca por evidências científicas de qualidade, que avaliam de forma adequada a eficácia de intervenções médicas. Existem inúmeros sites de busca, como “PubMed.gov”, “SciELO.br”, “Cochranelibrary.com” e outros. As evidências científicas de qualidade para esta pergunta são os ensaios clínicos randomizados, desde que apresentem baixo risco de erros aleatório e sistemático.

A avaliação crítica dos ensaios clínicos randomizados (ver Capítulo 7) deve ser independente da hipótese gerada *a priori*, da pergunta objetiva. Devemos colocar de lado nossas preferências e avaliar cuidadosamente o estudo. Este, por sua vez, vai moldar ou influenciar a probabilidade pré-teste desta hipótese, reduzindo ou não a probabilidade de ela ser verdade (a depender do resultado do estudo). Expomos assim, a função da ciência moderna: provar a existência de um fenômeno, e não a inexistência, por meio da experimentação. Nesse caso, o ônus está em provar que o tratamento com antibióticos beneficia os pacientes, utilizando experimentos controlados (ensaios clínicos randomizados).

Assim, ao buscar as evidências científicas de qualidade, iremos perceber como os estudos de boa qualidade refutaram a hipótese do uso de antibióticos em reduzir complicações infecciosas ou morte na pancreatite aguda grave.<sup>41-46</sup> Portanto, chegamos à conclusão que não devemos utilizar esta terapia.

Apesar disto, é alta a prevalência de prescrição de antibióticos na prática clínica,<sup>47-49</sup> podendo chegar a mais de 60%. Mas e por que não fazer? Afinal estamos falando de uma terapia que parece relativamente “inócua”, e que provavelmente não irá fazer mal ao paciente. Esse é um dos maiores equívocos na medicina moderna, uma vez que além de não trazer benefício (probabilisticamente), abrimos a guarda para uma série de consequências não intencionais que podem causar danos ao paciente. De fato, a terapia antimicrobiana utilizada de forma fútil, pode levar ao desenvolvimento de bactérias multirresistentes e infecções fúngicas graves.<sup>50</sup> Parte desse fenômeno pode ser explicado em como somos seduzidos pela fisiopatologia e raciocínio lógico mecanicista, além de subestimarmos iatrogenias. Torna-se muito plausível que antibióticos reduziram o risco de complicações nesta grave tempestade inflamatória pancreática, principalmente associada à

necrose, e essa intuição pode ser confirmada de forma equivocada através de análises retrospectivas de observações do mundo real, devido aos nossos vieses cognitivos.

Agora, imagine o seguinte caso clínico hipotético: uma paciente do sexo feminino, de 60 anos de idade, deu entrada no hospital com o mesmo quadro de pancreatite aguda grave, necrotizante. Ao ser tratada clinicamente com medidas racionais, ela evolui no vigésimo primeiro dia de internação com febre, dor abdominal e leucocitose. Durante a investigação da provável complicação, foi evidenciado infecção da necrose pancreática. Apesar da utilização de antibióticos, a infecção persiste e, portanto, essa paciente precisa ser submetida a uma drenagem operatória. A pergunta se estabelece: qual deve ser a intervenção cirúrgica para tratar a necrose pancreática infectada?

Esta é uma situação crítica, pois a própria doença e o eventual procedimento tradicional utilizado por décadas (necrosectomia por laparotomia), está associado a uma alta taxa de complicações (30% a 95%) e morte (11 a 39%).<sup>51-55</sup> Entretanto, com a evolução tecnológica, técnicas minimamente invasivas (laparoscopia, cateter percutâneo, drenagem endoscópica) surgem, e estas, por sua vez, causam menor trauma cirúrgico. Apesar do entusiasmo frente a esta nova tecnologia, precisamos de um teste de hipótese para avaliar a eficácia dessas intervenções. Mais uma vez, o que encontramos como evidência científica de boa qualidade testando a eficácia desta intervenção?

Se buscarmos ensaios clínicos randomizados, iremos encontrar um estudo publicado em 2010, mostrando maior eficácia da técnica minimamente invasiva. Nele, a composição do desfecho primário analisado: incidência de complicações (como falência multiorgânica, perfuração de vísceras ocas) e morte foi maior na técnica convencional (69%) quando comparada com a técnica minimamente invasiva (40%).<sup>56</sup> O risco relativo com esta técnica foi de 0,57, portanto, mostrando proteção desta nova técnica com redução absoluta do risco de 29%. Temos um número necessário para tratar (NNT) médio de três ( $100/29=3,4$ ), ou seja, tratamos aproximadamente três pacientes para um ter o benefício. É uma conduta altamente eficaz, pois estamos falando de um NNT muito pequeno. Assim, chegamos à conclusão baseada em racionalidade que devemos utilizar a técnica minimamente invasiva para nossa paciente hipotética. Perceba que

não é uma garantia que ela irá se beneficiar, e sim de uma probabilidade estimada.

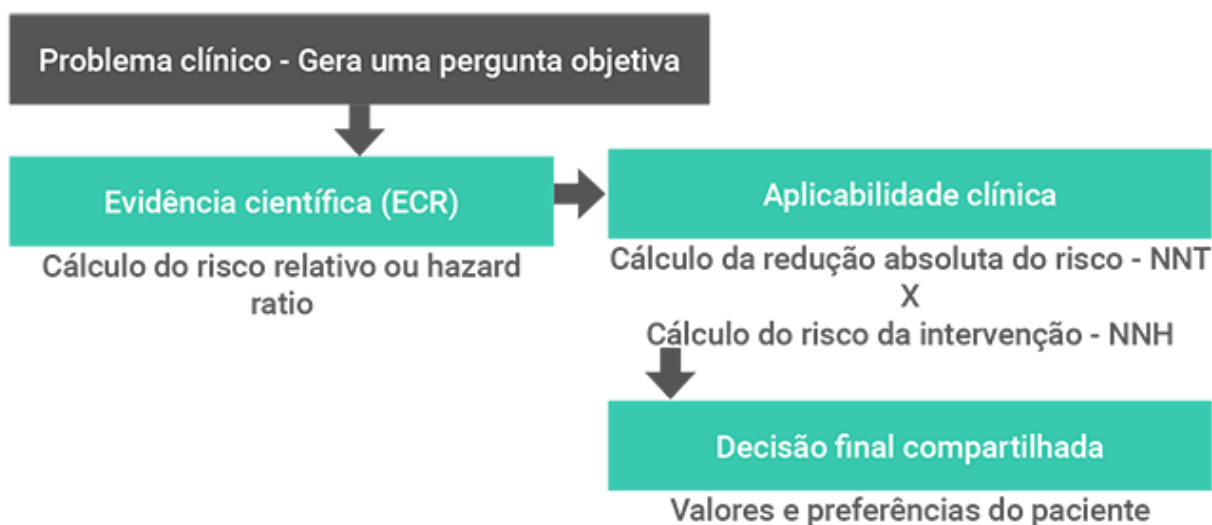
E no cenário da oncologia cirúrgica, como lidar com essa perspectiva probabilística? Seguimos o mesmo raciocínio. Imagine um homem de 70 anos, que realizou ingenuamente a pedido de um médico, rastreio para câncer de próstata (ver capítulo sobre rastreamento). Seu PSA veio de 7,8 ng/mL, e ultrassonografia evidenciando aumento da glândula prostática. Ao realizar uma biópsia guiada por imagem, foi identificado câncer de próstata localizado. O paciente fica angustiado com a incerteza de seu prognóstico, afinal recebe uma notícia de uma doença oncológica, porém estando assintomático. Ao ser encaminhado para um especialista, o mesmo com sua intuição e formação tradicional de cirurgião oncológico, com sua experiência adquirida durante anos neste tipo de intervenção, acessa sua memória e lembra de casos prévios que operou e obteve imenso sucesso na cirurgia prostatectomia radical, afinal aprendeu de forma dogmática que o melhor tratamento para um câncer localizado é a sua extirpação. De fato, esse pensamento mecanicista parece ter plausibilidade biológica, entretanto, sabemos que o câncer de próstata localizado é uma doença de comportamento indolente, e uma pergunta surge a nossa mente: a cirurgia de prostatectomia reduz mortalidade em pacientes com câncer de próstata localizado?

Ao acessarmos ensaios clínicos randomizados testando esta hipótese, identificamos um resultado interessante: além de não existir diferença significativa na sobrevida câncer específica, a sobrevida global em 12 anos ao se comparar prostatectomia radical com tratamento conservador (não operatório) foi a mesma (hazard ratio, 0,84 com intervalo de confiança 95%, 0,70 a 1,01;  $P=0,06$ ).<sup>57</sup> Mesmo em mãos experientes, toda intervenção cirúrgica pode causar danos ao paciente, e nesse estudo controlado, a prostatectomia radical causou mais incontinência urinária (40% contra 10%) e disfunção erétil. Perceba que ao calcularmos o risco de incontinência urinária, encontramos um risco absoluto maior (30%) no grupo que foi submetido à cirurgia. Assim, temos um número necessário para causar dano (“*number needed to harm*”, NNH) médio de três ( $100/30=3,3$ ), ou seja, tratamos aproximadamente três pacientes para um ter esse malefício. É uma conduta ineficaz, pois estamos observando que probabilisticamente além de não aumentar a sobrevida, a cirurgia pode causar sérios danos ao paciente. É importante conhecer o paciente por meio de uma relação médico-paciente

adequada, entendendo seus valores e preferências, que vão influenciar a nossa tomada de decisão. Assim, chegamos à conclusão baseada em racionalidade que devemos utilizar uma estratégia mais conservadora para esse paciente, fundamentada em uma decisão compartilhada.

Diante dos exemplos expostos, criamos um arcabouço sofisticado de pensamento clínico probabilístico, que pode nortear intervenções em diversas situações encontradas na prática clínica cirúrgica (Figura 6).

**Figura 6.**



Fluxograma da aplicação de intervenções em cirurgia, baseada na interpretação da evidência científica, ponderando valores e preferências do paciente.

NNT = Número necessário para tratar.

NNH = Número necessário para causar dano (“*Number needed to harm*”).

## Cirurgia baseada em evidências – as dificuldades

As dificuldades de implementação desse conceito são inúmeras. A cirurgia tradicionalmente foi mais baseada em experiência do que em evidências, e a maioria dos grandes avanços terapêuticos desta área nos últimos 200 anos, foi alicerçado sem estudos científicos rigorosos.<sup>58</sup> Décadas atrás, Richard Horton ridicularizou o quão primitiva era a pesquisa em cirurgia, comparando-a à uma ópera cômica, onde cirurgiões achavam que suas cirurgias se comportavam como verdadeiros testes estatísticos, bastando simples observações para se chegar a conclusões.<sup>59</sup> Parte dessa crítica se deve ao que discutimos no início deste capítulo: o conhecimento tradicional dogmático na cirurgia fez com que criássemos uma escola de cirurgiões

enviesados por sua experiência. De fato, a cirurgia depende de habilidade técnica e de detalhes práticos, os quais os cirurgiões estão constantemente ajustando em busca de melhores resultados. Lawrence Bonchek reiterou essas diferenças entre cirurgia e outras áreas médicas, colocando a perspectiva de que os medicamentos são compostos estáveis, cuja eficácia independe da habilidade do médico, enquanto cirurgias dependem do refinamento da arte cirúrgica, tendo uma relação linear crescente entre treinamento e eficácia.<sup>60</sup> Assim, se torna primoroso obter o maior número de cirurgiões aprimorando e realizando procedimentos, porém cria-se inconscientemente a ideia de que o mais importante é sofisticar a expertise manual, deixando de lado algo de extrema relevância: a observação cuidadosa dos fenômenos nessa área precisa passar por experimentação controlada das hipóteses geradas.

Existem outras possíveis dificuldades impostas à racionalidade alicerçada na interpretação de evidências científicas, mantendo o cirurgião no paradigma da medicina tradicional. Stirrat<sup>3</sup> descreveu algumas:

- A cirurgia envolve ação, portanto o papel do cirurgião sempre foi agir, atraindo médicos com este perfil cognitivo. Isso talvez torne o cirurgião menos reflexivo.
- Gerações sucessivas de cirurgiões aprenderam técnicas e habilidades passadas por um chefe, e o viés da autoridade impedia qualquer questionamento.
- A maioria dos cirurgiões se sente convencida de que seus atos são sempre em prol do melhor para seus pacientes, subestimando eventuais danos.
- A cirurgia, por definição, invade a integridade corporal do paciente no intuito de tratá-lo na seguinte perspectiva: é tudo ou nada (afinal, não se pode realizar uma cirurgia pela metade).
- Ter de confessar incertezas sobre uma cirurgia pode minar a confiança do paciente.



- O cirurgião se torna vaidoso, por meio de elogios pessoais graças à sua expertise e inovação cirúrgica. Testes científicos podem ameaçar esse status.

De fato, são teorias plausíveis, que perpetuam um pensamento baseado na tradição cirúrgica. Acrescento mais um item a esta questão: a falta de entendimento probabilístico na cirurgia, mantendo o médico com a convicção de que suas práticas garantem sempre desfechos favoráveis aos seus pacientes, quando na realidade ele apenas oferece um “bisturi de probabilidades”.

Existem ainda uma série de vieses cognitivos que de fato foram demonstrados por estudos de psicologia comportamental.<sup>11, 61-66</sup> Esses, por sua vez, podem estar presentes tanto em médicos generalistas, cirurgiões e até mesmo nos pacientes, podendo confundir o julgamento clínico. Eric Andrew Finkelstein, estudou parte dessas heurísticas em pacientes oncológicos e encontrou resultados interessantes (Tabela 2).<sup>65</sup>

**Tabela 2.**

Vieses Cognitivos	Efeito Provável
Escolha ao risco	Pacientes tendem a ter maior expectativa em riscos apresentados como perdas, do que como ganhos.
Risco de atribuição	Pacientes tendem a escolher mais tratamentos quando apresentados de forma positiva (exemplo: a cirurgia tem 80% de sobrevivência), do que quando apresentados de forma negativa (exemplo: a cirurgia tem mortalidade de 20%).
Viés do risco relativo e risco absoluto	Espera-se que os pacientes tenham mais (ou menos) probabilidade de buscar tratamento quando as reduções (ou aumentos) de risco são apresentadas em termos relativos do que em termos absolutos
Viés do padrão	Pacientes tendem a buscar tratamentos que são apresentados como a opção “padrão”.
Viés otimista	Espera-se que os pacientes percebam seu risco individual de uma forma mais positiva do que outros pacientes na mesma circunstância
Viés de projeção	Espera-se que os pacientes presumam que seus gostos ou preferências permanecerão os mesmos com o tempo (exemplo: mesmo após uma cirurgia mutilante). Fatores viscerais, como dor e depressão, podem afetar as escolhas de tratamento dos pacientes.

Vieses Cognitivos	Efeito Provável
Viés do presente	Os pacientes tendem a colocar um peso desproporcional no momento presente (decisão impulsiva), do que avaliar ganhos na perspectiva futura. Efeitos adversos podem levar à não adesão da terapia.

Vieses cognitivos e seus potenciais efeitos.

Perceba como o comportamento humano presente até mesmo nos pacientes pode explicar também como a intuição faz com que tenhamos percepções equivocadas sobre riscos, benefícios e performance, dificultando o aprimoramento da cirurgia baseada em evidências.

Uma revisão sistemática publicada em 2017 por Tammy Hoffman e colaboradores, avaliando aproximadamente 13.000 médicos, concluiu que eles raramente tinham expectativas precisas de benefícios ou prejuízos de suas intervenções (exames, rastreios e tratamentos). Existiam imprecisões em ambas as direções. No entanto, os médicos subestimaram com maior frequência os danos e superestimaram com maior frequência os benefícios de suas ações.<sup>67</sup> Com isso, percepções imprecisas sobre os benefícios e danos na área da cirurgia, podem resultar em opções de manejo clínico inadequados.

Recentemente, o viés otimista foi demonstrado na área da cirurgia ortopédica. Um estudo avaliou a percepção dos ortopedistas e dos pacientes no controle da dor, ao realizar um procedimento operatório (artroplastia de ombro). Mais anos de prática, posição hierárquica hospitalar e maior número de cirurgias realizadas por ano causaram uma percepção ilusória nos cirurgiões. Eles percebiam que seus pacientes obtinham um maior alívio da dor após a cirurgia, do que a melhora real da dor sentida por estes pacientes.<sup>68</sup> Assim, os cirurgiões que operavam mais pensavam que tinham um melhor desempenho.

Concluimos, portanto, que não basta apenas abrir os olhos e buscar um maior letramento científico, deixando o dogmatismo de lado. É importante também que cirurgiões e médicos de uma forma geral, reconheçam como a mente pode nos trair, entendendo como vieses cognitivos se comportam, buscando uma maior racionalidade na tomada de decisões.

**Cirurgia baseada em evidências – a reforma iluminista**

O Iluminismo, um dos movimentos intelectuais e filosóficos mais importantes do século XVIII, incluiu um conjunto de ideias centradas na razão. Assim, a função primordial desse capítulo é apresentar uma visão diferente da cirurgia alicerçada em fatos e inspirada nos seguintes ideais iluministas: razão, ciência e humanismo aplicados à prática clínica.

A cirurgia talvez seja a especialidade que mais exige do paciente. Cabe ao cirurgião, dentro de uma perspectiva humanista, lembrar que os pacientes depositam nele extraordinária confiança, na promessa de que ao cortar seus corpos irão, ao final, proporcionar a cura ou alívio de seu sofrimento. Portanto, antes de olharmos o que as evidências informam sobre as intervenções em cirurgia, é fundamental lembrarmos de dois aforismas que remetem integridade moral e ética:

1. Hipocrático: “*primum non nocere*”. Ou seja, a fundamentação primordial da cirurgia é não causar dano.
2. Talebiano (Nassim Nicholas Taleb):
  - Via negativa: a razão pela qual iatrogenias ocorrem, em parte, se deve ao viés da intervenção, a via positiva. Ou seja, cirurgiões pensando em soluções rápidas e mágicas, na propensão de querer ajudar, agem causando danos ao paciente. Nós não somos lembrados pelo que evitamos ou pela “não ação”. Por isso a importância de agir pela via negativa (quando não temos evidência), deixando claro que o conhecimento negativo é mais robusto ao erro e à iatrogenia.
  - “*Skin in the game*”. Ou seja, ter a pele no jogo. Conceito criado no mercado financeiro, sobre ter a pele no jogo na tomada de decisões, ao correr riscos em investimentos.<sup>69</sup> Poderíamos fazer uma analogia na área cirúrgica, onde o cirurgião faz uma reflexão moral sobre qual a melhor decisão a ser tomada se a pele dele estivesse no jogo.

Bem, essa tentativa de reducionismo moral e ético auxilia extraordinariamente em criarmos um freio contra ações fúteis, que não são respaldadas por evidências científicas de qualidade e que podem trazer sérios danos ao paciente.

É papel fundamental das pesquisas em cirurgia e também da decisão clínica do cirurgião escrutinar a magnitude do benefício de suas intervenções sem deixar de lado os danos eventuais causados por essas mesmas intervenções. Diferentemente de terapias medicamentosas, um dano associado a um procedimento operatório pode causar grande impacto na qualidade de vida ou eventual morte de um paciente. Assim, precisamos entender que evidências científicas observacionais são tão importantes quanto ensaios clínicos randomizados, a depender de qual pergunta objetiva realizamos. Estudos observacionais são suficientes para avaliar fatores de risco para doenças, escores de risco em cirurgias (por exemplo: escore de risco para fístulas pancreáticas), modelos epidemiológicos preditores ou em situações de plausibilidade extrema (um princípio que constitui em observar a eficácia do óbvio). Exemplos de situações de plausibilidade extrema são: o uso de antibióticos em infecções abdominais, suturas de ferimentos extensos, drenagem de abscessos, esplenectomia em paciente com trauma esplênico grave e choque hipovolêmico, fixações de fraturas expostas, drenagem de hematoma extradural e outras. Nestes cenários não precisamos de experimentos controlados para constatar o óbvio, é uma questão de eficácia auto evidente.

Por sua vez, à medida que racionalizamos criticamente as evidências que foram construídas na história da cirurgia, percebemos que a maioria das situações que precisamos inferir eficácia necessita de verificação através de ensaios clínicos randomizados. É espantoso como muitos colegas cirurgiões resistem a essa ideia, e argumentos foram feitos para indicar que as circunstâncias do ecossistema da cirurgia, impedem que ensaios clínicos randomizados respondam essas questões.<sup>70-73</sup> De fato, em 2003 uma revisão de 134.689 ensaios clínicos randomizados publicados em 34 anos, demonstrou que apenas 15% deles eram de cirurgia.<sup>36</sup> É evidente que em especialidades cirúrgicas, existem peculiaridades que dificultam (mas não inviabilizam) a realização de estudos controlados. São eles:

- Cirurgia vs. “*sham*”: Equivalente a uma cirurgia placebo. Esse tipo de experimento é bastante utilizado em modelos animais.
- Vieses: seleção e observador (alguns cirurgiões argumentam que determinada cirurgia é feita para um determinado tipo de paciente).

- Cegamento: de pacientes e cirurgiões.
- Curva de aprendizado: em casos de técnicas operatórias, quanto maior a complexidade, maior a curva de aprendizado.
- Eficácia e efetividade: um procedimento eficaz, quando for aplicado por um cirurgião inábil no mundo real, pode ser ineficiente.
- Padronização de técnicas.
- Falta de educação em epidemiologia clínica.

Esses obstáculos não devem impedir a realização de estudos científicos controlados, porém exigirão que todos os cirurgiões reflitam e busquem soluções para construir um ecossistema científico aprimorado, íntegro e que responda objetivamente problemas clínicos.

