

Estudos observacionais: quando confiar, desconfiar ou descartar?

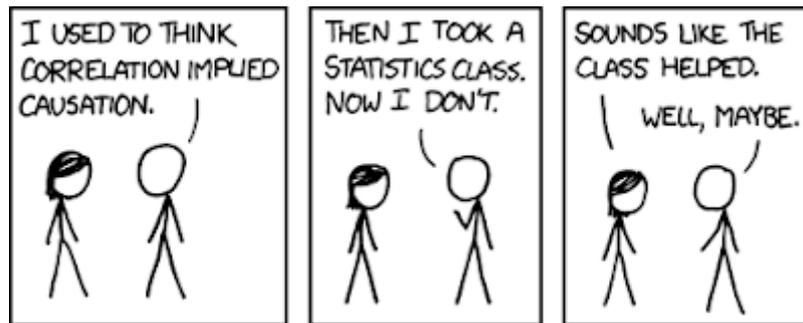
Por: Luis Claudio Correia

A popular utilização de coortes observacionais para testar benefício ou segurança de condutas e hábitos de vida talvez seja a mais importante fonte de falsas notícias médicas, mediadas principalmente por fatores de confusão. Assim cabe a pergunta: quando estes trabalhos poderiam trazer algum tipo de contribuição científica e quando nem deveriam existir?

Atualmente, um grande número de publicações existe em função da manutenção do ecossistema da pesquisa na era industrial (produção em quantidade), dificultando a evolução para a era do conhecimento. Para conceito de eficácia ou para atribuir efeitos adversos a condutas médicas, muitas publicações observacionais servem mais para impulsionar carreiras acadêmicas do que para gerar valor científico.

Para complicar, vários destes estudos provêm de coortes imensas, com alta precisão estatística. Este é o caso no estudo [PURE](#), que utilizou uma amostra maior que 100.000 pacientes para concluir a favor da dieta com baixo teor de carboidrato na prevenção de eventos cardiovasculares. É paradoxal, pois imensas coortes observacionais possuem grande precisão para descrever resultados enviesados. Grande precisão significa baixa aleatoriedade, o que eleva a consistência do achado. No entanto, vieses (neste caso, viés de confusão) são erros que se repetem sistematicamente na amostra. Teremos portanto a consistência de um resultado potencialmente falso, com alto grau de significância estatística. É como uma mentira bem contada, que parece se tornar verdade por ser muito consistente.

Meses depois, foi publicada a [coorte do ARIC](#), sugerindo o contrário, dieta com baixo teor de carboidrato pode se deletéria. Em qual acreditar? Em nenhum dos dois. Sorte que a preguiça me impede de aprofundar a leitura desses trabalhos e sair com *insights* pseudo-elegantes.



Quando confiar em coortes observacionais?

Coortes observacionais representam o desenho de primeira linha para construção de modelos probabilísticos preditores de risco; também são estes os estudos primordiais para avaliar efetividade, ou seja, o comportamento de uma conduta no mundo real, que depende da variabilidade do ambiente, do processo de decisão médica (no mundo real não se sorteia a decisão) e da preferência do paciente. Porém no caso da inferência de causalidade (fatores de proteção ou risco), a discussão quanto à veracidade da evidência precisa ser aprofundada. É o que faremos nesta postagem.

Sabemos que para explorar benefício promovido por hábitos ou condutas, ensaios clínicos randomizados representam o único desenho de estudo que pode ser considerado confirmatório, o único capaz de eliminar efeitos de confusão. Embora haja muitas afirmações neste sentido, apenas ensaios clínicos randomizados seriam capazes de comprovar que uma taça de vinho ao dia tenha efeito cardiovascular protetor.

Por outro lado, na avaliação de “dano”, coortes observacionais podem ser consideradas evidências confirmatórias em casos especiais, quando três critérios forem preenchidos: **grande probabilidade pré-teste (plausibilidade biológica), grande força de associação e impossibilidade ética de randomização.** É o caso da associação entre etilismo e cirrose

hepática, cujo **risco relativo é em torno de 20**. Esta associação é tão forte que se torna pouco provável que seja toda mediada por fatores de confusão.

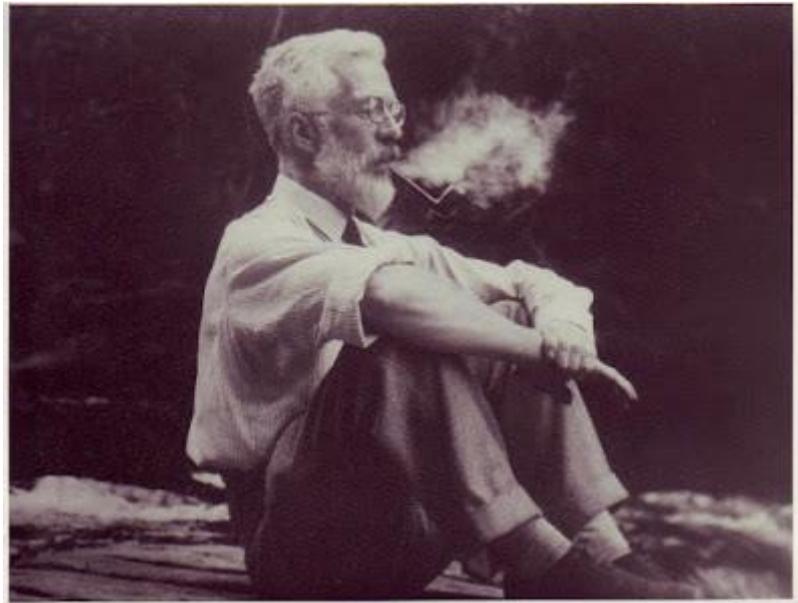
Observem que o efeito do álcool no organismo humano traz os dois lados da moeda. A obrigatoriedade do ensaio clínico para propor vinho como conduta cardiovascular protetora, e suficiência de evidências observacionais para propor etilismo como hábito danoso ao fígado. Por trás dessa maior flexibilidade na aceitação de uma evidência observacional está a constatação de que a **eliminação de um fator provavelmente deletério tolera melhor a incerteza do que a implementação de uma estratégia**.

Toda ação (implementação de conduta) tem maior "ônus da prova" do que uma não ação (deixar de fazer algo).

A relação de tabagismo e câncer de pulmão é outro exemplo de conceito confirmado por estudos observacionais. O risco de quem fuma é 20 vezes maior do que o risco de quem não fuma (risco relativo = 20). Isto é diferente da força de associação entre tabagismo e doença cardiovascular, cujo **risco relativo para um tabagista** (40 cigarros ao dia) é "apenas" 2.0. A doença cardiovascular é multifatorial e tabagismo representa apenas um fator, com peso semelhante aos demais. Um dado que reforça esta relação causal é o gradiente dose-resposta de tabagismo e eventos cardiovasculares. Além disso, há inúmeros motivos para se recomendar contra o tabagismo. Mas se não fosse por estas questões, algum ceticismo se aplicaria a esta relação cuja força de associação é semelhante à da hipercolesterolemia e hipertensão. Diferente de tabagismo, estes dois fatores de risco possuem a prova do conceito confirmada por ensaios clínicos randomizados com anti-hipertensivos e drogas redutoras de colesterol.

Ronald Fisher, tabagista e pai da estatística, provocou a comunidade médica ao questionar a associação causal entre tabagismo e câncer de pulmão, devido às conhecidas limitações de estudos observacionais. A intenção deve ter sido chamar a atenção para o conceito de efeito de confusão, originalmente descrito por ele próprio, que também propôs o

artifício da randomização como forma de ajuste perfeito para confusão. Do ponto de vista epidemiológico um melhor exemplo seria a associação entre tabagismo e doença cardiovascular, pois a força de associação com câncer de pulmão é grande demais para não ter causalidade implícita. Mas ele queria causar desconforto, surpresa, controvérsia, chamando atenção para o fenômeno da confusão. Neste sentido, câncer de pulmão foi um exemplo mais provocativo. Vale a pena [ler sobre esse estória](#).