O conhecimento científico, associado à ética em pesquisa na atualidade, já venceu alguns destes obstáculos. Alguns exemplos provocantes de intervenções cirúrgicas, que foram testadas por meio de experimentos controlados com grupo “*sham*”, falharam em mostrar benefício em ensaios clínicos randomizados. Por exemplo: artroplastia de joelho para osteoartrite degenerativa ou laceração meniscal,<sup>74-75</sup> descompressão subacromial para síndrome do impacto do ombro,<sup>76</sup> desnervação renal para tratamento de hipertensão arterial,<sup>77</sup> cirurgia vertebral para fraturas vertebrais osteoporóticas,<sup>78</sup> angioplastia coronária percutânea para melhorar capacidade funcional à exercícios<sup>79</sup> e injeção peridural com glicocorticoides para tratamento de dor ocasionada por estenose da coluna lombar.<sup>80</sup> De fato, um número considerável de estudos *sham-controlados* em cirurgia falharam em demonstrar benefício de procedimentos invasivos.<sup>81</sup> Isso significa que cirurgias funcionam menos do que imaginamos? Ou apenas estes exemplos mostram cenários de fácil execução de experimentos placebo-controlados na cirurgia, mas que já se mostravam como hipóteses de baixa eficácia *a priori*? Não sabemos, mas de qualquer forma, isto nos remonta a importância de avaliarmos com cuidado nossas intervenções, e caso haja evidências escassas ou conflitantes, devemos testar nossas ideias.

O cegamento é outro problema na área da cirurgia, que pode afetar a avaliação crítica de eficácia, entretanto, um estudo recente conseguiu de uma forma engenhosa, cegar os pacientes submetidos a duas intervenções cirúrgicas distintas. O estudo comparou a técnica minimamente invasiva (laparoscópica) contra a técnica convencional (cirurgia aberta), em pacientes submetidos à pancreatemia distal devido a tumores situados na cauda do pâncreas. O cegamento dos pacientes foi realizado utilizando um grande curativo e vestimenta que cobria toda a cavidade abdominal, impedindo a visualização da incisão. Esse experimento controlado, por sua vez, concluiu que a cirurgia laparoscópica foi mais eficaz, trazendo uma recuperação funcional mais rápida dos pacientes.<sup>82</sup>

Devemos permanecer curiosos e céticos a respeito de técnicas ou novas tecnologias que vão surgindo na área da cirurgia. Toda modernidade e inovação é muito bem-vinda, entretanto, devemos submetê-las ao escrutínio da experimentação rigorosa. No início dos anos 1990, a laparoscopia surgiu com muito entusiasmo, e foi amplamente utilizada na prática clínica durante décadas sem passar por estudos controlados. Bastavam apenas evidências observacionais para justificar a utilização desta técnica minimamente invasiva em todas as especialidades cirúrgicas. Por sorte, ela se mostrou de fato eficaz na maioria dos experimentos realizados na cirurgia. Algumas exceções foram verificadas, especialmente na cirurgia pancreática de alta complexidade (duodenopancreatemia). O mesmo grupo que mostrou a eficácia da técnica laparoscópica para tumores pancreáticos distais, testou em um ensaio clínico randomizado se a duodenopancreatemia minimamente invasiva seria superior a técnica convencional (cirurgia aberta) em pacientes com tumores da cabeça do pâncreas. O estudo foi interrompido pelo comitê de ética, devido à maior mortalidade no grupo cirurgia minimamente invasiva, mesmo em mãos experientes.<sup>83</sup> Isso caracteriza o que chamamos de “*medical reversal*” ou reversão médica, que é quando um ensaio clínico mais recente e metodologicamente superior produz resultados que contradizem a prática clínica existente, passando a ser inadequada. Realmente, essa cirurgia era feita amplamente antes deste estudo, uma vez que muitos cirurgiões acreditavam que realmente estavam beneficiando seus pacientes.<sup>84-87</sup>

Mais recentemente, surgiu o robô “Da Vinci”, uma máquina de última geração, extraordinariamente sofisticada, que utiliza braços articulados e imagem tridimensional de alta resolução para realizar cirurgias

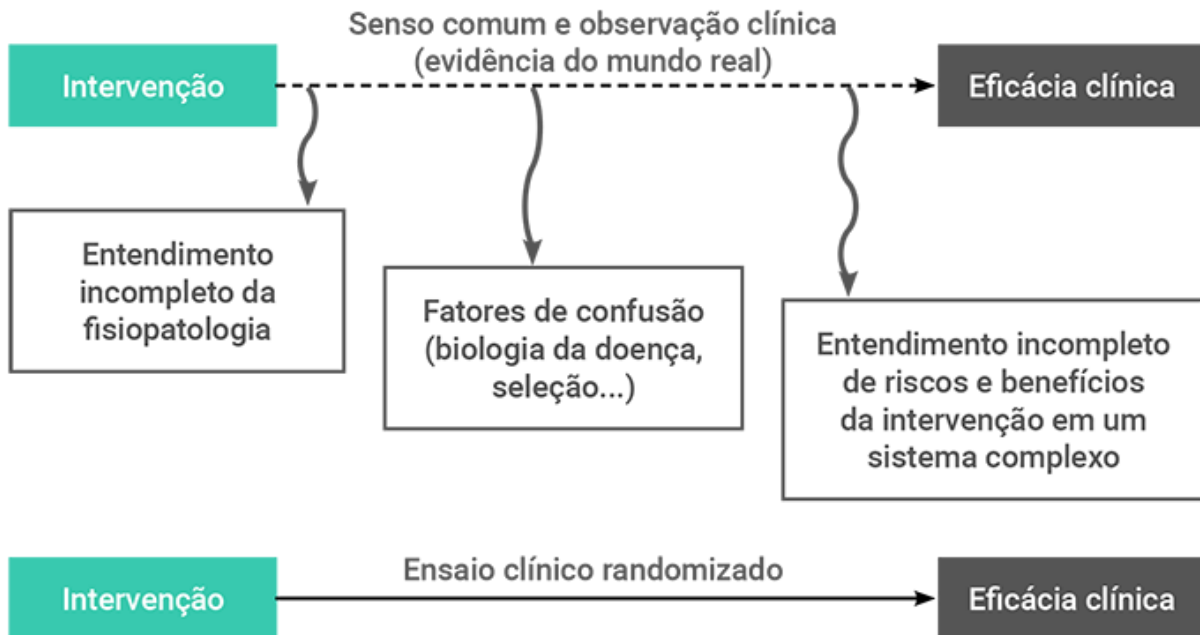
minimamente invasivas. Sem dúvida é um grande avanço na cirurgia, mas como discutimos neste capítulo, é fundamental testarmos nossas engenhosidades avaliando seu real benefício. Ela provavelmente substituirá a técnica laparoscópica tradicional, por trazer mais conforto ao cirurgião, mantendo os mesmos benefícios da laparoscopia. De fato, ensaios clínicos randomizados já mostraram a não inferioridade da técnica robótica em relação a técnicas convencionais, porém o maior problema ainda é o alto custo.<sup>88-91</sup>

Apesar do grande ímpeto dessa nova modalidade, alguns grupos de urologia no mundo fizeram falsas inferências, por meio de estudos observacionais. Urologistas relataram uma menor incidência de complicações urológicas após cirurgias radicais da próstata utilizando a plataforma robótica, principalmente incontinência urinária e disfunção erétil.<sup>92-98</sup> Entretanto, um ensaio clínico randomizado recente, comparando essa técnica com a cirurgia aberta, mostrou resultados equivalentes destas complicações, contrariando observações prévias.<sup>99</sup> É por isso que a verificação de uma hipótese é fundamental, para entendermos como o mundo da cirurgia funciona.

Espero que o exposto ajude a compreender a importância de uma reforma iluminista na arte de pensar de cirurgiões, sabendo como utilizar a observação empírica para gerar hipóteses, tendo a curiosidade de buscar a experimentação controlada dessas conjecturas. Um estudo recente na área de cardiologia, ilustrou muito bem como apenas o senso comum em conjunto com a observação do mundo real usualmente apresentam fatores de confusão, gerando falsas associações causais (ver Figura 7).<sup>100</sup> Esta perspectiva pode ser muito bem implementada em todas as especialidades, incluindo a cirurgia. Por exemplo: um cirurgião que utiliza a via positiva pode ingenuamente achar que um determinado paciente, com uma doença oncológica avançada peritoneal, teve uma maior sobrevida quando comparado a um paciente que ele não operou. O cirurgião pode fazer uma falsa associação causal que foi a sua cirurgia que promoveu mais tempo de vida, quando na realidade inúmeros fatores podem ter contribuído para isso (idade ou um tumor biologicamente mais indolente por exemplo).

### **Figura 7.**

## A randomização é fundamental para avaliar eficácia de intervenções:



A evidência do mundo real (observação) não descreve com precisão o efeito de um tratamento nos resultados por várias razões. A randomização contorna esses mecanismos de falha, criando um experimento controlado para entender o verdadeiro efeito de uma intervenção. Quando o bom senso e a observação clínica são a base para as decisões de tratamento, há grande incerteza sobre o benefício e risco, de forma que os tratamentos podem ser ineficazes e/ou colocar o paciente em risco. Em contraste, a mágica da randomização presente em experimentos controlados de alta qualidade, avalia adequadamente benefício e risco de uma intervenção, permitindo uma tomada de decisão eficaz.<sup>100</sup>

## Conclusões

Espero que as reflexões deste capítulo ajudem o leitor a compreender que a cirurgia vem evoluindo, se desgarrando de uma medicina tradicional e dogmática. Em breve, cada vez mais veremos cirurgiões abraçando a incerteza da medicina baseada em evidências, transformando uma cultura cirúrgica em uma forma mais intelectualizada e racional. Talvez a especialidade tenha demorado a aceitar novos padrões do pensamento, devido a conceitos criados pela prática inquestionável, passada hierarquicamente por gerações, além das dificuldades trazidas à tona por nossos vieses cognitivos. Percebemos a importância da observação, principalmente em situações de plausibilidade extrema, em que a evidência do óbvio é suficiente para determinar eficácia de algumas intervenções. Em situações que fogem a isso, é fundamental avaliarmos a possibilidade da verificação através de experimentos controlados. Assim, surge algo



interessante: um cenário no qual uma intervenção cirúrgica precisa ser testada para avaliar sua eficácia, e caso se prove benéfica, um cirurgião habilidoso, bem treinado, pode fazer com que sua expertise transforme um procedimento eficaz, em um altamente efetivo na prática clínica.

Inúmeras perguntas e dúvidas surgirão, não apenas na cirurgia, mas na história da medicina e da ciência. Cabe a nós, dentro de uma perspectiva humilde, buscar a razão e os valores do humanismo para encontrar o melhor caminho.

## Referências bibliográficas

1. Tubbs RS. The Greatest obstacle to discovery. *Clin Anat.* 2018;31(947).
2. Parker S. Kill or Cure. *An Illustrated History of Medicine.* 2013.
3. Stirrat G. Ethics and evidence based surgery. *J Med Ethics.* 2004;30:160–5.
4. Correia LC. Princípios da medicina baseada em evidências [Internet]. 2011 [[acesso](#) em: 6 abr. 2021].
5. Rooney A. A história da medicina. Das primeiras curas aos milagres da medicina moderna. 2013.
6. Greenstone G. The history of bloodletting. *B C Med J.* 2010;52(1):12–4.
7. Vieira R, Pacheco RL, Latorraca CO, Martimbianco AL. Catálogo de vieses [Internet] [[acesso](#) em: 6 abr. 2021].
8. Tversky A, Kahneman D. The framing of decisions and the psychology of choice. *Exp Environ Econ.* 2018;1:173–8.
9. Tversky A, Kahneman D. Judgment under uncertainty: Heuristics and biases. *Judgm Under Uncertain Heuristics Biases.* 2017;185:1–92.
10. Read JD. The availability heuristic in person identification: The sometimes misleading consequences of enhanced contextual information. *Appl Cogn Psychol.* 1995;9(2):91–121.
11. Kahneman D. Rápido e devagar. Duas formas de pensar. 2011.
12. Tubre DJ, Schroeder AD, Estes J, Eisenga J, Fitzgibbons RJ. Surgical site infection: the “Achilles Heel” of all types of abdominal wall hernia reconstruction. *Hernia* [Internet]. 2018;22(6):1003–13 [[acesso](#) em: 6 abr. 2021].
13. Lockhart K, Dunn D, Teo S, Ng JY, Dhillon M, Teo E, et al. Mesh versus non-mesh for inguinal and femoral hernia repair. *Cochrane Database Syst Rev.* 2018;2018(9).
14. Ran K, Wang X, Zhao Y. Open tensionless repair techniques for inguinal hernia: a meta-analysis of randomized controlled trials. *Hernia* [Internet]. 2020;24(4):733–45 [[acesso](#) em: 6 abr. 2021].
15. Wu JJ, Way JA, Eslick GD, Cox MR. Transabdominal Pre-Peritoneal Versus Open Repair for Primary Unilateral Inguinal Hernia: A Meta-analysis. *World J Surg.* 2018;42(5):1304–11.

16. Bhangu A, Singh P, Pinkney T, Blazeby JM. A detailed analysis of outcome reporting from randomised controlled trials and meta-analyses of inguinal hernia repair. *Hernia*. 2015;19(1):65–75.
17. Bellhouse DR. The reverend thomas bayes, FRS: A biography to celebrate the tercentenary of his birth. *Stat Sci*. 2004;19(1):3–43.
18. Gillies DA. Was Bayes a Bayesian? *Hist Math*. 1987;14(4):325–46.
19. Miller D. *Textos escolhidos Popper*. 2010.
20. Popper K. *A lógica da pesquisa científica*. 2013.
21. Sacket DL. Evidence based medicina: what is and what it isn't. *B C Med J*. 1996;321:72.
22. Borghi L. William Osler , Medical History and the Origins of the International Society for the History of Medicine William Osler , medical history and the origins of the International Society for the History of Medicine. *Vesalius*. 2019;(December 2018).
23. Osler W. *The Evolution of Modern Medicine*, Yale University Press, New Haven 1921. 1921.
24. Bhangu A, Søreide K, Di Saverio S, Assarsson JH, Drake FT. Acute appendicitis: Modern understanding of pathogenesis, diagnosis, and management. *Lancet*. 2015;386(10000):1278–87.
25. Cartwright SL, Knudson MP. Evaluation of acute abdominal pain in adults. *Am Fam Physician*. 2008;77(7):971–8.
26. Frossard JL, Steer ML, Pastor CM. Acute pancreatitis. *Lancet*. 2008;371(9607):143–52.
27. Gans SL, Pols MA, Stoker J, Boermeester MA. Guideline for the diagnostic pathway in patients with acute abdominal pain. *Dig Surg*. 2015;32(1):23–31.
28. McNamara R, Dean AJ. Approach to Acute Abdominal Pain. *Emerg Med Clin North Am*. 2011;29(2):159–73.
29. Natesan S, Lee J, Volkamer H, Thoureen T. Evidence-Based Medicine Approach to Abdominal Pain. *Emerg Med Clin North Am* [Internet]. 2016;34(2):165–90. Available from: <http://dx.doi.org/10.1016/j.emc.2015.12.008>
30. Bhuiya FA, Pitts SR, McCaig LF. Emergency department visits for chest pain and abdominal pain: United States, 1999-2008. *NCHS Data Brief*. 2010;(43):1–8.
31. Alvarado A. A practical score for the early diagnosis of acute appendicitis. *Ann Emerg Med*. 1986;15(5):557–64.
32. Wagner JM, McKinney WP, Carpenter JL. Does this patient have appendicitis? *J Am Med Assoc*. 1996;276(19):1589–94.
33. Hwang ME. Sonography and Computed Tomography in Diagnosing Acute Appendicitis. *Radiol Technol*. 2018;89(3):224–37.
34. Gorter RR, Eker HH, Gorter-Stam MAW, Abis GSA, Acharya A, Ankersmit M, et al. Diagnosis and management of acute appendicitis. EAES consensus development conference 2015. *Surg Endosc*. 2016;30(11):4668–90.

35. Hoffmann TC, Del Mar C. Clinicians' expectations of the benefits and harms of treatments, screening, and tests: A systematic review. *JAMA Intern Med.* 2017;177(3):407–19.
36. Wente MN, Seiler CM, Uhl W, Büchler MW. Perspectives of evidence-based surgery. *Dig Surg.* 2003;20(4):263–9.
37. Toledo-Pereyra LH. Evidence-based surgery. *J Investig Surg.* 2005;18(5):219–22.
38. Meakins JL. Evidence-based surgery. *Surg Clin North Am.* 2006;86(1):1–16.
39. Duran-Vega HC. Evidence based surgery: A necessary tool. *Cirugía y Cir (English Ed [Internet].* 2015;83(3):265–70 [[acesso](#) em: 6 abr. 2021].
40. Shearman AD, Shearman CP. Evidence-based surgery. *Surg (United Kingdom) [Internet].* 2018;36(9):448–53 [[acesso](#) em: 6 abr. 2021].
41. Boxhoorn L, Voermans RP, Bouwense SA, Bruno MJ, Verdonk RC, Boermeester MA, et al. Acute pancreatitis. *Lancet [Internet].* 2020;396(10252):726–34 [[acesso](#) em: 6 abr. 2021].
42. García-Barrasa A, Borobia FG, Pallares R, Jorba R, Poves I, Busquets J, et al. A double-blind, placebo-controlled trial of ciprofloxacin prophylaxis in patients with acute necrotizing pancreatitis. *J Gastrointest Surg.* 2009;13(4):768–74.
43. Lim CLL, Lee W, Liew YX, Tang SSL, Chlebicki MP, Kwa ALH. Role of Antibiotic Prophylaxis in Necrotizing Pancreatitis: A Meta-Analysis. *J Gastrointest Surg.* 2015;19(3):480–91.
44. Vege SS, DiMagno MJ, Forsmark CE, Martel M, Barkun AN. Initial Medical Treatment of Acute Pancreatitis: American Gastroenterological Association Institute Technical Review. *Gastroenterology [Internet].* 2018;154(4):1103–39 [[acesso](#) em: 6 abr. 2021].
45. Xue P, Deng LH, Zhang Z Da, Yang XN, Wan MH, Song B, et al. Effect of antibiotic prophylaxis on acute necrotizing pancreatitis: Results of a randomized controlled trial. *J Gastroenterol Hepatol.* 2009;24(5):736–42.
46. Villatoro E, Mulla M, Larvin M. Antibiotic therapy for prophylaxis against infection of pancreatic necrosis in acute pancreatitis. *Cochrane Database Syst Rev.* 2010;
47. Baltatzis M, Mason JM, Chandrabalan V, Stathakis P, McIntyre B, Jegatheeswaran S, et al. Antibiotic use in acute pancreatitis: An audit of current practice in a tertiary centre. *Pancreatology [Internet].* 2016;16(6):946–51 [[acesso](#) em: 6 abr. 2021].
48. Talukdar R, Ingale P, Choudhury HP, Dhingra R, Shetty S, Joshi H, et al. Antibiotic use in acute pancreatitis: An Indian multicenter observational study. *Indian J Gastroenterol.* 2014;33(5):458–65.
49. Tan JW, Gao Y, Kow AWC, Bonney G, Madhavan K, Windsor JA, et al. Clinical management and outcomes of acute pancreatitis: Identifying areas for quality improvement in a tertiary Asian setting. *Pancreatology [Internet].* 2019;19(4):507–18 [[acesso](#) em: 6 abr. 2021].
50. Lee HS, Lee SK, Park DH, Lee SS, Seo DW, Kim MH, et al. Emergence of multidrug resistant infection in patients with severe acute pancreatitis. *Pancreatology [Internet].*

2014;14(6):450–3 [[acesso](#) em: 6 abr. 2021].

51. Büchler MW, Gloor B, Müller CA, Friess H, Seiler CA, Uhl W. Acute necrotizing pancreatitis: Treatment strategy according to the status of infection. *Ann Surg*. 2000;232(5):619–22.
52. Connor S, Alexakis N, Raraty MGT. Early and late complications after pancreatic necrosectomy. *Surgery*. 2003;499–505.
53. Howard TJ, Patel JB, Zyromski N, Pitt HA, Lillemoe KD. Declining Morbidity and Mortality Rates in the Surgical Management of Pancreatic Necrosis. *J Gastrointest Surg*. 2007;43–9.
54. Rau B, Bothe A, Beger HG. Surgical treatment of necrotizing pancreatitis by necrosectomy and closed lavage : Changing patient characteristics and outcome in a. *Surgery*. 2001;28–39.
55. Rodriguez JR, Razo AO, Targarona J, Thayer SP, Rattner DW, Warshaw AL, et al. Debridement and Closed Packing for Sterile or Infected Necrotizing Pancreatitis Insights into Indications and Outcomes in 167 Patients. *Ann Surg*. 2008;294–9.
56. Bhatti UF, Alam HB. A Step-up Approach or Open Necrosectomy for Necrotizing Pancreatitis. 50 Landmark Pap [Internet]. 2019 Nov 26;108–10 [[acesso](#) em: 6 abr. 2021].
57. Wilt TJ, Jones KM, Barry MJ, Andriole GL, Culkin D, Wheeler T, et al. Follow-up of Prostatectomy versus Observation for Early Prostate Cancer. *N Engl J Med* [Internet]. 2017 Jul 13;377(2):132–42 [[acesso](#) em: 6 abr. 2021].
58. Gawande A. Two Hundred Years of Surgery. *N Engl J Med* [Internet]. 2012 May 3;366(18):1716–23 [[acesso](#) em: 6 abr. 2021].
59. Horton R. Surgical research or comic opera: questions, but few answers. *Lancet*. 1996;347(1):984–5.
60. Bonchek LI. Are Randomized Trials Appropriate for Evaluating New Operations? *N Engl J Med* [Internet]. 1979 Jul 5;301(1):44–5 [[acesso](#) em: 6 abr. 2021].
61. Weinstein ND. Health Risk Appraisal and Optimistic Bias. *Int Encycl Soc Behav Sci*. 2001;6612–5.
62. Jansen KA. The Optimistic Bias and Illusions of Control in Clinical Research. *IRB Ethics Hum Res* [Internet]. 2016 Aug 1;38(2):8–14 [[acesso](#) em: 6 abr. 2021].
63. Tversky A, Kahneman D. The framing of decisions and the psychology of choice. *Science* (80- ) [Internet]. 1981 Jan 30;211(4481):453–8 [[acesso](#) em: 6 abr. 2021].
64. Kane HL, Halpern MT, Squiers LB, Treiman KA, McCormack LA. Implementing and evaluating shared decision making in oncology practice. *CA Cancer J Clin*. 2014;64(6):377–88.
65. Ozdemir S, Finkelstein EA. Cognitive Bias: The Downside of Shared Decision Making. *JCO Clin Cancer Informatics*. 2018;(2):1–10.
66. Vogel P, Vogel DHV. Cognition errors in the treatment course of patients with anastomotic failure after colorectal resection. *Patient Saf Surg*. 2019;13(1):1–10.

67. Hoffmann TC, Del Mar C. Clinicians' Expectations of the Benefits and Harms of Treatments, Screening, and Tests. *JAMA Intern Med* [Internet]. 2017 Mar 1;177(3):407 [[acesso](#) em: 6 abr. 2021].
68. Torrens C, Miquel J, Santana F. Optimistic bias: the more you do, the better you think it goes. *Survey analysis of reverse shoulder arthroplasty. Patient Relat Outcome Meas.* 2019; 10:277–82.
69. Taleb NN. *Arriscando a própria pele: Assimetrias ocultas no cotidiano.* Ed. Objetiva, editor. 2018.
70. Lilford R, Braunholtz D, Harris J, Gill T. Trials in surgery. *Br J Surg* [Internet]. 2004 Jan;91(1):6–16 [[acesso](#) em: 6 abr. 2021].
71. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: Problems and possible solutions. *Br Med J.* 2002;324(7351):1448–51.
72. Meakins JL. Innovation in surgery: the rules of evidence. *Am J Surg* [Internet]. 2002 Apr;183(4):399–405 [[acesso](#) em: 6 abr. 2021].
73. Sutton DN, Wayman J, Griffin SM. Learning curve for oesophageal cancer surgery. *Br J Surg* [Internet]. 1998 Oct;85(10):1399–402 [[acesso](#) em: 6 abr. 2021].
74. Johnson LL. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *Arthroscopy.* 2002;18(7):683–7.
75. Sihvonen R, Paavola M, Malmivaara A, Itälä A, Joukainen A, Nurmi H, et al. Arthroscopic Partial Meniscectomy versus Sham Surgery for a Degenerative Meniscal Tear. *N Engl J Med.* 2013;369(26):2515–24.
76. Paavola M, Malmivaara A, Taimela S, Kanto K, Inkinen J, Kalske J, et al. Subacromial decompression versus diagnostic arthroscopy for shoulder impingement: randomised, placebo surgery controlled clinical trial. *BMJ.* 2018;362:k2860.
77. Bhatt DL, Kandzari DE, O'Neill WW, D'Agostino R, Flack JM, Katzen BT, et al. A Controlled Trial of Renal Denervation for Resistant Hypertension. *N Engl J Med.* 2014;370(15):1393–401.
78. Buchbinder R, Osborne RH, Ebeling PR, Wark JD, Mitchell P, Wriedt C, et al. A Randomized Trial of Vertebroplasty for Painful Osteoporotic Vertebral Fractures. *N Engl J Med* [Internet]. 2009 Aug 6;361(6):557–68 [[acesso](#) em: 6 abr. 2021].
79. Al-Lamee R, Thompson D, Dehbi HM, Sen S, Tang K, Davies J, et al. Percutaneous coronary intervention in stable angina (ORBITA): a double-blind, randomised controlled trial. *Lancet.* 2018;391(10115):31–40.
80. Friedly JL, Comstock BA, Turner JA, Heagerty PJ, Deyo RA, Sullivan SD, et al. A Randomized Trial of Epidural Glucocorticoid Injections for Spinal Stenosis. *N Engl J Med.* 2014;371(1):11–21.
81. Probst P, Grummich K, Harnoss JC, Hüttner FJ, Jensen K, Braun S, et al. Placebo-Controlled Trials in Surgery. *Med (United States).* 2016;95(17):1–8.

82. De Rooij T, Van Hilst J, Van Santvoort H, Boerma D, Van Den Boezem P, Daams F, et al. Minimally Invasive Versus Open Distal Pancreatectomy (LEOPARD): A Multicenter Patient-blinded Randomized Controlled Trial. *Ann Surg.* 2019;269(1):2–9.
83. van Hilst J, De Rooij T, Bosscha K, Brinkman DJ, Van Dieren S, Dijkgraaf MG, et al. Laparoscopic versus open pancreatoduodenectomy for pancreatic or periampullary tumours (LEOPARD-2): a multicentre, patient-blinded, randomised controlled phase 2/3 trial. *Lancet Gastroenterol Hepatol.* 2019;4(3):199–207.
84. Umemura A, Nitta H, Takahara T, Hasegawa Y, Sasaki A. Current status of laparoscopic pancreaticoduodenectomy and pancreatectomy. *Asian J Surg [Internet].* 2018;41(2):106–14 [[accesso](#) em: 6 abr. 2021].
85. Mohammed S, Van Buren G, Fisher WE. Pancreatic cancer: Advances in treatment. *World J Gastroenterol.* 2014;20(28):9354–60.
86. Jacobs MJ, Kamyab A. Total laparoscopic pancreaticoduodenectomy. *J Soc Laparoendosc Surg.* 2013;17(2):188–93.
87. Capretti G, Boggi U, Salvia R, Belli G, Coppola R, Falconi M, et al. Application of minimally invasive pancreatic surgery: an Italian survey. *Updates Surg [Internet].* 2019;71(1):97–103 [[accesso](#) em: 6 abr. 2021].
88. Prabhu AS, Carbonell A, Hope W, Warren J, Higgins R, Jacob B, et al. Robotic inguinal vs transabdominal laparoscopic inguinal hernia repair the RIVAL randomized clinical trial. *JAMA Surg.* 2020;155(5):380–7.
89. Kim MJ, Park SC, Park JW, Chang HJ, Kim DY, Nam BH, et al. Robot-assisted Versus Laparoscopic Surgery for Rectal Cancer: A Phase II Open Label Prospective Randomized Controlled Trial. *Ann Surg.* 2018;267(2):243–51.
90. Parekh DJ, Reis IM, Castle EP, Gonzalgo ML, Woods ME, Svatek RS, et al. Robot-assisted radical cystectomy versus open radical cystectomy in patients with bladder cancer (RAZOR): an open-label, randomised, phase 3, non-inferiority trial. *Lancet [Internet].* 2018;391(10139):2525–36 [[accesso](#) em: 6 abr. 2021].
91. van der Sluis PC, van der Horst S, May AM, Schippers C, Brosens LAA, Joore HCA, et al. Robot-assisted Minimally Invasive Thoracoscopic Esophagectomy Versus Open Transthoracic Esophagectomy for Resectable Esophageal Cancer: A Randomized Controlled Trial. *Ann Surg.* 2019;269(4):621–30.
92. Shah AA, Gahan JC, Sorokin I. Comparison of Robot-Assisted Versus Open Simple Prostatectomy for Benign Prostatic Hyperplasia. *Curr Urol Rep.* 2018;19(9):1–8.
93. Haglind E, Carlsson S, Stranne J, Wallerstedt A, Wilderäng U, Thorsteinsdottir T, et al. Urinary Incontinence and Erectile Dysfunction after Robotic Versus Open Radical Prostatectomy: A Prospective, Controlled, Nonrandomised Trial. *Eur Urol.* 2015;68(2):216–25.
94. Du Y, Long Q, Guan B, Mu L, Tian J, Jiang Y, et al. Robot-assisted radical prostatectomy is more beneficial for prostate cancer patients: A system review and meta-analysis. *Med Sci*

Monit. 2018;24:272–87.

95. Cockrell R, Lee DI. Robot-Assisted Simple Prostatectomy: Expanding on an Established Operative Approach. *Curr Urol Rep*. 2017;18(5):1–7.
96. Saika T, Miura N, Fukumoto T, Yanagihara Y, Miyauchi Y, Kikugawa T. Role of robot-assisted radical prostatectomy in locally advanced prostate cancer. *Int J Urol*. 2018;25(1):30–5.
97. Matei D V., Brescia A, Mazzoleni F, Spinelli M, Musi G, Melegari S, et al. Robot-assisted simple prostatectomy (RASP): Does it make sense? *BJU Int*. 2012;110(11 C):972–9.
98. Long JA, Poinas G, Fiard G, Leprêtre M, Delaitre-Bonnin C, Rébillard X, et al. Prostatectomie radicale laparoscopique robot-assistée : quelles sont les preuves à l’heure d’une demande de nomenclature spécifique ? *Prog en Urol [Internet]*. 2017;27(3):146–57 [[acesso](#) em: 6 abr. 2021].
99. Coughlin GD, Yaxley JW, Chambers SK, Occhipinti S, Samaratunga H, Zajdlewicz L, et al. Robot-assisted laparoscopic prostatectomy versus open radical retropubic prostatectomy: 24-month outcomes from a randomised controlled study. *Lancet Oncol [Internet]*. 2018;19(8):1051–60 [[acesso](#) em: 6 abr. 2021].
100. Fanaroff AC, Califf RM, Harrington RA, Granger CB, McMurray JJV, Patel MR, et al. Randomized Trials Versus Common Sense and Clinical Observation: JACC Review Topic of the Week. *J Am Coll Cardiol*. 2020;76(5):580–9.

# 14. DERMATOLOGIA BASEADA EM EVIDÊNCIAS

## Introdução

A Dermatologia como especialidade médica deve ser pautada pelos mesmos padrões de rigor científico e preocupação com o melhor cuidado ao paciente. No entanto, fica claro que a MBE ainda se encontra em fases iniciais nessa especialidade, por alguns motivos que analisaremos neste capítulo.

A falta de evidências não pode limitar o cuidado aos pacientes nesta especialidade, porém o médico dermatologista precisa estar atento para encontrar o equilíbrio entre tratar da melhor forma seus pacientes, com as melhores evidências disponíveis, ao mesmo tempo em que a avidez por novidades não o faça cair na tentação de superestimar estudos inadequados ou sucumbir às falácias de vieses.

O caminho da Dermatologia baseada em evidências (DBE) não é fácil, tampouco ágil, uma vez que se estima um interstício de cerca de 17 anos na incorporação de uma evidência na rotina médica,<sup>1</sup> e como consequência, as condutas médicas que são oferecidas diariamente na prática da medicina são frequentemente divergentes daquelas baseadas em evidências.<sup>2</sup>

O futuro da DBE será muito importante para procurar o padrão ouro da ciência no cuidado das afecções cutâneas e traz consigo um grande horizonte para o médico que esteja bem orientado na busca por manter-se atualizado e valorizar o aprendizado contínuo.

## Dermatologia baseada em evidência

A pele é o maior órgão do corpo humano, está sujeito a mais de 3.000 doenças dermatológicas listadas no CID-10.<sup>3</sup> Estima-se que aproximadamente 20%-30% da população terá uma doença dermatológica ao longo da vida.<sup>4</sup>

A dermatologia baseada em evidência vem crescendo paulatinamente nos últimos anos, com o incentivo à valorização da metodologia científica e sua interpretação, por esforços como o livro – *“Evidence-based Dermatology”* –



de Hywel C. Williams e colaboradores,<sup>5</sup> e grupos de pesquisa como o *Cochrane Skin Group*.<sup>6</sup> Além disso, existem inúmeras diretrizes baseadas em evidência no âmbito da dermatologia, incluindo as apresentadas pela *American Academy of Dermatology* e pela *British Association of Dermatologists*, no entanto, o grau de adesão dos dermatologistas ao redor do mundo às diretrizes e melhores práticas é desconhecido, porque poucos estudos avaliaram a concordância entre a evidência clínica e a prática real nos consultórios, clínicas e hospitais.<sup>7</sup>

Uma solução abrangente para ultrapassar as lacunas de investigação-prática se fundamenta na identificação bem sucedida dessas lacunas e numa abordagem sistêmica para as preencher. A ciência da implementação, tal como definida pelo *National Institutes of Health* (NIH, nos EUA), é o estudo sistemático de processos e fatores concebidos para promover a “adoção e integração de práticas, intervenções e políticas baseadas em evidências científicas nos cuidados de saúde de rotina e nos contextos de saúde pública”.<sup>7-8</sup>

Semelhante à ciência da implementação, a desimplementação pode efetivamente evitar ou mitigar danos aos pacientes, melhorar os processos de atendimento e reduzir os custos de saúde. “*Choosing Wisely*” é uma campanha educacional sobre saúde com sede nos Estados Unidos, liderada pela *American Board of Internal Medicine Foundation* que fez parceria com sociedades de especialidades médicas, incluindo a *American Academy of Dermatology*, para formular e divulgar 10 recomendações para evitar exames, tratamentos e procedimentos desnecessários (*American Academy of Dermatology Workgroup*, 2019) – um esforço de desimplementação. De forma similar, em 2020, a Sociedade Brasileira de Dermatologia aderiu ao “*Choosing Wisely*” e estabeleceu 10 recomendações em via de submissão para publicação. Existem muitas oportunidades para realizar a desimplementação em dermatologia; por exemplo, há uma necessidade de abordar o uso excessivo de antibióticos perioperatórios para cirurgias dermatológicas<sup>9</sup> e uma necessidade de reduzir o uso inadequado da cirurgia de Mohs.<sup>10-12</sup>

As barreiras para a realização de estudos de implementação ou desimplementação em dermatologia incluem a falta de compreensão da utilidade e dos métodos para a realização desse tipo de pesquisa, o equívoco de que simplesmente gerar evidências é suficiente para mudar a prática clínica e a necessidade de evidências mais robustas fundamentadas em

ensaios clínicos randomizados e estudos de eficácia comparativa, que possam definir melhor a lacuna entre as evidências e as práticas médicas atuais.<sup>7</sup>

A complexidade do entendimento da dermatologia também compreende características da especialidade como o fato de lidar com: (a) cenários não só relacionados a estados de doença; (b) manutenção da pele saudável em pessoas que buscam intervenções; (c) a variabilidade dos diferentes tipos de pele humana; (d) a grande quantidade de doenças que não apresenta ainda consensos diagnósticos ou terapêuticos; (e) um enorme número de doenças raras e crônicas que impõem obstáculos à pesquisa médica; (f) o uso de drogas que estão no mercado farmacológico há muitas décadas e, portanto, não apresentam vantagem econômica para serem objeto de vultosos investimentos para ensaios clínicos de qualidade científica elevada, dentre outros obstáculos no caminho da dermatologia baseada em evidências.

Nos próximos parágrafos vamos analisar alguns destes tópicos e buscar formas de melhor contorná-los ou compreendê-los.

## O ensino da Dermatologia

De maneira geral, a Dermatologia clínica requer um baixo custo tecnológico. O início do raciocínio terapêutico depende da habilidade de reconhecer alterações de pele rapidamente e de forma mais acurada possível, o que, por sua vez, depende largamente do reconhecimento de padrões de lesões baseado nas experiências e treinamentos prévios. O desenvolvimento dessa habilidade visual, de um “olhar treinado” é um processo muito pouco compreendido e pouco reprodutível, e dessa forma não é ensinado formalmente.<sup>5</sup>

Grandes variações de modalidades de tratamentos para a mesma doença refletem, frequentemente, tradições locais,<sup>13</sup> preferências pessoais, e mesmo a ainda presente cultura de valorização da conduta sugerida por especialistas eminentes mesmo sem o conhecimento de seu embasamento.

## Doenças crônicas

Diversas doenças da Dermatologia são crônicas, e como tal, acompanharão os pacientes por longos períodos, senão por toda sua vida. Essa característica dificulta grandemente o desenho de estudos que reflitam condições de vida real. Por exemplo, uma revisão de literatura sobre

*guidelines* terapêuticos para dermatite atópica infantil avaliou que a maioria dos estudos conduzidos tem duração de menos de 12 semanas de acompanhamento dos pacientes, sendo apenas 4 das opções tratamento analisadas (corticoides tópicos, inibidores de calcineurina tópicos, imunoterapia sistêmica e tratamento combinado) chegaram a ser investigadas por intervalos maiores de 52 semanas.<sup>14</sup>

É uma característica de muitas doenças crônicas ter períodos de melhora e de piora, às vezes a despeito dos tratamentos instituídos; dessa forma um estudo a curto prazo não pode ser extrapolado. Como se tratam de intervenções que possivelmente serão realizadas por anos, deve-se atentar em especial aos efeitos a longo prazo e seu risco-benefício.

## Doenças raras

A Dermatologia conta com um grande número de doenças raras. Em pelo menos mil doenças raras ou muito raras, nenhum estudo randomizado foi descrito em literatura.<sup>5</sup> Esforços conjuntos e multicêntricos seriam necessários para suprir essa carência, porém há pouco esforço nesse sentido, atualmente.

Diante da falta de evidências robustas, dermatologistas se amparam apenas por estudos com baixo poder confirmatório, ou até mesmo apenas por relatos de casos ou séries de relatos.

## Desenho dos estudos disponíveis

Muitos dos desfechos usados nas análises de eficácia são desfechos “moles”, por exemplo para avaliar a intensidade de um sintoma (ex.: prurido), ou sinais (ex.: eritema ou descamação), que são muito difíceis de mensurar de forma objetiva e reprodutível. Essas medidas reduzem a qualidade das conclusões produzidas por estes estudos.<sup>15</sup> Vale lembrar que são raros os desfechos “duros” disponíveis na dermatologia, afinal, até avaliações como extensão de acometimento de lesões na psoríase (PASI – *psoriasis area and severity index*),<sup>16</sup> são feitas de forma sujeita à variabilidade do pesquisador por não possuírem um aparelho de medida padronizado.

A motivação e a expectativa do paciente ou do examinador podem influenciar nos resultados de qualquer tratamento, porém essa interferência é

maior ainda em estudos com desfechos moles, o que compromete a confiabilidade destes.

O desenho dos estudos também é frequentemente dificultado pelo fato de que é muito difícil cegar os grupos de intervenções, pois os pacientes monitoram de perto seus próprios resultados,<sup>5</sup> afinal trata-se do tegumento externo facilmente avaliado todos os dias. Dessa forma é muito difícil separar o efeito puro da intervenção sem fatores de confusão.

## Drogas antigas

Muitas das drogas mais empregadas na dermatologia, como por exemplo os corticoides e os imunossuppressores, são medicações antigas. Oriundos de uma época em que não se falava em estudos duplo-cego randomizados, meta-análises ou níveis de evidência.

Os glicocorticoides foram descobertos no fim dos anos 1940,<sup>17</sup> sendo que o uso sistêmico de cortisona e hidrocortisona data de 1950-51<sup>18</sup> para insuficiência adrenal e artrite reumatoide, e o primeiro uso tópico em dermatologia data de 1952,<sup>19</sup> sendo comercializado com sucesso apenas em 1960. Naquela ocasião a medicina baseada em evidências ainda estava a 20 anos de nascer.<sup>20</sup>

O metotrexato foi inicialmente descoberto em 1940 como quimioterápico, e usado pela primeira vez em artrite reumatoide em 1951 devido à descoberta de seu potencial anti-inflamatório e inibidor da resposta proliferativa dos tecidos conectivos.<sup>21</sup> Seu emprego na dermatologia se deu por acaso, por uma observação de pacientes portadores de psoríase que melhoravam após início de seu tratamento antineoplásico com a nova droga.<sup>22</sup> Seu uso foi consolidado para o tratamento, mesmo que a imensa maioria das evidências sejam de menor nível de confiabilidade com estudos retrospectivos e séries de casos, e também *trials* não randomizados.<sup>22</sup> Hoje em dia essa droga também é muito usada no tratamento de vasculites, afecções bolhosas e linfoproliferativas.

A ciclosporina foi descoberta em meio a controvérsias em 1976, primeiramente para uso como imunossupressora pós-transplante.<sup>23</sup> Apenas em 1986 foi realizado o primeiro estudo controlado para seu uso em dermatologia,<sup>24</sup> e em 1997 chegou a ser aprovada para tratamento de psoríase pelo FDA americano.<sup>25</sup> No entanto é usada com frequência também para tratamento de diversas outras doenças inflamatórias, incluindo dermatite atópica, doenças bolhosas e doenças do tecido conjuntivo.

Outro ônus carregado pelas drogas antigas que usamos em dermatologia é o fato de que os poucos estudos disponíveis sobre elas não costumam abranger todas as doenças em que, na prática, são usadas para tratar. O uso *off label* é extremamente comum e aceito dentro da dermatologia clínica. Evidências mais fracas e a consolidação do tempo foram permitidas para justificar que essas drogas antigas dificilmente vão despertar o interesse (científico e financeiro) em atualizar suas evidências para o século XXI.

## O crescente papel das pesquisas financiadas pela indústria

É inegável que o maior fomentador de estudos atualmente é a indústria farmacêutica. Isso ocorre por alguns motivos, como a indisponibilidade de recursos públicos ou mesmo de outras fontes privadas para investir em grandes *trials*; assim como a pouca valorização que ainda é dada aos pesquisadores em termos de remuneração e investimento do tempo laboral. Esse último ponto vem, aos poucos, mudando principalmente nos países mais desenvolvidos. No Brasil ainda está longe de ser uma realidade.

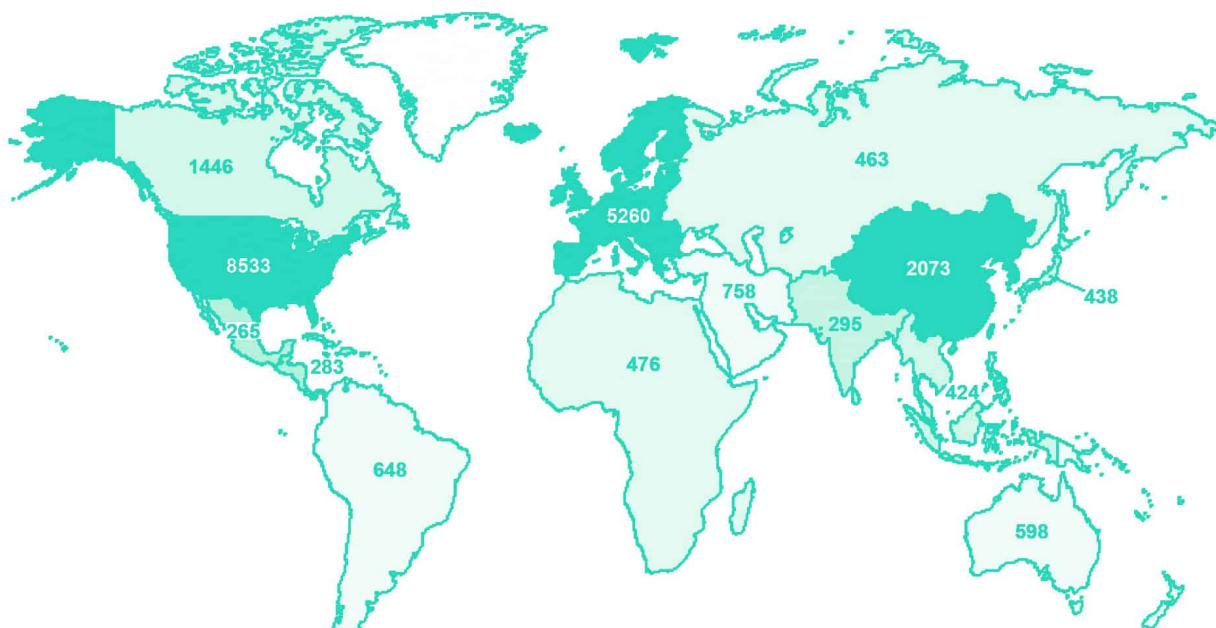
A empresa dificilmente terá interesse em conduzir um estudo custoso sobre uma medicação que já perdeu a patente, para que seu concorrente também possa se beneficiar do resultado. O objetivo é provar que as novas moléculas trazidas ao mercado são boas como tratamento, inclusive causando um fenômeno prevalente de desenhar estudos que comparam drogas a placebos apenas, e não aos demais tratamentos existentes ou ao padrão ouro de terapia. Dificultando ainda mais sua implementação no dia a dia do dermatologista.

Talvez no dia em que a dermatologia baseada em evidência esteja bem integrada na prática diária, torne-se mais fácil e valorizado conduzir pesquisas baseadas nas necessidades clínicas, mais do que nos interesses comerciais.

Infelizmente, países em desenvolvimento ou subdesenvolvidos, por inúmeras questões, muitas delas envolvendo logística ou questões de agências reguladoras, têm ocupado papel coadjuvante no processo de ensaios clínicos com novas drogas. Esse cenário pode ser visualizado na Figura 1, obtida no site [clinicaltrials.gov](http://clinicaltrials.gov),<sup>26</sup> que registra os estudos clínicos em sua plataforma e que demonstra que enquanto o Canadá tem 1.446 estudos clínicos envolvendo medicamentos para doenças cutâneas (*skin diseases*), a

América Latina inteira registrou 648 estudos apenas. Os trials registrados podem ser consultados no [Clinicaltrials.gov](https://clinicaltrials.gov)<sup>27</sup>.

**Figura 1**



Mapa disponível de todos os estudos registrados no ClinicalTrials.gov até 18 de fevereiro de 2021 (18.444 estudos) usando o termo de pesquisa “Dermatologic Disease”. O Brasil tem atualmente 405 estudos registrados na plataforma.

## O risco do mecanicismo

Neste livro você já aprendeu sobre os pensamentos mecanicista, frequentista e bayesiano. Na dermatologia encontramos um consagrado exemplo para nos lembrar dos riscos do mecanicismo na tomada de decisão terapêutica e a necessidade de testes de qualidade antes da adoção de novas terapias.

A talidomida foi criada em 1953 inicialmente para o tratamento de irritabilidade, ansiedade, dificuldade de concentração, depressão, hipotireoidismo e náusea; porém logo foi descoberto seu potencial anti-inflamatório e imunomodulador ao inibir seletivamente o fator de necrose tumoral (TNF-alfa), reduzir a quantidade de T-helper e aumentar as células T supressoras.<sup>28</sup>

A necrose epidérmica tóxica (NET) é uma reação imunológica cutânea grave associada a uma letalidade de aproximadamente 30%,<sup>29-30</sup> geralmente causada por reação adversa a alguma droga. O paciente tem um intenso

acometimento do estado geral, avulsão de pele e mucosas assemelhando-se a um grande queimado. O fator de necrose tumoral (TNF-alfa) tem sido implicado na patogênese da NET.

Em 1998 foi publicado no *The Lancet* um estudo que pretendia testar a eficácia e segurança da talidomida em controlar o processo de necrose, devido ao seu conhecido efeito anti-TNF-alfa e promissor resultado em outras patologias inflamatórias. Foi desenhado um estudo duplo-cego, randomizado, placebo-controlado para investigar essa questão.

O estudo<sup>31</sup> foi interrompido precocemente devido a uma mortalidade excessiva no grupo que recebeu a talidomida: 10 dos 12 pacientes evoluíram a óbito, comparado com 3 dos 10 do grupo placebo (RR= 2,78, p= 0,03). Paradoxalmente, foi observado que o plasma dos pacientes do grupo talidomida tiveram uma concentração mais elevada de TNF-alfa do que o grupo placebo já no segundo dia de tratamento.

Essa é mais uma situação da medicina baseada em evidências que nos relembra e reafirma a importância de não se ater exclusivamente à lógica quando lidamos com processos médicos, especialmente aqueles que não são visíveis a olho nu, como são as vias metabólicas.

## Cochrane skin group

O *Cochrane skin group*<sup>32</sup> é uma colaboração internacional que faz parte da *Cochrane Collaboration*, cujo objetivo é garantir que informações da melhor qualidade estejam disponíveis sobre os efeitos de intervenções de saúde para todos os médicos do mundo.<sup>33</sup> O *Skin group* é focado em realizar revisões sistemáticas sobre diversos manejos de doenças de pele que julguem importantes para profissionais dermatologistas, ou mesmo leigos com interesse na área.

Os grupos de revisão colaborativa são compostos por um corpo internacional e multidisciplinar de voluntários que se dedicam principalmente à produção de revisões sistemáticas e à disseminação dessas revisões para estimular a aplicação da medicina baseada em evidências na prática clínica e na informação de pacientes.

Antes de iniciar uma revisão, os grupos formados são obrigados a submeter um protocolo declarando os métodos propostos para levantamento de literatura, critérios de inclusão de referências, os tipos e métodos de análise da qualidade dos trabalhos incluídos. Esse protocolo é então publicado na *Cochrane Library* de forma que todos possam avaliar a

proposta de pesquisa. Essa política reduz a possibilidade de análises *post-hoc*<sup>34</sup> e reduz a tentação para revisores de serem guiados pelas evidências ao invés de pela pergunta proposta.

Os revisores, então, extraem as evidências consideradas relevantes de acordo com o protocolo já declarado e as inserem no software do *Cochrane Review Manager* (RevMan).<sup>35</sup> Quando relevante, meta-análises dos estudos são também realizadas. Ao final, as revisões sistemáticas finalizadas são publicadas no *Cochrane Database of Systematic Reviews*<sup>36</sup> com acesso aberto imediato. Idealmente as revisões são regularmente atualizadas, conforme a disponibilidade de novas evidências de qualidade; porém isso é altamente dependente na disponibilidade dos revisores voluntários.

Outra importante atividade do *Skin group* é construir um banco de dados de ensaios clínicos randomizados e não randomizados que foram aprovados por um rigoroso controle de qualidade científica. Este banco de dados é conhecido como *Specialized Skin Register* e visa tornar disponíveis eletronicamente as evidências publicadas, não publicadas, e as disponíveis apenas da forma impressa.

Vários papéis estão disponíveis para interessados em participar como membros do *Skin group*, como pesquisadores, tradutores, revisores e correvisores, disseminadores e subeditores, avaliadores de conteúdo, consumidores e voluntários. Estes papéis são resumidos da Tabela 1.<sup>37</sup>

**Tabela 1.**

Categoria	Descrição
Pesquisadores	Pesquisa manual da literatura para identificar todos os <i>trials</i> relevantes (o <i>Cochrane Collaboration</i> está trabalhando com outras instituições para criar um registro internacional de todos os ensaios clínicos randomizados na área da saúde).
Tradutores	Traduzir <i>trials</i> relevantes não publicados em inglês e deixá-los disponíveis para revisão.
Revisores e correvisores	Sumarizar as evidências em revisões de qualidade, mantê-las atualizadas e responder às críticas. Isso envolve desenvolver os protocolos, identificar <i>trials</i> relevantes, decidir quais incluir e extrair as informações importantes, inseri-las no software de revisão RevMan, adicionar novas evidências à medida que se tornam disponíveis e responder a comentários e críticas da comunidade.
Disseminadores e subeditores	Disponibilizar os resultados das revisões de forma rápida e acessível com linguagem não técnica para consumidores, gestores e outros grupos.



Categoria	Descrição
Avaliadores de conteúdo	Averiguar que o conteúdo dos protocolos e das revisões está à altura dos rigorosos critérios da Cochrane sobre validade, relevância, aplicabilidade e generalizabilidade.
Consumidores	O <i>Cochrane Skin Group</i> incentiva os consumidores e seus representantes a participar como revisores, co-revisores, avaliadores, tradutores e disseminadores.
Voluntários	Suporte administrativo e apoio para o <i>Cochrane Skin Group</i> .

Papéis dos membros do “Skin Group”.<sup>37</sup>

## Boas aplicações de MBE na Dermatologia

A psoríase foi uma das patologias que recebeu grandes investimentos recentemente, com a descoberta de novos agentes de tratamentos nas últimas duas décadas. As drogas imunobiológicas anti-TNF e anti-interleucinas receberam a atenção especial dos pesquisadores e o aporte da indústria farmacêutica.

A maioria das pesquisas realizadas sobre imunobiológicos na psoríase atendeu às expectativas da MBE. Foram realizados estudos duplo-cegos, randomizados, placebo-controlados, comparativos com outras terapias; tudo que o dermatologista esperava para a dermatologia baseada em evidências. O adalimumabe foi o objeto do estudo CHAMPION,<sup>38</sup> mostrando os resultados de eficácia e segurança quando randomizado com placebo e metotrexato, e analisado após 16 semanas. Outro trabalho avaliou sua eficácia em aplicações por até 48 semanas num estudo multicêntrico, randomizado, duplo-cego, placebo-controlado.<sup>39</sup> O etanercepte foi estudado não apenas contra placebo, mas também contra diferentes doses dele próprio.<sup>40-41</sup> Estudos satisfatórios foram conduzidos para definir também as doses de indução mais eficazes para o infliximab.<sup>42-43</sup> O ustekinumab foi amplamente estudado no PHOENIX 1<sup>44</sup> e PHOENIX 2<sup>45</sup> em estudos duplo-cegos, randomizados, placebo-controlados durante 76 e 52 semanas respectivamente.

Outra droga com efeitos bem documentados é o Omalizumabe, um anticorpo monoclonal anti-IgE aplicado no tratamento da urticária com cada vez mais frequência. Uma excelente revisão sistemática foi publicada em 2019<sup>46</sup> mostrando a força de diversos estudos<sup>47-52</sup> em recomendar o

omalizumabe para pacientes refratários a altas doses de anti-histamínicos; uma prática que já se mostra cada dia mais sedimentada na prática clínica dermatológica.

Não apenas no emprego de novas drogas é aplicada a dermatologia baseada em evidências. O tratamento cirúrgico dos cânceres de pele melanoma e não melanoma tem suas margens cirúrgicas de excisão bem pesquisadas,<sup>53</sup> havendo recomendações precisas e atualizações periódicas das evidências.<sup>54</sup>

Em estudo de meta-análise, Riddle e colaboradores<sup>56</sup> abordaram o manejo cirúrgico da hidradenite supurativa (HS). Uma meta-análise sobre as taxas de recorrência foi realizada nos estudos selecionados. Dos 715 estudos selecionados, 59 foram incluídos na revisão e 33 na meta-análise. Vinte e dois estudos sobre excisão ampla tiveram as menores taxas de recorrência, de 8% (IC 95% 2-16%); excisão local teve as maiores taxas em 34% (IC 95% 20-59%). Para estudos de rescisão ampla ou radical, técnicas de retalho tiveram a menor recorrência com 0% (IC 95% 0-4%); fechamento primário teve as maiores com 38% (IC 95% 20-59%). Excisões amplas e reparos em retalho foram associados com as menores recorrências de HS pós-cirúrgico, no entanto, isso deve ser pesado contra uma morbidade potencial mais alta de procedimentos extensos. Heterogeneidade e limitações metodológicas de evidências limitam a habilidade de tirar uma conclusão concreta sobre riscos de recorrência associados a procedimentos cirúrgicos na hidradenite supurativa.

Uma doença comum na prática dermatológica é a acne vulgar. As abordagens terapêuticas medicamentosas e não medicamentosas são inúmeras. Shi e colaboradores<sup>57</sup> realizaram uma meta-análise destas abordagens. Foram incluídos *trials* controlados e randomizados, que comparassem antibióticos tópicos (AT), peróxido de benzoíla (BP), retinoides tópicos (RT), antibióticos orais (AO), lasers, dispositivos de luz incluindo os de LED (LED), terapia fotodinâmica (TF), luz intensa pulsada, *peelings* químicos (PQ), outras terapias diversas ou medicina alternativa e complementar (TDMAD); ou a combinação destas. Foi realizada uma meta-análise bayesiana com efeitos randomizados para todos os tratamentos comparados com placebo e entre si. Uma meta-análise aditiva frequentista foi realizada para avaliar a robustez dos resultados e interações potenciais. Uma análise de sensibilidade foi realizada com diferentes antecedentes, e uma metarregressão para ajustar 9 potenciais modificadores de efeitos. Para

efeitos colaterais, os antibióticos orais tiveram mais risco em tratamentos combinados. Para regressão de lesões não inflamatórias, 17 intervenções tiveram diferenças significativas quando comparadas com placebo, e 3 intervenções (RT+BP; MD = -21.89, [IC 95%: -28.97, -14.76]; RT+PB+TDMAD: -22.48 [IC 95%: -34.13, -10.70]; AT+PB+PQ: -20.63 [IC 95%: -33.97, -7.13]) mostraram melhores resultados. Para redução de lesões inflamatórias, 19 intervenções foram estatisticamente melhores do que placebo, e 3 (RT+PB: MD = -12.13, [IC 95%: -18.41, -5.80]; RT+PB+TDMAD: -13.21 [IC 95%: -23.39, -3.04]; LED: -11.30 [IC 95%: -18.34, -4.42]) foram superiores às demais. Em resumo, nas lesões não inflamatórias e nas inflamatórias, RT+PB e AT+PB foram as melhores opções quando comparadas, e devem ser consideradas como as melhores indicadas para tratamento de acne leve a moderada. LED é uma opção para lesões inflamatórias quando há resistência a drogas. Todas as combinações envolvendo AO mostraram maior risco de reações adversas.

De forma similar em relação à acne vulgar Liu e colaboradores<sup>58</sup> avaliaram outros agentes tópicos no tratamento da acne, como o ácido azelaico, ácido salicílico, nicotinamida, enxofre, zinco, e alfa-hidróxi-ácido; através de uma meta-análise de estudos randomizados controlados, seguindo os passos recomendados na MBE. A conclusão foi de que as evidências disponíveis eram de qualidades moderada, baixa e muito baixa para estes tratamentos tópicos, e o risco de confusão e imprecisão limitaram a confiabilidade das evidências.

A Dermatite Atópica (DA) é uma condição que acomete até 20% das crianças e 3% dos adultos. Siegels e colaboradores<sup>59</sup> analisaram os estudos controlados e randomizados para tratamento da doença moderada a grave até fevereiro de 2020. Os desfechos primários analisados foram melhora de sinais clínicos, sintomas de DA e questionários de qualidade de vida. Foram encontradas evidências para apremilast, azatioprina (AZA), baricitinib, ciclosporina (CSA), corticosteroides, dupilumab, interferon-gamma, imunoglobulina intravenosa (IGIV), mepolizumab, metotrexate (MTX), omalizumab, upadacitinib e ustekinumab. Meta-análises também indicaram eficácia para o uso de baricitinib e dupilumab a curto prazo. Análises de eficácia de AZA e CSA indicaram superioridade ao placebo no curto prazo, porém o uso de escores não validados utilizados impede a comparação. A conclusão foi de que o estudo mais robusto e de maior qualidade de evidência foi de eficácia e segurança do uso de dupilumab por até 1 ano em

adultos. Evidências satisfatórias também foram encontradas para AZA, baricitinib e CSA. Restrições de metodologias usadas nos demais tratamentos impediram conclusões baseadas em evidências.

As revisões sistemáticas são importantes a fim de esclarecer debates sobre controvérsias. Um exemplo é o uso do método de acupuntura aplicado ao tratamento dermatológico. Hwang & Lio publicaram uma revisão deste tema em 2021.<sup>60</sup> Uma revisão sistemática havia sido publicada em 2015 concluindo que a acupuntura trazia benefícios em diversas doenças dermatológicas.<sup>61</sup> Hwang & Lio apresentaram uma revisão dos novos estudos publicados desde então para atualizar as evidências. A conclusão foi de que, apesar da existência de alguns estudos que apoiassem o uso da acupuntura, ainda era necessário a realização de estudos em larga escala, randomizados e controlados para apoiar esta terapia como evidência de alto nível.

Mesmo em condições raras em dermatologia, como no caso de reações medicamentosas graves como a Necrólise Epidérmica Tóxica (NET) e a Síndrome de Stevens-Johnson (SSJ), o obstáculo à obtenção de evidências robustas não está dependente apenas da dificuldade de reunir-se um número adequado de doentes no estudo de uma intervenção, mas também a qualidade destes estudos. Torres-Navarro e colaboradores<sup>62</sup> nos demonstram isso em uma recente revisão sistemática da literatura que envolveu um número expressivo de 2.466 doentes com NET/SSJ. Seis meta-análises foram realizadas em pacientes que receberam suporte clínico apenas, ou em combinação com drogas imunomoduladoras: corticosteroides, ciclosporina, etanercept, imunoglobulinas ou uma combinação de corticosteroides e imunoglobulinas. Uma metarregressão multivariada também foi realizada. Os resultados encontrados foram: -0.13 (IC 95%, -0.42,0.16) para corticosteroides, -0.39 (IC 95%, -0.87,0.09) para imunoglobulinas, 0.13 (IC 95%, -0.15,0.40) para terapia de suporte, -0.88 (IC 95%, -1.47, -0.29) para ciclosporina, -0.95 (IC 95%, -1.82, -0.07) para etanercept e - 0.56 (IC 95%, -0.94, -0.19) para imunoglobulinas associadas a corticosteroides. A metarregressão confirmou que a ciclosporina e imunoglobulinas associadas a corticosteroides estavam associadas a menor mortalidade do que a predita por SCORTEN.<sup>63</sup> A conclusão foi de que não há evidências suficientes para recomendar nenhuma terapia para uso em NET/SSJ.

Cosmiatria

A cosmiatria é a área da dermatologia que realiza procedimentos e tratamentos que tenham como finalidade a manutenção da beleza e a melhora da aparência da pele e seus anexos segundo a Sociedade Brasileira de Dermatologia.<sup>64</sup> A cosmiatria é, por definição, uma área de atuação médica, e como tal deveria buscar ser pautada nos mesmos padrões de conhecimento técnico e científico que exigimos para todas as demais áreas da medicina.

Os riscos da falta de pesquisas no tema também nos levam à falta de padronização da aplicação de diversas técnicas, deixando a segurança do paciente em risco e a cargo de cada médico individualmente.

Galanin e colaboradores<sup>65</sup> levantou bons exemplos de porquê devemos manter a cautela no estudo da dermatologia estética. Em primeiro lugar, o envelhecimento da face não é um processo completamente compreendido. Tem sido descrito como “perda de volume”, “queda gravitacional”, “flacidez da pele”, “remodelamento ósseo” e “qualidade do tecido adiposo”, porém a verdade é que o melhor entendimento do envelhecimento provém de uma mistura de todos estes conceitos, sabendo que nenhum por si só explica o complexo processo do envelhecimento cutâneo. Por esse motivo, não há nenhum método de tratamento que possa ser considerado como superior, visto que cada tecnologia busca agir em um modelo diferente. O fato de tratar-se de um desfecho extremamente subjetivo, e inclusive, sujeito a variabilidades culturais, que é a aparência de saúde e jovialidade da pele, combinado ao fato de as técnicas disponíveis serem substancialmente divergentes; é muito difícil imaginar um desenho de estudo satisfatório nessa área na DBE.

Outro item lembrado por Galanin é o risco de iatrogenias, pela falta de estudos e análises de segurança a longo prazo. Por exemplo, a criolipólise é uma técnica que causa apoptose de adipócitos localizados pelo resfriamento com o aparelho; apesar de liberada pelo FDA americano, vêm sendo cada vez mais relatados efeitos colaterais tardios como hiperplasia/hipertrofia adiposa paradoxal (HAP). A HAP é um aumento tardio de tecido adiposo no local do tratamento anterior, com adipócitos desorganizados, de diferentes tamanhos e formas, com aumento da espessura de septos e da vascularização; provavelmente causado pela indução de hipertrofia pela hipóxia nos adipócitos sobreviventes.

A cosmiatria trata de uma pele, a princípio, saudável ou sem graves comprometimentos, porém que busca intervenções de melhoria. No contexto

da medicina estética, o princípio que nos guia deveria abranger e ir além da não maleficência, sendo extremamente importante a demonstração de benefício positivo para o paciente.<sup>66</sup>

Em oposição, em relação aos tratamentos para cicatrizes e sequelas da acne a literatura tem fornecido níveis de evidência científica para embasar alguns procedimentos e abordagens terapêuticas. Gupta e colaboradores<sup>67</sup> revisaram este tema com o objetivo de escolher as melhores técnicas para cada tipo de cicatriz e cada tipo de pele. Essa revisão analisou criticamente os tratamentos disponíveis para definir o nível de indicação de cada uma de acordo com suas evidências. Os melhores níveis de evidência encontrados foram para microagulhamento, radiofrequência fracionada, CO<sub>2</sub> fracionado, e laser erbium-doped yttrium aluminium garnet (Er:YAG) para cicatrizes leves a moderadas. Os níveis de evidência estão descritos na Figura 2.

### **Figura 2.**

Manejo de cicatrizes de acne baseado em todos os tipos de lesões	
Cicatrizes pigmentadas superficiais	1. Peelings químicos: Ácidos lático, salicílico, Jessner, Pirúvico, Glicólico, Mandélico + Salicílico (E3-4)
Cicatrizes superficiais eritematosas	1. Lasers – Laser de corante pusado (E2-3) 2. Luz intensa pulsada (E2-3)
Cicatrizes em furador de gelo ( <i>icepick</i> )	1. Peelings químicos – Mandélico + Salicílico (E1), Ácido Glicólico (E1) 2. CROSS ATA (E2-3) 3. Micoagulamento (E2-3) 4. Radiofrequência micoagulhada (E2-3) 5. Lasers-ER: YAG (E1-2), CO2 (E1-2), NAFL (E1-2) 6. Subcisão (E1-2) 7. Punch (E2-3) 8. PRP + micoagulamento ou lasers ablativos (E2-3) 9. Terapias combinadas (E1-4)
Cicatrizes em forma de caixa ( <i>boxcar</i> )	1. Peelings químicos – Mandélico_AS (E1-2), AG(E1-2) 2. CROSS ATA (E2-3) 3. Micoagulamento (E2-3) 4. Dermoabrasão (E1-3) 5. Microdermoabrasão (E1-2) 6. Radiofrequência micoagulhada (E2-3) 7. Lasers-ER: YAG (E2-3), CO2(E2-3), NAFL (E2-4) 8. Subcisão (E2-3) 9. Punch (E2-4) 10. PRP + micoagulamento ou lasers ablativos (E2-3) 11. Terapias combinadas (E1-4)
Cicatrizes onduladas ( <i>rolling</i> )	1. CROSS TCA (E2-3) 2. Micoagulamento (E2-3) 3. Dermoabrasão (E1-3) 4. Microdermoabrasão (E1-2) 5. Micoagulamento (E3-4) 6. Radiofrequência micoagulhada (E2-3) 7. Lasers-ER: YAG (E3-4), CO2(E3-4), NAFL (E2-3) 8. Subcisão (E3-4) 9. PRP + micoagulamento ou lasers ablativos (E2-3) 10. Terapias combinadas (E1-4)
Cicatrizes atróficas	1. Enxerto autólogo de gordura com ou sem PRP (E1-3) 2. Preenchedores sintéticos (E1-3) 3. Injeção de células dérmicas (E1-3) 4. Fibroblastos cultivados (E1-3)
Cicatrizes hipertróficas	1. Injeção intralesional de triancinolona e 5-FU (E2-3) 2. Crioterapia com nitrogênio líquido (E2-3) 3. Crioterapia fracionada (E2-3) 4. Laseres ablativos fracionados (E1-2) 5. Excisão e fechamento (E2-3) 6. Toxina botulínica (E1-2)
Cicatrizes lienares	1. Excisão e fechamento (E2-3)

Níveis de evidência dos tratamentos de cicatrizes de acne.

Muitos dos tratamentos estéticos disponíveis no mercado hoje prometem rejuvenescer ou melhorar a pele sem evidências científicas para apoiá-los. A crescente oferta de procedimentos com benefícios não comprovados traz um grande risco de prejudicar a confiança nos médicos dermatologistas, não apenas no que se refere à área de atuação da cosmiaatria, mas também da dermatologia clínica e cirúrgica. Porém a boa aplicação do uso de evidências nos tratamentos cosmiátricos tende a consolidar a MBE como padrão a ser seguido por toda a dermatologia.

## Conclusão

O que fazer quando confrontados pela necessidade de oferecer conselhos e tratamentos médicos *versus* a limitação das evidências falhas disponíveis? O melhor que podemos fazer, na prática da dermatologia baseada em evidências, é conhecer os limites da compreensão atual, aconselhar e tratar com a melhor ciência disponível, e alertar sempre sobre as limitações.

Devemos contribuir para a valorização da MBE formulando perguntas clínicas bem construídas, encontrando as melhores respostas disponíveis para essa pergunta, avaliando criticamente as evidências, aplicando as conclusões aos pacientes propostos e sempre zelando pela disseminação das boas evidências científicas.<sup>68</sup>

## Referências bibliográficas

1. Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: understanding time lags in translational research. *J R Soc Med.* 2011 Dec;104(12):510–20.
2. Ashrafzadeh S, Metlay JP, Choudhry NK, Emmons KM, Asgari MM. Using Implementation Science to Optimize the Uptake of Evidence-Based Medicine into Dermatology Practice.
3. ICD-10 Version:2016.
4. Nijsten T, Stern RS. How epidemiology has contributed to a better understanding of skin disease. Vol. 132, *Journal of Investigative Dermatology.* Nature Publishing Group; 2012. p. 994–1002.
5. Williams H, Bigby M, Diepgen T, Herxheimer A, Naldi L, Rzany B. Evidence-Based Dermatology. Wiley; 2009. (Evidence-Based Medicine).
6. Reddi A, Prescott L, Doney E, Delamere F, Kollipara R, Dellavalle RP, Williams HC. The Cochrane Skin Group: a vanguard for developing and promoting evidence-based dermatology. *J Evid Based Med.* 2013 Nov;6(4):236–42.



7. Ashrafzadeh S, Metlay JP, Choudhry NK, Emmons KM, Asgari MM. Using Implementation Science to Optimize the Uptake of Evidence-Based Medicine into Dermatology Practice. *J Invest Dermatol*. 2020 May;140(5):952–8.
8. Implementation science news, resources and funding for global health researchers – Fogarty International Center @ NIH.
9. Bae-Harboe Y-SC, Liang CA. Perioperative antibiotic use of dermatologic surgeons in 2012. *Dermatologic Surg Off Publ Am Soc Dermatologic Surg [et al]*. 2013 Nov;39(11):1592–601.
10. Connolly SM, Baker DR, Coldiron BM, Fazio MJ, Storrs PA, Vidimos AT, et al. AAD/ACMS/ASDSA/ASMS 2012 appropriate use criteria for Mohs micrographic surgery: a report of the American Academy of Dermatology, American College of Mohs Surgery, American Society for Dermatologic Surgery Association, and the American Society for Mohs S. *J Am Acad Dermatol*. 2012 Oct;67(4):531–50.
11. Asgari MM, Olson JM, Alam M. Needs assessment for Mohs micrographic surgery. *Dermatol Clin*. 2012 Jan;30(1):167–75, x.
12. Blechman AB, Patterson JW, Russell MA. Application of Mohs micrographic surgery appropriate-use criteria to skin cancers at a university health system. *J Am Acad Dermatol*. 2014 Jul;71(1):29–35.
13. Eedy DJ, Griffiths CEM, Chalmers RJG, Ormerod AD, Smith CH, Barker JNWN, Potter J, Ingham J, Lowe D, Burge S. Care of patients with psoriasis: an audit of U.K. services in secondary care. *Br J Dermatol*. 2009 Mar;160(3):557–64.
14. Siegfried EC, Jaworski JC, Mina-Osorio P. A Systematic Scoping Literature Review of Publications Supporting Treatment Guidelines for Pediatric Atopic Dermatitis in Contrast to Clinical Practice Patterns. Vol. 8, *Dermatology and Therapy*. Springer Healthcare; 2018. p. 349–77.
15. Eaglstein WH, Corcoran G. New Drugs and New Molecular Entities in Dermatology. Vol. 147, *Arch Dermatol*. 2011.
16. Robinson A, Kardos M, Kimball AB. Physician Global Assessment (PGA) and Psoriasis Area and Severity Index (PASI): Why do both? A systematic analysis of randomized controlled trials of biologic agents for moderate to severe plaque psoriasis. *J Am Acad Dermatol*. 2012;66(3):369–75.
17. Chast F. Histoire de la corticothérapie. Vol. 34, *Revue de Medecine Interne*. *Rev Med Interne*; 2013. p. 258–63.
18. Benedek TG. History of the development of corticosteroid therapy. *Clin Exp Rheumatol*. 2011;29(5 Suppl 68):S-5-12.
19. Camarasa JG, Giménez-Arnau A. Corticosteroids: topical. In: *European Handbook of Dermatological Treatments*. Springer Berlin Heidelberg; 2003. p. 731–8.
20. Sur RL, Dahm P. History of evidence-based medicine. In: *Indian Journal of Urology*. Wolters Kluwer -- Medknow Publications; 2011. p. 487–9.

21. Bedoui Y, Guillot X, Sélambarom J, Guiraud P, Giry C, Jaffar-Bandjee MC, Ralandison S, Gasque P. Methotrexate an old drug with new tricks. Vol. 20, *International Journal of Molecular Sciences*. MDPI AG; 2019.
22. Shen S, O'Brien T, Yap LM, Prince HM, McCormack CJ. The use of methotrexate in dermatology: A review. Vol. 53, *Australasian Journal of Dermatology*. Blackwell Publishing; 2012. p. 1–18.
23. Cheng M. Hartmann Stahelin (1925-2011) and the contested history of cyclosporin A. Vol. 27, *Clinical Transplantation*. Blackwell Publishing Ltd; 2013. p. 326–9.
24. Ellis CN, Gorsulowsky DC, Hamilton TA, Billings JK, Brown MD, Headington JT, Cooper KD, Baadsgaard O, Duell EA, Annesley TM. Cyclosporine improves psoriasis in a double-blind study. *JAMA*. 1986 Dec;256(22):3110–6.
25. Amor KT, Ryan C, Menter A. The use of cyclosporine in dermatology: Part i. Vol. 63, *Journal of the American Academy of Dermatology*. *J Am Acad Dermatol*; 2010. p. 925–46.
26. Search of: Skin diseases – Results on Map – [ClinicalTrials.gov](https://clinicaltrials.gov).
27. Search of: Skin diseases | (Map: Brazil) – List Results – [ClinicalTrials.gov](https://clinicaltrials.gov).
28. Hassan I, Dorjay K, Anwar P. Thalidomide in dermatology: Revisited. *Indian J Dermatol*. 2015 Apr;60(2):213.
29. Diphoorn J, Cazzaniga S, Gamba C, Schroeder J, Citterio A, Rivolta AL, Vighi GD, Naldi L. Incidence, causative factors and mortality rates of Stevens-Johnson syndrome (SJS) and toxic epidermal necrolysis (TEN) in northern Italy: data from the REACT registry. *Pharmacoepidemiol Drug Saf*. 2016 Feb;25(2):196–203.
30. Estrella-Alonso A, Aramburu JA, González-Ruiz MY, Cachafeiro L, Sánchez MS, Lorente JA. Toxic epidermal necrolysis: A paradigm of critical illness. Vol. 29, *Revista Brasileira de Terapia Intensiva*. Associação de Medicina Intensiva Brasileira – AMIB; 2017. p. 499–508.
31. Wolkenstein P, Latarjet J, Roujeau JC, Duguet C, Boudeau S, Vaillant L, Maignan M, Schuhmacher MH, Milpied B, Pilorget A, Bocquet H, Brun-Buisson C, Revuz J. Randomised comparison of thalidomide versus placebo in toxic epidermal necrolysis. *Lancet*. 1998 Nov;352(9140):1586–9.
32. [Welcome | Cochrane Skin](https://www.cochrane.org/skin).
33. Collier A, Johnson KR, Delamere F, Leonard T, Dellavalle RP, Williams H. The Cochrane Skin Group: Promoting the best evidence. Vol. 9, *Journal of Cutaneous Medicine and Surgery*. 2005. p. 324–31.
34. Zhang Y, Hedro R, Rivera A, Rull R, Richardson S, Tu XM. Post hoc power analysis: Is it an informative and meaningful analysis? *Gen Psychiatry*. 2019 Aug;32(4):100069.
35. [RevMan | Cochrane Training](https://www.cochrane.org/training).
36. [Search | Cochrane Library](https://www.cochrane.org/search).
37. Williams H, Adetugbo K, Li Wan Po A, Naldi L, Diepgen T, Murrell D. The Cochrane Skin Group: Preparing, maintaining, and disseminating systematic reviews of clinical

interventions in dermatology. Vol. 134, Archives of Dermatology. American Medical Association; 1998. p. 1620–6.

38. Saurat JH, Stingl G, Dubertret L, Papp K, Langley RG, Ortonne JP, Unnebrink K, Kaul M, Camez A. Efficacy and safety results from the randomized controlled comparative study of adalimumab vs. methotrexate vs. placebo in patients with psoriasis (CHAMPION). *Br J Dermatol.* 2008 Mar;158(3):558–66.
39. Gordon KB, Langley RG, Leonardi C, Toth D, Menter MA, Kang S, Heffernan M, Miller B, Hamlin R, Lim L, Zhong J, Hoffman R, Okun MM. Clinical response to adalimumab treatment in patients with moderate to severe psoriasis: Double-blind, randomized controlled trial and open-label extension study. *J Am Acad Dermatol.* 2006 Oct;55(4):598–606.
40. Van De Kerkhof PCM, Segaert S, Lahfa M, Luger TA, Karolyi Z, Kaszuba A, Leigheb G, Camacho FM, Forsea D, Zang C, Boussuge MP, Paolozzi L, Wajdula J. Once weekly administration of etanercept 50 mg is efficacious and well tolerated in patients with moderate-to-severe plaque psoriasis: A randomized controlled trial with open-label extension. *Br J Dermatol.* 2008 Nov;159(5):1177–85.
41. Leonardi CL, Powers JL, Matheson RT, Goffe BS, Zitnik R, Wang A, Gottlieb AB. Etanercept as Monotherapy in Patients with Psoriasis. *N Engl J Med.* 2003 Nov;349(21):2014–22.
42. Menter A, Feldman SR, Weinstein GD, Papp K, Evans R, Guzzo C, Li S, Dooley LT, Arnold C, Gottlieb AB. A randomized comparison of continuous vs. intermittent infliximab maintenance regimens over 1 year in the treatment of moderate-to-severe plaque psoriasis. *J Am Acad Dermatol.* 2007;56(1):31.e1-31.e15.
43. Gottlieb AB, Evans R, Li S, Dooley LT, Guzzo CA, Baker D, Bala M, Marano CW, Menter A. Infliximab induction therapy for patients with severe plaque-type psoriasis: A randomized, double-blind, placebo-controlled trial. *J Am Acad Dermatol.* 2004 Oct;51(4):534–42.
44. Leonardi CL, Kimball AB, Papp KA, Yeilding N, Guzzo C, Wang Y, Li S, Dooley LT, Gordon KB. Efficacy and safety of ustekinumab, a human interleukin-12/23 monoclonal antibody, in patients with psoriasis: 76-week results from a randomised, double-blind, placebo-controlled trial (PHOENIX 1). *Lancet.* 2008;371(9625):1665–74.
45. Papp KA, Langley RG, Lebwohl M, Krueger GG, Szapary P, Yeilding N, Guzzo C, Hsu MC, Wang Y, Li S, Dooley LT, Reich K. Efficacy and safety of ustekinumab, a human interleukin-12/23 monoclonal antibody, in patients with psoriasis: 52-week results from a randomised, double-blind, placebo-controlled trial (PHOENIX 2). *Lancet.* 2008;371(9625):1675–84.
46. Rubini NPM, Ensina LFC, Silva EMK, Sano F, Solé D. Effectiveness and safety of Omalizumab in the treatment of chronic spontaneous urticaria: Systematic review and meta-analysis. *Allergol Immunopathol (Madr).* 2019 Nov;47(6):515–22.

47. Kaplan A, Ledford D, Ashby M, Canvin J, Zazzali JL, Conner E, Veith J, Kamath N, Staubach P, Jakob T, Stirling RG, Kuna P, Berger W, Maurer M, Rosén K. Omalizumab in patients with symptomatic chronic idiopathic/spontaneous urticaria despite standard combination therapy. *J Allergy Clin Immunol*. 2013 Jul;132(1):101–9.
48. Saini S, Rosen KE, Hsieh H-J, Wong DA, Conner E, Kaplan A, Spector S, Maurer M. A randomized, placebo-controlled, dose-ranging study of single-dose omalizumab in patients with H1-antihistamine-refractory chronic idiopathic urticaria. *J Allergy Clin Immunol*. 2011 Sep;128(3):567-73.e1.
49. Staubach P, Metz M, Chapman-Rothe N, Sieder C, Bräutigam M, Maurer M, Weller K. Omalizumab rapidly improves angioedema-related quality of life in adult patients with chronic spontaneous urticaria: X-ACT study data. *Allergy*. 2018 Mar;73(3):576–84.
50. Finlay AY, Kaplan AP, Beck LA, Antonova EN, Balp M-M, Zazzali J, Khalil S, Maurer M. Omalizumab substantially improves dermatology-related quality of life in patients with chronic spontaneous urticaria. *J Eur Acad Dermatol Venereol*. 2017 Oct;31(10):1715–21.
51. Hide M, Park H-S, Igarashi A, Ye Y-M, Kim T-B, Yagami A, Roh J, Lee J-H, Chinuki Y, Youn SW, Lee S-K, Inomata N, Choi J-H, Fukunaga A, Wang J, Matsushima S, Greenberg S, Khalil S. Efficacy and safety of omalizumab in Japanese and Korean patients with refractory chronic spontaneous urticaria. *J Dermatol Sci*. 2017 Jul;87(1):70–8.
52. Tonacci A, Billeci L, Pioggia G, Navarra M, Gangemi S. Omalizumab for the Treatment of Chronic Idiopathic Urticaria: Systematic Review of the Literature. Vol. 37, *Pharmacotherapy*. Pharmacotherapy Publications Inc.; 2017. p. 464–80.
53. Wright FC, Souter LH, Kellett S, Easson A, Murray C, Toye J, McCready D, Nessim C, Ghazarian D, Hong NJL, Johnson S, Goldstein DP, Petrella T. Primary excision margins, sentinel lymph node biopsy, and completion lymph node dissection in cutaneous melanoma: a clinical practice guideline. *Curr Oncol*. 2019 Aug;26(4):e541–50.
54. Perez MC, Orcutt ST, Zager JS. Current standards of surgical management in primary melanoma. *G Ital di dermatologia e Venereol organo Uff Soc Ital di dermatologia e Sifilogr*. 2018 Feb;153(1):56–67.
55. Gresham LM, Marzario B, Dutz J, Kirchhof MG. An Evidence-Based Guide to SARS-CoV-2 Vaccination of Patients on Immunotherapies in Dermatology. *J Am Acad Dermatol*. 2021;
56. Riddle A, Westerkam L, Feltner C, Sayed C. Current Surgical Management of Hidradenitis Suppurativa: A Systematic Review and Meta-Analysis. *Dermatol Surg*. 2021 Jan;
57. Shi Q, Tan L, Chen Z, Ge L, Zhang X, Yang F, Liu C, Zhang J. Comparative Efficacy of Pharmacological and Nonpharmacological Interventions for Acne Vulgaris: A Network Meta-Analysis. *Front Pharmacol*. 2020;11:592075.
58. Liu H, Yu H, Xia J, Liu L, Liu G, Sang H, Peinemann F. Evidence-based topical treatments (azelaic acid, salicylic acid, nicotinamide, sulfur, zinc, and fruit acid) for acne: an abridged version of a Cochrane systematic review. *J Evid Based Med*. 2020 Nov;13(4):275–83.

59. Siegels D, Heratizadeh A, Abraham S, Binnmyr J, Brockow K, Irvine AD, Halken S, Mortz CG, Flohr C, Schmid-Grendelmeier P, Van der Poel L-A, Muraro A, Weidinger S, Werfel T, Schmitt J. Systemic treatments in the management of atopic dermatitis: A systematic review and meta-analysis. *Allergy*. 2020 Oct.
60. Hwang J, Lio PA. Acupuncture in Dermatology: An Update to a Systematic Review. *J Altern Complement Med*. 2021 Jan;27(1):12–23.
61. Ma C, Sivamani RK. Acupuncture as a Treatment Modality in Dermatology: A Systematic Review. *J Altern Complement Med*. 2015 Sep;21(9):520–9.
62. Torres-Navarro I, Briz-Redón Á, Botella-Estrada R. Systemic therapies for Stevens-Johnson Syndrome and Toxic Epidermal Necrolysis: a SCORTEN-based systematic review and meta-analysis. *J Eur Acad Dermatol Venereol*. 2020 Jun;
63. Bastuji-Garin S, Fouchard N, Bertocchi M, Roujeau JC, Revuz J, Wolkenstein P. SCORTEN: a severity-of-illness score for toxic epidermal necrolysis. *J Invest Dermatol*. 2000 Aug;115(2):149–53.
64. Campos de atuação – Sociedade Brasileira de Dermatologia.
65. Galanin I, Nicu C, Tower JI. Facial Fat Fitness: A New Paradigm to Understand Facial Aging and Aesthetics. *Aesthetic Plast Surg*. 2020 Sep;
66. Goh C. The need for evidence-based aesthetic dermatology practice. *J Cutan Aesthet Surg*. 2009;2(2):65.
67. Gupta A, Kaur M, Patra S, Khunger N, Gupta S. Evidence-based Surgical Management of Post-acne Scarring in Skin of Color. *J Cutan Aesthet Surg*. 2020;13(2):124–41.
68. Bigby M. Evidence-based medicine in dermatology. *Dermatol Clin*. 2000 Apr;18(2):261–76.

# 15. ORTOPEDIA BASEADA EM EVIDÊNCIAS

## Introdução

Imagine uma situação clínica: um paciente apresenta uma lesão no joelho e procura um ortopedista que prescreve, além das medicações, compressa de gelo e fisioterapia. Após um período, o paciente retorna com melhora parcial da dor. O ortopedista indica fortalecimento muscular, *kinesiotaping* e prescreve ácido hialurônico intra-articular. Parece um cenário rotineiro, não é? Mas os questionamentos que surgem são: compressa funciona mesmo? Se sim, qual a melhor? Quente ou fria? Fortalecimento muscular? *Kinesiotaping*? Ácido hialurônico? Esses tratamentos têm realmente respaldo científico?

## Compressa quente versus fria

O uso de termoterapia é bastante frequente na prática médica. No consultório ortopédico, a pergunta é rotineira “doutor, é melhor colocar compressa quente ou fria?”. Nos jogos de futebol televisionados que mostram atletas recebendo gelo logo após lesões aguçam ainda mais a curiosidade de profissionais e leigos. Afinal, termoterapia funciona? Se sim, qual a melhor? Fria ou quente? Qual o racional do seu uso?

O primeiro e mais importante passo é sabermos os efeitos do calor e do frio nas partes moles. Uma revisão sistemática sobre o efeito dos agentes termais em partes moles foi realizada por Bleakley.<sup>1</sup> Nesse estudo, se afirma que o calor altera as propriedades viscoelásticas dos tecidos colágenos, mostrando tendências de que o aquecimento aumenta o arco de movimento em uma variedade de articulações, o que não é visto na crioterapia. O autor critica o fato de que os trabalhos estudados têm alto risco de viés e amostra não adequada.

A crioterapia é comumente descrita como um procedimento para aliviar a dor e diminuir a inflamação em problemas musculoesqueléticos. O mecanismo da terapia fria para recuperação após o exercício é atribuído predominantemente ao seu efeito vasoconstritor, que reduz as reações

inflamatórias por meio da diminuição do metabolismo celular.<sup>2</sup> Esse mecanismo é o oposto do calor que gera vasodilatação tecidual.

A partir desse racional fisiológico, buscamos estudos que reforcem o uso de determinado agente termal na prática ortopédica. No entanto, há poucos estudos na literatura com forte evidência. Além disso, mesmo nas revisões sistemáticas e meta-análises, há críticas em relação à metodologia dos trabalhos que levam a riscos de vieses.

O tratamento que parece ser universal é o uso da crioterapia em condições agudas, principalmente após trauma direto com processo inflamatório e risco de hematoma, em especial, nas primeiras 72 horas, com ampla literatura respaldando esse uso.<sup>3-6</sup> Esse raciocínio é o mesmo para lesões musculares que ocorrem em atletas. Nas lesões musculares, mostrou-se que o uso precoce de crioterapia está associado a um hematoma significativamente menor no gap das fibras musculares rompidas, menor inflamação e regeneração acelerada.<sup>7,8</sup>

Em relação a patologias ortopédicas em geral, os estudos tendem a ter resultados divergentes e com baixa qualidade metodológica. As revisões sistemáticas sempre, ao discutirem os resultados, citam que há necessidade de ensaios clínicos randomizados (ECRs) com melhor metodologia quando se estuda a termoterapia e sua aplicabilidade dentro das mais variadas patologias (inflamatórias, degenerativas e pós-operatórias).<sup>9-12</sup> A falta de padronização no protocolo da termoterapia e dos tratamentos associados são outras dificuldades.

Nos ECRs, a falta de uma amostra significativa associada a uma metodologia inadequada levam a resultados divergentes com fatores confundidores com outras terapias e os próprios autores sinalizam que o tipo de compressa, fria ou quente, deve ser baseado na aceitação do paciente e na experiência clínica do profissional.<sup>13,14</sup>

Por fim, estudamos uma revisão sistemática que trata sobre o contraste de imersão quente e fria. Essa técnica visa realizar um efeito vascular que promoveria um aumento de fluxo sanguíneo superficial e intramuscular. Nessa revisão, Stanton et al.<sup>15</sup> colocam como evidência pobre de que ocorra uma alteração de fluxo sanguíneo com essa técnica dado resultados divergentes dos trabalhos, apesar de todos apresentarem um nível 2 de evidência, mas com uma qualidade metodológica fraca segundo os autores.

Na opinião do autor máster do capítulo, há uma tendência geral de melhora clínica com compressas frias ou quentes em diversas afecções

ortopédicas. Sugere-se que a prescrição seja pautada na melhora clínica sintomática do paciente sempre embasado pelo racional fisiológico do mecanismo. De maneira global, traumas agudos e controle do edema são teoricamente melhor tratadas com crioterapia. Por sua vez, efeitos de relaxamento muscular são obtidos com o calor. A literatura atual é bastante escassa e, mesmo quando presente, tem evidência científica fraca.

Quando não temos literatura forte que nos respalde, buscar um raciocínio mecanicista e sintomático parece a atitude mais assertiva. O uso de compressas não muda, muitas vezes, a história natural da doença ortopédica, mas visa um controle sintomático que é bem palpável na prática clínica, como controle de edema e melhora da dor. Logo, na ortopedia, por termos uma relação causa-efeito evidente com o tratamento, podemos basear nossa escolha na resposta clínica.

## Fortalecimento muscular

Há inúmeros estudos de forte impacto científico, como revisões sistemáticas, meta-análises e ECRs bem delineadas, que afirmam o benefício do fortalecimento em afecções ortopédicas. Esses estudos incluem articulações como a coluna, ombro, cotovelo, quadril, joelho, tornozelo e pé.<sup>16-23</sup>

Situações especiais como prevenção de lesões em corredores, arremessadores, pré e pós-operatórios e em grupos especiais, como amputados, também têm publicações com forte evidência de melhora.<sup>24-28</sup>

A importância do fortalecimento, em termos genéricos, é explicada por compartilhamento de carga e fatores biomecânicos.<sup>28</sup>

O controle algico por meio do fortalecimento muscular é explicado, muitas vezes, à luz de eventos biomecânicos. Na dor femoropatelar, por exemplo, o fortalecimento dos abdutores e rotadores externos do quadril diminuem o valgo dinâmico, um movimento que gera desalinhamento dinâmico na patela que ocasiona, em última análise, dor.<sup>29</sup> Outras situações biomecânicas também contribuem para a redução de dor, em patologias do quadril, ombro, entre outras.<sup>19, 24</sup>

A prevenção de lesões por meio de fortalecimento muscular também é bem descrita por estudos bem delineados, como no estudo de Petersen,<sup>30</sup> que analisou, em um ECR, 942 jogadores de futebol, sendo 461 atletas no grupo de intervenção e 481 atletas no grupo controle, avaliando desfechos sobre



prevenção de lesão muscular. O grupo da intervenção teve 15 lesões musculares *versus* 52 no grupo controle. Na discussão dos dados, o autor ainda afirma que 9 das lesões do primeiro grupo ocorreram em um momento inicial de treinamento (sem o completo efeito protetivo do fortalecimento), enquanto no grupo controle as lesões foram em todos os momentos do estudo. Prevenção de lesão em arremessadores,<sup>31</sup> em pós-operatórios<sup>32</sup> e prevenção de sobrecargas articulares,<sup>33</sup> entre outras condições são bastante presentes em publicações.

O aumento de performance em atividades esportivas também tem evidência forte. Trowell et al.,<sup>33</sup> em uma revisão sistemática e meta-análise que envolveu 25 estudos, todos controlados, com um total de 571 participantes que realizaram fortalecimento em média por 10,4 semanas. Eles analisaram o fortalecimento junto com atividades de resistência e concluíram que o treinamento de fortalecimento em corredores melhorou significativamente a mecânica da flexão do joelho, flexão plantar do tornozelo, extensão do joelho e agachamento, sugerindo uma maior capacidade de geração de força no tríceps sural, quadríceps, isquiotibiais e musculatura glútea. Esses músculos desempenham um papel crucial na propulsão do centro de massa e aceleração da perna durante a corrida gerando transferência de força para passos maiores e corrida com velocidades mais rápidas. Esse aumento de performance também foi debatido em diversos outros esportes.

Como vimos, a produção científica é vasta quando se trata de fortalecimento muscular em patologias ortopédicas. Há inúmeros estudos com nível de evidência forte, ficando a discussão do seu benefício para determinados tipos de exercícios em determinadas patologias. No entanto, no aspecto global, o fortalecimento muscular é uma condição que tem significativa associação com melhora da dor e do desempenho articular, bem como na prevenção de lesões.

## Kinesiotaping em ortopedia

É comum vermos o uso de *kinesiotaping* (KT) em afecções articulares, principalmente entre atletas profissionais e amadores. Essa modalidade de tratamento é realmente efetiva para esses atletas? Trouxemos, estrategicamente, as articulações que mais são causas de dor nos atletas: coluna, joelho e ombro para debate.

Em uma pesquisa que envolveu uma revisão sistemática de 5 estudos<sup>34</sup> envolvendo o uso do *KT* na síndrome de dor patelofemoral, os autores concluíram que o uso da *KT* isolada não tem benefício e que o seu uso poderia até ser associado ao fortalecimento tradicional. O *KT* foi utilizado com o intuito de medializar a patela e, quando presente, grupos-controle eram feitos com o *KT* sem tensão. Na análise dos trabalhos isoladamente, o *taping* isolado não obteve melhora significativa, mas o *KT* associado a fortalecimento foi mais efetivo que o fortalecimento isolado, no entanto, o *KT*-placebo (sem tensão) também foi mais efetivo, podendo ter efeito placebo apenas.

Em uma revisão sistemática com meta-análise, com avaliação de 11 ECRs em portadores de dor lombar,<sup>35</sup> o uso do *KT* não foi mais eficaz do que nenhuma intervenção ou placebo. Além disso, não foi eficaz quando usado como um adjuvante a Fisioterapia convencional no tratamento de pacientes com lombalgia crônica inespecífica.

Em relação às patologias do ombro, em uma revisão sistemática e meta-análise de ECRs, Ghozy et al.<sup>36</sup> avaliaram 12 estudos com 555 participantes. Os autores destacaram que a eficácia de *KT* foi investigada como um tratamento isolado, adjuvante para exercício, e em comparação com outras modalidades de tratamento usuais. Os resultados mostraram que o tape resultou em uma melhora significativa de dor e disfunção no ombro quando combinadas com exercícios. No entanto, assim como na dor femoropatelar, não produziu melhor resultados do que placebo ou quando comparado ao tratamento com corticosteroides. Os autores concluíram que o *KT* só pode ser usado com benefício como tratamento complementar.

Quando se analisa o uso da *KT* nas habilidades de desempenho esportivo em atletas, Reneker et al.,<sup>37</sup> em uma revisão sistemática, demonstrou que não há evidência de melhora no desempenho esportivo com o uso da *KT*. Diversos esportes foram analisados, como futebol, handebol, ciclismo, entre outros. Dentro de 193 comparações realizadas no estudo dentre os mais variados aspectos, apenas 2 foram em favor uso de *KT*.

Na busca de estudos com alto poder de evidência científica em lesões musculoesqueléticas em geral,<sup>38, 39</sup> os estudos parecem convergir para uma mesma tendência: *KT* só tem efeito benéfico se aplicadas com outras formas de tratamento, mas não conseguem superar os resultados do grupo placebo e não devem serem utilizados isoladamente.

Uma visão crítica que temos de ter é que é bastante difícil uma padronização de terapias por *KT*, pois há diversas técnicas em sua aplicação nas mais diversas patologias. Talvez, o método até possa ser efetivo, mas não se descobriu ainda o melhor método de aplicação dos tapes que proporciona melhora biomecânica efetiva para determinada articulação.

À luz da evidência científica atual, a prescrição de *KT* em lesões ortopédicas não tem embasamento científico suficiente na literatura, apesar de ser comum sua prescrição principalmente em lesões esportivas. Os estudos sobre o tema são mais escassos e com baixo poder estatístico de uma forma geral.

## Evidência no uso de ácido hialurônico

O ácido hialurônico (AH) é um polissacarídeo composto de repetidas unidades de dissacarídeos.<sup>40</sup> Possui diversas funções nas articulações normais, como efeitos condroprotetores, anti-inflamatórios, mecânicos, analgésicos, além de reduzir o impacto no osso subcondral e estimular a síntese de glicosaminoglicanos e proteoglicanos.<sup>41</sup> Em articulações degeneradas, pode ter concentração e tamanho diminuídos em até 50%.<sup>40</sup>

Articulações saudáveis possuem um líquido sinovial variando entre 6.500 e 10.900 kDa. Já nas artroses, o hialuronato sinovial é despolimerizado com massa molecular entre 2.700 e 4.500 kDa.<sup>42</sup> Com esse raciocínio em prática, o uso de ácido hialurônico exógeno intra-articular vem ganhando espaço em procedimentos de infiltração articular realizado principalmente por ortopedistas, reumatologistas e radiologistas. Afinal, há evidência científica desse uso? Como interpretar os trabalhos acerca desse tema?

Os trabalhos sobre o AH em patologias articulares são inúmeros e bem variados, porém a maior parte é sobre o controle da dor e dos escores funcionais principalmente em patologias intra-articulares, em especial, na artrose.

A principal articulação envolvida e com maior número de publicações, sem dúvidas, é o joelho. As articulações do quadril e do ombro também apresentam trabalhos com bom nível de evidência. Separamos aqui 3 trabalhos de forte impacto científico para discutir o benefício do AH nessas articulações.

No ombro, uma revisão sistemática e meta-análise de 15 estudos,<sup>43</sup> sendo 5 ensaios clínicos randomizados e 6 coortes prospectivas, evidenciou que a infiltração do AH é segura e que melhora dor nas artroses glenoumerais. O autor aborda risco de viés: 7 dos estudos foram com infiltração não guiada por ultrassonografia que pode estar relacionados a piores resultados pela menor acurácia no local da injeção da medicação. Fato esse que só aumenta o benefício do uso do AH que tenderia a melhorar os resultados ao ser injetado no local preciso.

No quadril, em um ensaio clínico randomizado<sup>44</sup> que comparou o uso de um grupo de pacientes com artrose de quadril que receberam AH de alto peso molecular em comparação com um grupo que recebeu AH de médio peso molecular. O autor afirma que uma única dose do primeiro grupo e duas doses do segundo grupo tinham similar redução de dor e disfunção. Essa conclusão mostra que mesmo AHs de menor peso molecular também são alternativas ao tratamento de artrose de quadril.

No joelho, em um ensaio clínico randomizado<sup>45</sup> que comparou infiltração intra-articular de AH em comparação com plasma rico em plaquetas e mostrou que ambos têm significativa melhora na dor e função em até 24 semanas com melhora do PRP em relação ao AH após 24 semanas.

Em uma análise global, baseado na literatura, há indícios que o AH tem bons resultados no controle da dor e melhora da função no tratamento da artrose. No entanto, há variáveis que temos que analisar e quando olhamos os detalhes metodológicos de cada trabalho, vemos aspectos que são dignos de nota.

O principal aspecto, abordado pelos 3 artigos citados, é a questão do placebo. No estudo do ombro, houve uma melhora significativa também no grupo que recebeu placebo. No quadril, uma crítica do autor cita como limitação de ausência do placebo e ainda cita que pode existir um efeito placebo importante na redução da dor, principalmente na primeira semana. No estudo do joelho, da mesma forma, o autor cita a falta de um grupo placebo para efeitos comparativos.

Fazendo uma busca mais abrangente, há trabalhos de forte impacto que mostram que o AH servem para diversas patologias.<sup>46-49</sup> Ao analisarmos esses estudos, sobre as variadas patologias, há uma uniformidade de melhora com uso de solução salina e/ou corticosteroides no curto prazo. Nos médio e longo prazos, o AH tende a ter melhores resultados. Esse

efeito inicial pode ser explicado pelo efeito mecânico em estruturas neurovasculares e efeito nas aderências entre o tendão e tecido peritendinoso conforme citado Boesen et al.<sup>50</sup> Logo, muito dos efeitos de curto prazo podem ser mais relacionados a hidrodissociação do que pelo AH propriamente dito.

Além disso, há outro aspecto importante que é o tipo de AH. A classe de alto peso molecular tende a ter resultados mais consistentes na literatura, como descrito na recente meta-análise de Phillips et al.,<sup>51</sup> que colocaram resultados mais consistentes dessa classe em relação aos AH de baixo peso molecular no tratamento da artrose do joelho, apesar de ambas tenderem para uma melhora clínica. No quadril, como citado previamente, os resultados foram similares.

Outro aspecto que deve ser avaliado é qual articulação está sendo estudada. Em articulações como tornozelo, pé e mão, após ampla pesquisa de literatura, poucos estudos com forte impacto científico foram vistos e a maioria dos trabalhos que citam outras articulações são compostas geralmente com resultados divergentes.<sup>52, 53</sup>

Em resumo, parece que o AH tem realmente tendência de bons resultados na literatura no controle de dor e melhora da função em patologias ortopédicas, principalmente intra-articulares, mais notadamente a articulação do joelho. No entanto, há diversos aspectos a serem avaliados e trabalhos devem ser analisados de maneira criteriosa, pois todos os estudos apresentam limitações e nem mesmo os estudos com maior poder de evidência são isentos de limitações.

## Conclusão

Dentro de um cenário rotineiro apresentado na introdução do capítulo, percebemos que o uso de compressa fria ou quente tem pouco número de publicações e com resultados divergentes na literatura; analisamos o fortalecimento muscular, que é amplamente aceito com inúmeros estudos com forte nível de evidência. Não encontramos evidências positivas para o uso de *KT* na literatura atual, apesar de uma boa quantidade de publicações; por fim, vimos que o uso de AH, apesar de uma tendência geral de bons resultados, deve ser individualizado, pois depende de inúmeros fatores, como articulação infiltrada, tipo de ácido hialurônico utilizado e a forma de administração. Obtivemos quatro cenários diferentes, o que só reforça a

importância da busca pela Medicina Baseada em Evidências para melhorarmos a assertividade em nossas prescrições como ortopedistas.

## Referencias bibliográficas

1. Bleakley CM, McDonough SM, MacAuley DC. Cryotherapy for acute ankle sprains: A randomised controlled study of two different icing protocols. *Br J Sports Med*. 2006;40(8):700-5.
2. Hohenauer E, Taeymans J, Baeyens JP, Clarys P, Clijsen R. The effect of post-exercise cryotherapy on recovery characteristics: A systematic review and meta-analysis. *PLoS One* [Internet]. 2015;10(9):1-22 [[acesso](#) em: 19 mar. 2021].
3. Airaksinen O V., Kyrklund N, Latvala K, Kouri JP, Grönblad M, Kolari P. Efficacy of cold gel for soft tissue injuries: A prospective randomized double-blinded trial. *Am J Sports Med*. 2003;31(5):680-4.
4. Bleakley C, McDonough S, MacAuley D. The Use of Ice in the Treatment of Acute Soft-Tissue Injury: A Systematic Review of Randomized Controlled Trials. *Am J Sports Med*. 2004;32(1):251-61.
5. Bleakley CM, Costello JT. Do thermal agents affect range of movement and mechanical properties in soft tissues? A systematic review. *Arch Phys Med Rehabil* [Internet]. 2013;94(1):149-63 [[acesso](#) em: 19 mar. 2021].
6. Park YH, Song JH, Kim TJ, Kang SH, Chang AS, Kim HJ. Comparison of the use of evaporative coolants and ice packs for the management of preoperative edema and pain in ankle fractures: a prospective randomized controlled trial. *Arch Orthop Trauma Surg* [Internet]. 2019;139(10):1399-405 [[acesso](#) em: 19 mar. 2021].
7. Hurme T, Rantanen J, Kaliomo H. Effects of early cryotherapy in experimental skeletal muscle injury. *Scand J Med Sci Sports*. 1993;3(1):46-51.
8. Fernandes TL, Pedrinelli A, Hernandez AJ. Lesão muscular - Fisiopatologia, diagnóstico, tratamento e apresentação clínica. *Rev Bras Ortop*. 2011;46(3):247-55.
9. Welch V, Brosseau L, Casimiro L, Judd M, Shea B, Tugwell P, et al. Thermotherapy for treating rheumatoid arthritis. *Cochrane Database Syst Rev*. 2002;(10).
10. French SD, Cameron M, Walker BF, Reggars JW, Esterman AJ. A cochrane review of superficial heat or cold for low back pain. *Spine (Phila Pa 1976)*. 2006;31(9):998-1006.
11. Brosseau L, Yonge K, Robinson V, Marchand S, Judd M, Wells G, et al. Thermotherapy for treatment of osteo-arthritis. *Physiotherapy*. 2003;89(12):694-5.
12. Adie S, Naylor J, Harris IA. Cryotherapy following total knee replacement. *Cochrane Database Syst Rev*. 2009;(3).
13. Garra G, Singer AJ, Leno R, Taira BR, Gupta N, Mathaikutty B, et al. Heat or cold packs for neck and back strain: A randomized controlled trial of efficacy. *Acad Emerg Med*.

2010;17(5):484-9.

14. Denegar CR, Dougherty DR, Friedman JE, Schimizzi ME, Clark JE, Comstock BA, et al. Preferences for heat, cold, or contrast in patients with knee osteoarthritis affect treatment response. *Clin Interv Aging*. 2010;5:199-206.
15. Breger Stanton DE, Lazaro R, MacDermid JC. A Systematic Review of the Effectiveness of Contrast Baths. *J Hand Ther [Internet]*. 2009;22(1):57-70 [[acesso em: 19 mar. 2021](#)]. .
16. Verbrugghe J, Agten A, Stevens S, Hansen D, Demoulin C, O Eijnde B, et al. Exercise Intensity Matters in Chronic Nonspecific Low Back Pain Rehabilitation. *Med Sci Sports Exerc*. 2019;51(12):2434-42.
17. Warby SA, Ford JJ, Hahne AJ, Watson L, Balster S, Lenssen R, et al. Comparison of 2 Exercise Rehabilitation Programs for Multidirectional Instability of the Glenohumeral Joint: A Randomized Controlled Trial. *Am J Sports Med*. 2018;46(1):87-97.
18. Raman J, MacDermid JC GR (Univ. Effectiveness of Different Methods of Resistance Exercises in Lateral Epicondylitis—A Systematic Review. In: *Yearbook of Hand and Upper Limb Surgery [Internet]*. Mosby, Inc.; 2012. p. 125-6 [[acesso em: 19 mar. 2021](#)].
19. Hoit G, Whelan DB, Dwyer T, Ajrawat P, Chahal J. Physiotherapy as an Initial Treatment Option for Femoroacetabular Impingement: A Systematic Review of the Literature and Meta-analysis of 5 Randomized Controlled Trials. *Am J Sports Med*. 2020;48(8):2042-50.
20. Nascimento L, Teixeira-Salmela LF, Souza RB, Resende RA. Hip and knee strengthening is more effective than knee strengthening alone for reducing pain and improving activity in individuals with patellofemoral pain: a systematic review with meta-analysis. *J Orthop Sport Phys Ther*. 2017;48(1):19-31.
21. Hölmich P, Nyvold P, Larsen K. Continued significant effect of physical training as treatment for overuse injury: 8- to 12-year outcome of a randomized clinical trial. *Am J Sports Med*. 2011;39(11):2447-51.
22. Silbernagel KG, Thomeé R, Eriksson BI, Karlsson J. Continued sports activity, using a pain-monitoring model, during rehabilitation in patients with achilles tendinopathy: A randomized controlled study. *Am J Sports Med*. 2007;35(6):897-906.
23. Rathleff MS, Mølgaard CM, Fredberg U, Kaalund S, Andersen KB, Jensen TT, et al. High-load strength training improves outcome in patients with plantar fasciitis: A randomized controlled trial with 12-month follow-up. *Scand J Med Sci Sport*. 2015;25(3):e292-300.
24. Taddei UT, Matias AB, Duarte M, Sacco ICN. Foot Core Training to Prevent Running-Related Injuries: A Survival Analysis of a Single-Blind, Randomized Controlled Trial. *Am J Sports Med*. 2020;48(14):3610-9.
25. Sakata J, Nakamura E, Suzuki T, Suzukawa M, Akaike A, Shimizu K, et al. Efficacy of a Prevention Program for Medial Elbow Injuries in Youth Baseball Players. *Am J Sports Med*. 2018;46(2):460-9.

26. Shaarani SR, O'Hare C, Quinn A, Moyna N, Moran R, O'Byrne JM. Effect of prehabilitation on the outcome of anterior cruciate ligament reconstruction. *Am J Sports Med.* 2013;41(9):2117-27.
27. Harput G, Ulusoy B, Yildiz TI, Demirci S, Eraslan L, Turhan E, et al. Cross-education improves quadriceps strength recovery after ACL reconstruction: a randomized controlled trial. *Knee Surgery, Sport Traumatol Arthrosc [Internet].* 2019;27(1):68-75 [[acesso](#) em: 19 mar. 2021].
28. Wong CK, Ehrlich JE, Ersing JC, Maroldi NJ, Stevenson CE, Varca MJ. Exercise programs to improve gait performance in people with lower limb amputation: A systematic review. *Prosthet Orthot Int.* 2016;40(1):8-17.
29. Khayambashi K, Mohammadkhani Z, Ghaznavi K, Lyle MA, Powers CM. The effects of isolated hip abductor and external rotator muscle strengthening on pain, health status, and hip strength in females with patellofemoral pain: A randomized controlled trial. *J Orthop Sports Phys Ther.* 2012;42(1):22-9.
30. Petersen J, Thorborg K, Nielsen MB, Budtz-Jørgensen E, Hölmich P. Preventive effect of eccentric training on acute hamstring injuries in Men's soccer: A cluster-randomized controlled trial. *Am J Sports Med.* 2011;39(11):2296-303.
31. Sakata J, Nakamura E, Suzuki T, Suzukawa M, Akeda M, Yamazaki T, et al. Throwing Injuries in Youth Baseball Players: Can a Prevention Program Help? A Randomized Controlled Trial. *Am J Sports Med.* 2019;47(11):2709-16.
32. Coppack RJ, Etherington J, Wills AK. The effects of exercise for the prevention of overuse anterior knee pain: A randomized controlled trial. *Am J Sports Med.* 2011;39(5):940-8.
33. Trowell D, Vicenzino B, Saunders N, Fox A, Bonacci J. Effect of Strength Training on Biomechanical and Neuromuscular Variables in Distance Runners: A Systematic Review and Meta-Analysis. *Sport Med [Internet].* 2020;50(1):133-50 [[acesso](#) em: 19 mar. 2021].
34. Logan CA, Bhashyam AR, Tisosky AJ, Haber DB, Jorgensen A, Roy A, et al. Systematic Review of the Effect of Taping Techniques on Patellofemoral Pain Syndrome. *Sports Health.* 2017;9(5):456-61.
35. Sheng Y, Duan Z, Qu Q, Chen W, Yu B. Kinesio taping in treatment of chronic non-specific low back pain: A systematic review and meta-analysis. *J Rehabil Med.* 2019;51(10):734-40.
36. Sherief Ghozy, Nguyen Minh Dung, Mostafa Ebraheem Morra, Sara Morsy, Ghadeer Gamal Elsayed, Linh Tran, Le Huu Nhat Minh, Alzhraa Salah Abbas, Tran Thai Huu Loc, Truong Hong Hieu, Truong Cong Dung NTH. Efficacy of kinesio taping in treatment of shoulder pain and disability: a systematic review and meta-analysis of randomised controlled trials Sherief. *Physiotherapy [Internet].* 2020;107:176-88 [[acesso](#) em: 19 mar. 2021].
37. Reneker JC, Latham L, McGlawn R, Reneker MR. Effectiveness of kinesiology tape on sports performance abilities in athletes: A systematic review. *Phys Ther Sport.* 2018;31:83-



98.

38. Montalvo AM, Cara E Le, Myer GD. Effect of kinesiology taping on pain in individuals with musculoskeletal injuries: Systematic review and meta-analysis. *Phys Sportsmed*. 2014;42(2):48-57.
39. Mostafavifar M, Wertz J, Borchers J. A systematic review of the effectiveness of kinesio taping for musculoskeletal injury. *Phys Sportsmed*. 2012;40(4):33-40.
40. Clegg TE, Caborn D, Mauffrey C. Viscosupplementation with hyaluronic acid in the treatment for cartilage lesions: A review of current evidence and future directions. *Eur J Orthop Surg Traumatol*. 2013;23(2):119-24.
41. Altman RD, Manjoo A, Fierlinger A, Niazi F, Nicholls M. The mechanism of action for hyaluronic acid treatment in the osteoarthritic knee: A systematic review. *BMC Musculoskelet Disord* [Internet]. 2015;16(1):1-10 [[acesso em: 19 mar. 2021](#)].
42. Rutjes AWS, Ju P, Costa BR, Trelle S, Nu E. Annals of Internal Medicine Viscosupplementation for Osteoarthritis of the Knee. *Ann Intern Med*. 2012;157:180-91.
43. Zhang B, Thayaparan A, Horner N, Bedi A, Alolabi B, Khan M. Outcomes of hyaluronic acid injections for glenohumeral osteoarthritis: a systematic review and meta-analysis. *J Shoulder Elb Surg*. 2019;28(3):596-606.
44. Clementi D, D'Ambrosi R, Bertocco P, Bucci MS, Cardile C, Ragni P, et al. Efficacy of a single intra-articular injection of ultra-high molecular weight hyaluronic acid for hip osteoarthritis: a randomized controlled study. *Eur J Orthop Surg Traumatol*. 2018;28(5):915-22.
45. Di Martino A, Di Matteo B, Papio T, Tentoni F, Selleri F, Cenacchi A, et al. Platelet-Rich Plasma Versus Hyaluronic Acid Injections for the Treatment of Knee Osteoarthritis: Results at 5 Years of a Double-Blind, Randomized Controlled Trial. *Am J Sports Med*. 2019;47(2):347-54.
46. Tosun HB, Gumustas S, Agir I, Uludag A, Serbest S, Pepele D, et al. Comparison of the effects of sodium hyaluronate-chondroitin sulphate and corticosteroid in the treatment of lateral epicondylitis: a prospective randomized trial. *J Orthop Sci*. 2015;20(5):837-43.
47. Kanchanathepsak T, Pichyangkul P, Suppaphol S, Watcharananan I, Tuntiyatorn P, Tawonsawatruk T. Efficacy Comparison of Hyaluronic Acid and Corticosteroid Injection in Treatment of Trigger Digits: A Randomized Controlled Trial. *J hand Surg Asian-Pacific Vol*. 2020;25(1):76-81.
48. Liu DH, Tsai MW, Lin SH, Chou CL, Chiu JW, Chiang CC, et al. Ultrasound-Guided Hyaluronic Acid Injections for Trigger Finger: A Double-Blinded, Randomized Controlled Trial. *Arch Phys Med Rehabil* [Internet]. 2015;96(12):2120-7 [[acesso em: 19 mar. 2021](#)].
49. Kumai T, Samoto N, Hasegawa A, Noguchi H, Shiranita A, Shiraishi M, et al. Short-term efficacy and safety of hyaluronic acid injection for plantar fasciopathy. *Knee Surgery, Sport Traumatol Arthrosc*. 2018;26(3):903-11.

50. Boesen AP, Hansen R, Boesen MI, Malliaras P, Langberg H. Effect of High-Volume Injection, Platelet-Rich Plasma, and Sham Treatment in Chronic Midportion Achilles Tendinopathy: A Randomized Double-Blinded Prospective Study. *Am J Sports Med.* 2017;45(9):2034-43.
51. Phillips M, Vannabouathong C, Devji T, Patel R, Gomes Z, Patel A, et al. Differentiating factors of intra-articular injectables have a meaningful impact on knee osteoarthritis outcomes: a network meta-analysis. *Knee Surgery, Sport Traumatol Arthrosc* [Internet]. 2020;28(9):3031-9 [[acesso em: 19 mar. 2021](#)].
52. Vannabouathong C, Del Fabbro G, Sales B, Smith C, Li CS, Yardley D, et al. Intra-articular Injections in the Treatment of Symptoms from Ankle Arthritis: A Systematic Review. *Foot Ankle Int.* 2018;39(10):1141-50.
53. Jantzen C, Ebskov LB, Andersen KH, Benyahia M, Rasmussen PB, Johansen JK. The Effect of a Single Hyaluronic Acid Injection in Ankle Arthritis: A Prospective Cohort Study. *J Foot Ankle Surg.* 2020;59(5):961-3.

# 16. JORNALISMO DE SAÚDE BASEADO EM EVIDÊNCIAS

## Introdução

“Jornalistas deveriam desconfiar da ciência”. A frase soa quase herética num tempo em que o negacionismo científico traz ameaças que afetam todos nós, incluindo a desconfiança de algumas pessoas nas vacinas e a dúvida sobre o papel do homem nas mudanças climáticas, por exemplo. Além do mais, se tem alguém que merece ser olhado com reserva, hoje, é a própria mídia – “sempre tão cheia de sensacionalismo e distorções!”, muitos apontam. Mas antes que o profissional de saúde, o jornalista, e qualquer membro do público que defenda a ciência ignore este capítulo, vamos detalhar para qual ciência o ceticismo deve ser apontado.

Por mais que usemos o termo ciência como entidade única, com fins de simplificação ou de retórica, não existe uma só ciência com C maiúsculo sendo praticada por aí. E não estou me referindo à miríade de áreas diferentes e métodos específicos, mas sim à qualidade, isto é, a capacidade dos diferentes estudos de gerar evidências fortes. Não saber diferenciar essa capacidade e ter o mesmo olhar para qualquer estudo – sem desconfiar da qualidade da ciência que vai usar como fonte de suas notícias – é talvez o primeiro erro na prática do jornalismo de saúde.

O jornalista não é cientista e por isso não é obrigado a saber? Ora, a minoria dos profissionais de saúde é também cientista, e nem por isso está desobrigada a saber avaliar a força da ciência em que se baseia para tomar suas decisões no dia a dia. Isso porque decisões de médicos, enfermeiros, nutricionistas etc. afetam a saúde de seus pacientes, e obviamente estes devem se pautar pelo que há de melhor na área de conhecimento em que atuam. Se o jornalismo afeta decisões das pessoas que também geram consequências na saúde delas, nós, jornalistas, também deveríamos nos pautar nas melhores evidências ao produzir notícias sobre saúde. Muitas vezes não é nem questão de ser ciência de má qualidade – pode ser ciência boa, porém em etapas ou abordagens que não são as melhores a serem apresentadas, ainda, ao público. Ou que, se forem, devem ser ainda mais contextualizadas em seus resultados.

Como isso pode ser feito na prática, ou seja, como identificar boas evidências científicas para apresentar ao seu público é o que procurei fazer neste capítulo, que pode ser útil tanto a jornalistas quanto a profissionais diversos que desejem ler com mais senso crítico o noticiário científico.

## Ciência cumulativa

O primeiro ponto a ter em mente é que em ciência não se tiram conclusões fechadas a partir de um único estudo – e nem um estudo mais recente tem necessariamente as melhores respostas para uma pergunta.<sup>1</sup> Ciências, as da saúde, inclusive, são construídas com base em acúmulo de evidências. E quanto mais um estudo contrariar as evidências que se têm até o momento sobre um assunto, mais cuidado se exige ao olhar para suas conclusões. Assim, se um jornalista se deparar com uma pesquisa que diga que o tabagismo não é um grande fator de risco para o câncer de pulmão, fato já largamente estudado e para o qual já há robustas evidências, é ainda mais importante verificar o que outros pesquisadores no tema têm a dizer a respeito antes de correr para escrever um texto com uma manchete chamativa, acreditando que conseguiu uma ótima pauta. O mesmo vale para temas que ainda estão sendo explorados e, por isso, podem parecer mais controversos – alguns controversos ao ponto de, numa mesma semana, serem publicados dois estudos com conclusões opostas sobre um tópico, por exemplo. A nutrição é uma área muito propícia a esse tipo de risco, e um dos motivos é a dificuldade de se acompanhar objetivamente hábitos alimentares (normalmente autorrelatados em pesquisas com humanos) e isolá-los de outros fatores (outros hábitos e diferenças individuais) ao tentar responder se o consumo de algum produto interfere – e como interfere – em aspectos da nossa saúde. Daí o estranhamento de alimentos como ovo serem retratados ora como vilões, ora como mocinhos. Ou outros como café, chocolate, vinho (que a maioria de nós adoramos, e sobre os quais as reportagens sempre “vendem bem”) terem estudos a respeito sempre nas manchetes, como se tivessem uma capacidade mágica de fazer bem a aspectos tão diversos como a saúde cardiovascular e o equilíbrio dos neurotransmissores.

Para outros temas que estão sempre em pauta, como o câncer, o diabetes, a hipertensão e o AVC, entre outros, muitas entidades como a Organização Mundial de Saúde (OMS), assim como sociedades médicas,

dispõem de guias sintetizando as evidências mais fortes acumuladas. A associação entre hábitos e doenças como o câncer é um exemplo. Guias com as melhores evidências até o momento sobre vários tópicos, inclusive a Covid-19, sempre são tornados públicos pelo Serviço Nacional de Saúde britânico (National Health Service – NHS), sendo constantemente atualizados. Mas mesmo que não exista material pronto acerca do tema sobre o qual um jornalista está noticiando, ele pode acessar estudos de revisão – que buscam sumarizar e discutir os resultados dos estudos mais importantes feitos até aquele momento. E, talvez o mais importante, consultar bons especialistas da área para pedir uma contextualização do chamado estado da arte naquele assunto: o que as últimas e mais robustas evidências sobre ele indicam.

## Onde o estudo foi publicado importa

Embora também possa acontecer de periódicos científicos de alta reputação na comunidade acadêmica trazerem estudos que depois se mostram equivocados, inclusive sendo retirados, estar publicada numa revista de alta qualificação fala a favor de uma pesquisa, assim como a publicação em algum periódico obscuro ou de baixa relevância fala contra. Para os periódicos brasileiros ligados aos programas de pós-graduação, temos o sistema Qualis,<sup>2</sup> que tem um sistema de classificação quanto à qualidade, por área de pesquisa. Para periódicos nacionais e também internacionais, pode-se observar o chamado fator de impacto, uma métrica que contabiliza as citações recebidas nos últimos 2 ou 3 anos, publicada anualmente. Na comunidade científica, ser amplamente citado é um dos indicativos da qualidade de um artigo; se um periódico costuma ter seus artigos amplamente citados, esse é um indício – mas não o único – da sua qualidade. A ferramenta para saber o fator de impacto de uma revista é o site InCites Journal Citations Report.<sup>3</sup> É recomendável utilizá-lo comparativamente (e não o fator isolado de cada revista), usando a busca por categoria para verificar dentro de cada área as revistas com maior fator.

Existe uma ameaça bem mais grave que periódicos de pouco impacto em se tratando de publicações científicas: os chamados periódicos predatórios. Tais revistas estão interessadas apenas em receber a taxa paga pelos autores, e não têm nenhum critério sobre a qualidade dos artigos submetidos. Na carreira acadêmica, pesquisadores têm constantemente sua

produtividade de pesquisa avaliada, seja ao se inscrever num concurso para docência, seja ao pleitear uma bolsa para financiar seus projetos. Esses periódicos exploram essa pressão, que muitas vezes é voltada equivocadamente à quantidade de artigos publicados – e não à qualidade. Disso resultam artigos sem revisão adequada e fraquíssimos em nível de evidência – e até mesmo artigos falsos (quando nenhuma pesquisa foi feita para produzi-los). Para comprovar a existência desse perigoso mercado, pesquisadores e jornalistas já fizeram algumas vezes o teste, submetendo textos com erros graves,<sup>4</sup> e inclusive textos sem sentido, gerados aleatoriamente por programas de computador, que foram prontamente aceitos por essas revistas.

Existem listas na internet que indicam periódicos potencialmente predatórios, que o jornalista pode acessar, a exemplo da Preda Qualis,<sup>5</sup> que traz publicações com essa característica ligadas aos programas de pós-graduação brasileiros. Uma lista internacional gratuita é a *Beall's list*,<sup>6</sup> que infelizmente parou de ser atualizada em 2017.

## Conflitos de interesse

Em geral, os autores costumam ser obrigados a informar se os resultados da pesquisa podem favorecer interesses de alguém – seja dos próprios autores do estudo, seja dos seus financiadores. Por exemplo: uma indústria farmacêutica que patrocina um estudo sobre o uso de uma medicação existente e que ela já comercializa, para um novo fim.

Conflito de interesse não quer dizer que alguém agiu de maneira corrupta nem que o estudo necessariamente foi enviesado por isso. Mesmo assim, quem tem acesso aos resultados daquela pesquisa tem o direito de saber que o conflito existe: que há, por trás do artigo, relações que aumentam as chances de o autor de um estudo não ser guiado somente pelo interesse científico, ainda que de maneira inconsciente, porque os resultados trazem implicações pessoais.

Como jornalistas, devemos também prestar atenção nesse ponto ao decidir se um estudo merece ou não ser noticiado. E se optarmos por noticiá-lo, mesmo sabendo que há alguma situação que suscite o conflito de interesse, é preciso que isso conste na matéria.

Contarei um exemplo pessoal. Em certa ocasião, noticiei numa rede social um estudo feito por pesquisadores de uma universidade conceituada

em um periódico insuspeito a respeito do efeito da terapia em câmara hiperbárica sobre os telômeros (parte dos cromossomos cujo tamanho reduzido é associado ao envelhecimento). A repercussão foi imediata e enorme – afinal, quem não ficaria interessado em algum método que talvez pudesse oferecer rejuvenescimento celular? Outros veículos<sup>7</sup> também deram a pesquisa. Eu, como outros, cometi um erro, que foi não me atentar para o item “conflito de interesses” que constava ao final do artigo: o vínculo de alguns autores do estudo com uma rede de clínicas que oferecia essa terapia (mesmo que com outros fins). E o pior, alertada por um pesquisador que foi mais a fundo no levantamento, fiquei sabendo que outros cientistas que assinaram o artigo não informaram conflitos de interesse existentes (eles também prestavam serviços para clínicas relacionadas, mas não expuseram isso no artigo). Em poucas horas, a postagem “viralizou”, e já havia várias pessoas me mandando mensagem perguntando se eu conhecia alguma clínica que oferecesse o tratamento no Brasil. Preferi apagar a postagem e informar o público por que o fiz. Infelizmente, outras notícias sobre o método saíram em veículos de grande alcance nacionais e internacionais, e jamais foram retiradas, ou mesmo informaram aos leitores a existência daqueles conflitos de interesses.

E já que tocamos na questão dos erros, é bom dizer que, por mais que os evitemos incessantemente, acreditemos que aplicamos rígidos critérios e tenhamos “as melhores das intenções”, uma hora ou outra eles acontecerão. Ora, se até cientistas e as publicações científicas, que são feitas de maneira bem mais metódica e sob maior revisão que as reportagens jornalísticas, erram, não devemos esperar que isso não aconteça no jornalismo, mas sim ter um plano de ação já em mente para quando ocorrer. Nesse quesito, a transparência é irmã da credibilidade. As ações de assumir o erro, informar ao público exatamente o que aconteceu, e corrigir o erro em questão devem sempre estar incluídas nesse plano. Mais do que o jornalista individualmente, os veículos noticiosos devem ter uma política firme e clara a respeito, até porque os prejuízos para o próprio veículo, para o cientista envolvido na matéria com erro e, mais importante, para o público podem ser grandes.

## Tipos de estudo

Existem muitos modos de estudar um mesmo tema ou tentar responder a uma pergunta de pesquisa parecida. Cada tipo de estudo tem vantagens, desvantagens, oferece mais ou menos dificuldade para quem pesquisa e, no final, gera conclusões com mais ou menos peso. Por exemplo, para o teste de terapêuticas, isto é, tratamentos (farmacológicos ou não), o chamado padrão ouro é o ensaio clínico randomizado (RCT, na sigla em inglês). Embora não sem valor, constituem evidências mais fracas os estudos observacionais, ou seja, em que a adoção do tratamento não foi feita no contexto sistematizado de um experimento, com variáveis controladas, mas sim observada em retrospectiva. Já os estudos de revisão sistemática reúnem evidências de outros estudos sobre um tema, buscando extrair as conclusões mais fortes. Podem ser úteis, mas também devem ser avaliados com cuidado. Se os estudos escolhidos para entrar na revisão são fracos, por exemplo, os resultados da revisão também o serão. Para ajudar a aumentar a força das evidências de uma revisão é possível adotar a meta-análise, utilizando métodos estatísticos para sumarizar resultados de pesquisas independentes. E mesmo a meta-análise pode ser feita de diferentes maneiras, produzindo resultados mais ou menos confiáveis.

Na prática, percebemos que a escolha de estudos a serem divulgados pela mídia não se atém a essas diferenças. Uma pesquisa<sup>8</sup> que analisou artigos originais cobertos por jornais de ampla circulação nos Estados Unidos verificou que esses jornais cobriam mais estudos observacionais e menos RCTs em comparação ao que era publicado em periódicos científicos considerados de alto impacto. Ainda, mostrou que os artigos sobre estudos observacionais cobertos pela mídia eram de qualidade inferior. Ou seja, os jornais davam maior atenção a pesquisas médicas com metodologia mais fraca.

Não é intenção deste texto descrever e comparar todos os tipos de estudo e níveis de evidência, mas é interessante que o jornalista conheça pelo menos o básico do assunto para evitar armadilhas na hora de apresentar as conclusões de uma pesquisa ao público (capítulo 3).

## In vitro e in vivo

Na avaliação de novas drogas e tratamentos, precedem os ensaios clínicos com pacientes pelo menos duas etapas de estudos. Primeiramente, são feitos estudos *in vitro*, quando, em geral, as drogas são testadas em



culturas de células dentro de um laboratório. Estudos *in vitro* são uma fase indispensável e têm enorme valor na ciência. Para serem noticiados ao público geral, nem sempre. Há uma imensa quantidade de estudos *in vitro* que indicam bons resultados para uma nova substância, mas não chegam nem a ir para o estágio final, que é o ensaio clínico. Mesmo dentre os estudos *in vitro* há variações, mas geralmente são evidências ainda fracas, ou pelo menos insuficientes para adoção do novo tratamento na prática clínica. Após estudos *in vitro*, um composto pode ser testado em modelos animais, ou *in vivo*, como se diz no jargão científico.

Novamente, pesquisas com animais são de grande relevância para a ciência, principalmente quando já existe um modelo animal adequado para aquela condição, por exemplo, um camundongo criado em laboratório para ter diabetes tipo 1, e que já se sabe que dá respostas mais ou menos parecidas com a fisiologia humana. Ainda assim, mesmo estudos bem-feitos usando modelos animais não fornecem as melhores evidências para transposição de seus resultados a pacientes humanos. Pode-se noticiar os resultados apresentando seu bom potencial etc. – mas nada além disso. Há matérias que nem sequer indicam ao público que estão tratando de um estudo realizado com animais, ou se indicam, não deixam claro que bons resultados em estudos com animais não são garantia de que um tratamento funcionará bem com humanos.

Além disso, praticamente inexistente o hábito na imprensa de acompanhar posteriormente as pesquisas sobre aquele tratamento noticiado para saber suas conclusões em etapas posteriores e, muito menos, o de informar ao público quando resultados que foram noticiados não se mostraram duradouros ou mesmo foram desmentidos por novos estudos.

Um estudo<sup>9</sup> que examinou o noticiário científico e verificou se as moléculas com atividade biológica apresentadas por reportagens como promissoras se tornaram medicamentos revelou que apenas duas (5%) de 40 moléculas identificadas no noticiário no período estudado cumpriram com sucesso todas as fases de teste e desenvolvimento, apesar de jornalistas e mesmo cientistas que falavam nas reportagens terem dado como certo que as moléculas se tornariam medicamentos em alguns anos.

Uma história sobre um medicamento promissor que não foi para a frente pode não brilhar os olhos de um repórter. Mas falar do assunto é importante, ainda que o diluindo dentro de outras histórias para mostrar como funciona a ciência.

## Estatística

Não conheço curso de Jornalismo que dê formação em estatística básica, muito menos em estatística para pesquisa científica. Com raríssimas exceções, não temos nem noções de como ela funciona. Assim, é normal termos dificuldades em ler um estudo e saber, numa olhada rápida, se os resultados que ele traz são significativos. Nesse aspecto, uma solução é contar com a ajuda de fontes da sua confiança que possam dar algumas dicas e indicar logo de cara que os números daquele artigo podem não dizer muita coisa, e que a pesquisa não merece ser noticiada. Sempre vamos precisar desse tipo de cooperação, ainda que procuremos aprender o básico (capítulo 6) – o que vale para uma especialidade e um tipo de estudo, em geral, não se aplica integralmente a outra.

Por exemplo, o tamanho da amostra em um ensaio clínico, chamado de  $n$ , se refere ao número de pacientes que receberam a intervenção de que se deseja avaliar a eficiência. Uma amostra muito pequena pode não ser suficiente para testar uma hipótese, trazendo resultados aleatórios, isto é, que podem ter se dado por mero acaso, sem relação com a intervenção que foi realizada. Isso vale tanto para resultados positivos quanto negativos. Mas o critério do que é ou não uma amostra pequena pode variar para cada tipo de estudo. Estudos pré-clínicos, por exemplo, para testar a segurança de uma intervenção, antes de verificar sua eficácia, podem ter amostras menores. Lembrando que, com raras exceções, estudos pré-clínicos nem mesmo constituem boas pautas para serem noticiadas ao público geral – apesar de serem fundamentais para a ciência.

Outro exemplo: o intervalo de confiança é um conceito relacionado ao fato do resultado de uma pesquisa nunca nos dar uma verdade absoluta, mas uma estimativa. Por exemplo, ao chegar ao resultado de que após a intervenção (o tratamento com um medicamento, por exemplo), os pacientes de um estudo apresentaram melhora de 27% a 30% em relação ao parâmetro pesquisado (digamos, a redução da glicemia, a taxa de açúcar no sangue), não podemos dizer que em toda a população que for submetida àquele tratamento haverá redução da glicemia de 27% a 30% – por mais bem desenhado que tenha sido o estudo. Podemos apenas estimar isso. Há uma fórmula que nos diz o quanto podemos confiar neste resultado, buscando medir a incerteza. E qual seria um índice tolerável para essa incerteza? Cinco por cento? Depende. Não existe um número mágico.

Depende do tipo de estudo, do que ele busca medir, das diferenças entre os critérios clínicos de uma dada especialidade. Em última análise, mesmo que um jornalista de ciências seja experiente em ler e interpretar artigos científicos, ajuda especializada é quase sempre necessária.

## Preprints

A divulgação de *preprints* é a última controvérsia da ciência na pandemia. Em tempos corriqueiros, a orientação seria não utilizar como fonte de pauta resultados divulgados em repositórios de artigos que ainda não passaram pela revisão dos pares. Durante a pandemia do novo coronavírus, porém, muitos cientistas preferiram divulgar os resultados dos seus estudos por esse meio, com a justificativa de não atrasar, com o processo de submissão a periódicos, o compartilhamento de conhecimentos em um contexto que pede urgência. Para os jornalistas em busca dos últimos desdobramentos na pesquisa por qualquer coisa que possa ajudar nessa crise de saúde, parece irresistível noticiar esses estudos, que muitas vezes são divulgados nas redes sociais pelos próprios autores.

No *Jornal da USP*, inicialmente ficamos inseguros sobre noticiar estudos ainda sem revisão. Mas percebemos que não era uma escolha nossa. Vários veículos da mídia externa encontravam os estudos em *preprint* dos pesquisadores da USP e os noticiavam. Então concluímos que era melhor tentar dar um tratamento mais adequado a esse tipo de pauta do que simplesmente ignorá-lo. Um dos principais cuidados necessários é explicar logo no início ao leitor o que é *preprint* e relativizar suas conclusões mais como um indicativo de potencial do que como um resultado. E explicar isso de uma maneira muito, muito simples mesmo. Não apenas dizer que “Ainda não foi revisado por pares”, quando a maior parte do público sequer entenderá que “pares” são esses.

Outro imperativo para noticiar um estudo em *preprint* é encontrar pesquisadores da mesma área do conhecimento e, de preferência de outra instituição, para comentarem o artigo e ajudar a dimensionar aquelas conclusões – coisas que, sozinho, um jornalista não consegue fazer. Se ainda não houve revisão da comunidade científica dentro dos ritos acadêmicos habituais, caberia a quem noticia ao menos passar o artigo por uma espécie de revisão extraoficial de um ou mais especialistas daquela área.

## Incerteza e linguagem

É um desafio necessário de enfrentar o tratar com naturalidade a incerteza que permeia a ciência sem que isso gere desconfiança sobre a ciência como um todo. Este precisa ser um trabalho contínuo nas reportagens: apresentar a ciência como um processo, e não como descobertas pontuais independentes. Muito da incompreensão tanto de jornalistas quanto do público que atrapalha entender o contexto de um resultado de pesquisa vem da falta de conhecimentos básicos sobre o funcionamento da ciência. Um jornalista não precisa ser especialista em método científico. Mas espera-se que um repórter que cobre um tema de ciência tenha bem claros conceitos como ensaio clínico, revisão por pares e replicação de experimentos, para citar alguns. E, principalmente, saber utilizar a linguagem para expressar essa incerteza com naturalidade. Um exemplo fictício: ao invés de passar uma certeza impossível de ser constatada na prática usando a manchete “Cafeína melhora desempenho cognitivo de idosos”, teríamos as seguintes opções:

- Cafeína tem potencial para melhorar desempenho cognitivo de idosos.
- Em experimento, cafeína melhorou desempenho de idosos em testes cognitivos.
- Em testes com animais idosos, cafeína melhorou desempenho em tarefas cognitivas (quando for o caso de testes com animais, evidentemente).
- Estudo traz evidências de que cafeína pode melhorar desempenho cognitivo de idosos.
- Cafeína e desempenho cognitivo (essa é para quem tem poucos caracteres disponíveis, uma reclamação constante dos jornalistas ao se defenderem sobre manchetes mal elaboradas).

- Cafeína pode melhorar desempenho cognitivo? (manchete em forma de pergunta – não é o melhor dos mundos para jornalistas, que gostam de trazer conclusões no título, mas funciona quando se tem pouco espaço).

## Sensacionalismo

O sensacionalismo na prática do jornalismo de ciência gera descrédito, primeiro, para a própria imprensa. Se dizemos hoje que “o ovo faz mal à saúde” (capítulo 5) sem contextualizar bem a pesquisa que mostrou esses resultados, falar das suas reais implicações e limitações, amanhã, quando surgir uma pesquisa falando que o consumo do ovo com determinada frequência é favorável para alguns aspectos da saúde, pouca gente levará a sério o que escrevermos sobre o tema nas páginas dos jornais. Em última análise, o sensacionalismo na mídia prejudica a confiança do público na própria ciência. E para evitar esse problema, a colaboração dos cientistas também é essencial. Na ânsia de serem citados, muitas vezes os próprios autores das pesquisas exageram os resultados de seus estudos ao comunicá-los para a mídia. Quando não eles, as instituições a que pertencem o fazem, ao enviar seus *press releases*.<sup>10</sup>

O sensacionalismo é uma armadilha porque traz ganhos imediatos, seja em cliques e *page views*, seja em repercussão pessoal para um pesquisador ou uma instituição. Mas sabemos que esses ganhos não são duradouros, e podem se converter em perdas maiores. A busca por um jornalismo de ciência e saúde baseado em evidências, portanto, deve ser uma tarefa que une cientistas e jornalistas em cooperação, com frutos em médio e longo prazo que beneficiarão ambas as partes. Mais do que isso, contribuirá para termos uma sociedade bem informada e mais saudável, em todas as acepções do termo.

## Referências bibliográficas

1. Belluz J. Health journalism has a serious evidence problem. Here’s a plan to save it. [Internet]. VOX. 2016 [[acesso](#) em: 19 mar. 2021].
2. Capes. Qualis Periódicos [Internet] [[acesso](#) em: 19 mar. 2021].
3. ANALYTICS C. InCites Journal Citations Report. 2021.

4. Bohannon J. Who's afraid of peer review? [Internet] Science 2013;342:60-5 [[acesso](#) em: 19 mar. 2021].
5. Prado PI, Kraenkel RA, Coutinho RM. Preda Qualis – Periódicos potencialmente predatórios no QUALIS-CAPES [Internet]. 2017 [[acesso](#) em: 19 mar. 2021].
6. Beall J. Beall's list of potential predatory journals and publishers. 2017.
7. Brito S. Estudo rejuvenesce células de idosos em 25 anos [Internet]. Veja. 2016 [[acesso](#) em: 19 mar. 2021].
8. Selvaraj S, Borkar DS, Prasad V. Media coverage of medical journals: Do the best articles make the news? PLoS One. 2014 Jan 17;9(1).
9. Fioravanti C, Fioravanti C. Otimismo em um mar de incertezas: a cobertura jornalística sobre a pesquisa de novos medicamentos no Brasil. J Sci Commun [Internet]. 2018;17(2):1-19 [[acesso](#) em: 19 mar. 2021].
10. Sumner P, Vivian-Griffiths S, Boivin J, Williams A, Venetis CA, Davies A, et al. The association between exaggeration in health related science news and academic press releases: retrospective observational study. BMJ [Internet]. 2014 Dec 9;349(dec09 7):g7015-g7015 [[acesso](#) em: 19 mar. 2021].

# Table of Contents

[Capa](#)

[Créditos](#)

[Como ler este livro?](#)

[Autores](#)

[Sumário](#)

[Prefácio](#)

[Seção 1. Fundamentos](#)

[1. Por que precisamos de evidências? Onde encontrá-las? Quando não precisamos delas?](#)

[Introdução](#)

[O que é medicina baseada em evidências?](#)

[Importância da medicina baseada em evidências](#)

[Onde encontrar a melhor evidência científica?](#)

[Quando não precisamos de evidências científicas?](#)

[Referências bibliográficas](#)

[2. Medicina enviesada por evidências](#)

[Introdução](#)

[Os riscos da interação das empresas de biotecnologia com as pesquisas clínicas](#)

[Os riscos da interação de empresas de biotecnologia com a elaboração de guidelines \(diretrizes\)](#)

[Os riscos do mau uso das evidências em benefício pessoal ou político](#)

[Retracted papers – artigos despublicados](#)

[O que fazer para evitar cair nesse jogo “sujo”?](#)

[Referências bibliográficas](#)

[3. A pirâmide fluida da Medicina Baseada em Evidências e o caminho de uma hipótese até a sua confirmação](#)

[Introdução](#)

[Critérios de Bradford Hill](#)

[Alerta para a fluidez da pirâmide se iniciando agora](#)

[Relatos de casos](#)

[Estudos ecológicos](#)

[Pesquisas transversais – observacionais](#)  
[Estudo de caso-controle – observacionais](#)  
[Estudos de coorte – observacionais](#)  
[Estudos observacionais – um resumo](#)  
[Estudos clínicos randomizados – o topo da pirâmide, mas nem sempre](#)  
[Meta-análises e revisões sistemáticas](#)  
[Conclusões](#)  
[Referências bibliográficas](#)

#### [4. Vieses em estudos científicos](#)

[Introdução](#)  
[Viés do franco-atirador texano](#)  
[Escolha de cerejas \(cherry-picking\)](#)  
[Viés de seleção](#)  
[Viés de aferição](#)  
[Vieses de relato](#)  
[Viés cronológico](#)  
[Vieses de intervenção](#)  
[Vieses inerentes ao rastreamento](#)  
[Rob-2](#)  
[Referências bibliográficas](#)

#### [5. A análise bayesiana dos estudos científicos](#)

[Introdução](#)  
[O que é essa tal análise bayesiana e o que isso tem a ver com o livro?](#)  
[Teorema de Bayes em poucas palavras e sem firulas](#)  
[Onde isso entra na análise de artigos científicos?](#)  
[Indo muito além do valor de p](#)  
[Indo muito mais além do valor de p](#)  
[Como incorporar o pensamento bayesiano na análise de pesquisas?](#)  
[Referências bibliográficas](#)

#### [6. Estatística para quem não gosta de matemática](#)

[Introdução](#)  
[Os primeiros passos – entendendo porcentagens](#)  
[Você não se complica pelo que você não sabe, mas pelo que você sabe com certeza, só que não é verdade](#)



[Medidas de tendência](#)

[Medidas de dispersão](#)

[Comparando tendências e dispersões](#)

[Interpretando o valor de p](#)

[Chance ou probabilidade? Uma introdução às medidas de associação](#)

[Entendendo o nnt e o papel limitado dos médicos na história natural das doenças](#)

[Referências bibliográficas](#)

## [7. Como interpretar artigos sobre terapias?](#)

[Introdução](#)

[Por que prescrevemos medicamentos?](#)

[Passo 1 – qual o tipo de pesquisa?](#)

[Passo 2 – o estudo foi cego?](#)

[Passo 3 – o estudo comparou o que com o quê?](#)

[Passo 4 – a hipótese é plausível? Tem boa probabilidade pré-teste?](#)

[Passo 5 – o estudo teve sensibilidade adequada?](#)

[Passo 6 – o que os desfechos do estudo têm a nos dizer?](#)

[Passo 7 – qual a especificidade do resultado encontrado?](#)

[Passo 8 – o estudo tem validação externa?](#)

[Analizando meta-análises](#)

[Revisões sistemáticas](#)

[Referências bibliográficas](#)

## [8. Como interpretar um teste diagnóstico?](#)

[Introdução](#)

[Alguns conselhos](#)

[Variabilidade inter e intraobservador](#)

[Entendendo a sensibilidade e a especificidade de um teste](#)

[Precisamos nos converter ao raciocínio do reverendo](#)

[Thomas Bayes](#)

[Um modelo \(adaptado\) de raciocínio clínico diagnóstico usando o método bayesiano](#)

[O modelo Bayesiano de atendimento](#)

[Referências bibliográficas](#)

## [Seção 2. Especialidades em Evidência](#)

### [9. Cardiologia baseada em evidências](#)

Introdução

Tratamento das arritmias baseado em evidências

Tratamento das valvopatias baseado em evidências

Tratamento da insuficiência cardíaca baseado em evidências

Tratamento da doença arterial coronariana baseado em evidências

Prevenção primária e secundária com AAS e estatinas na doença arterial coronariana

Tratamento da hipertensão arterial sistêmica baseada em evidências

Conclusões

Referências bibliográficas

## 10. Endocrinologia baseada em evidências

Introdução

Hipotireoidismo subclínico

Ultrassonografia como rastreio de câncer de tireoide

Obesidade

Perda de peso e redução de riscos de saúde

Diabetes mellitus

Pré-diabetes e síndrome metabólica

Conclusões

Referências bibliográficas

## 11. Oncologia baseada em evidências

Introdução

Os principais desfechos em oncologia são:

Questões adicionais a serem feitas nos estudos clínicos

Conclusão

Referências bibliográficas

## 12. Rastreamento de câncer baseado em evidências

Introdução

O que é rastreamento?

Os efeitos do rastreamento

Vieses em avaliações de rastreamento

Conclusão

Referências bibliográficas

## 13. Cirurgia baseada em evidências

A cirurgia baseada em dogma

O problema do dogmatismo

A mudança do paradigma tradicional para o paradigma científico na cirurgia

O raciocínio clínico probabilístico na cirurgia

Cirurgia baseada em evidências

Cirurgia baseada em evidências – as dificuldades

Cirurgia baseada em evidências – a reforma iluminista

Conclusões

Referências bibliográficas

#### 14. Dermatologia baseada em evidências

Introdução

Dermatologia baseada em evidência

O ensino da Dermatologia

Doenças raras

Desenho dos estudos disponíveis

Drogas antigas

O crescente papel das pesquisas financiadas pela indústria

O risco do mecanicismo

Cochrane skin group

Boas aplicações de mbe na Dermatologia

Cosmiatria

Conclusão

Referências bibliográficas

#### 15. Ortopedia baseada em evidências

Introdução

Compressa quente versus fria

Fortalecimento muscular

Kinesiotaping em ortopedia

Evidência no uso de ácido hialurônico

Conclusão

Referências bibliográficas

#### 16. Jornalismo de saúde baseado em evidências

Introdução

Ciência cumulativa

Onde o estudo foi publicado importa

Conflitos de interesse

Tipos de estudo

[In vitro e in vivo](#)

[Estatística](#)

[Preprints](#)

[Incerteza e linguagem](#)

[Sensacionalismo](#)

[Referências bibliográficas](#)