Quando desconfiar de coortes observacionais?

A maioria dos estudos observacionais que propõem fatores causadores de doenças não completam os critérios exemplificados pelo tabagismo ou alcoolismo. Nestes casos, devemos desconfiar.

Boa parte dos cânceres são desprovidos de causa genética ou ambiental. Motivados por um instinto de sobrevivência anticientífico, saímos a procura de pseudocausas. De forma caricatural, quase tudo pode causar câncer e quase tudo tem respaldo por trabalhos de coorte que avaliam múltiplas associações, com alto risco de viés de confusão e aferição, encontrando todos os tipos de resultados. Vivemos em um mundo de fantasia. O mesmo ocorre com fatores protetores de câncer.

Nesta semana, o *JAMA Internal Medicine* publicou coorte de 60.000 pacientes sugerindo que comidas orgânicas reduzem incidência de câncer. Dá arrepio em observar a tabela que compara as características clínicas da amostra dividida em quartis do escore de ingesta orgânica. Tudo é diferente, abrindo enorme possibilidade de efeito de confusão, nunca plenamente resolvido por artifícios estatísticos multivariados.

Ao desconfiar, partiremos para uma análise mais aprofundada de quando valorizar o estudo como gerador de hipótese ou quando descartar o estudo pelo seu grande potencial de gerar fake news científica.

Nesta análise, duas perguntas se tornam relevantes:

- Em que situações estes desenhos de estudo são menos susceptíveis a vieses, podendo ser usados pelo menos como geradores de hipótese?
- Estudos observacionais tendem mais a resultados falso positivos ou falso negativos?

A chave destas questões está na análise da potencial “**confusão por indicação**”. Ou seja, o quanto a indicação de um tratamento pelo médico ou opção de um hábito pelo paciente potencialmente se relaciona com variáveis confundidoras.

Situações de Baixo Potencial de Confusão

Ocorre quando a “indicação” não foi uma opção de acordo com o paciente, mas sim uma característica do ambiente. Por exemplo, um serviço tem a rotina de cirurgia laparoscópica e outro serviço a rotina de cirurgia aberta. A comparação dos resultados destas duas técnicas não sofre muito de confusão pois não é uma escolha feita com base na característica do paciente. Portanto a escolha não carrega variáveis de implicam em pacientes mais ou menos graves, desde que ambos os serviços tenham a mesma complexidade.

Ao compararmos a taxa de transfusão de sangue ou hemoderivados na cirurgia cardíaca de pacientes operados em uma época em que o

tromboelastograma não era disponível em nosso hospital *versus* a época em que ficou disponível, a indicação desse teste hematológico não foi baseada no paciente. Trata-se do mesmo hospital, da mesma complexidade. Menor probabilidade de efeito de confusão.

Nestes casos, após análise multivariada para ajusta diferenças confundidoras entre duas amostras que podem ter surgido do acaso (não por indicação), podemos considerar estes estudos suficientes para **gerar hipóteses**.

Lembrando que gerar hipóteses é diferente de gerar notícias ... apenas estudos confirmatórios devem gerar notícia para o público em geral.

Por outro lado, quando há “indicação” da conduta de acordo com o paciente, coortes observacionais não devem servir nem mesmo para gerar hipóteses. Muito menos para impulsionar carreiras científicas. O que adianta publicar algo de baixo valor preditivo, seja superestimando associações ou subestimando associações. Estes são os casos de estudos a serem descartados, que podem tender ao falso positivo ou falso negativo.

Quando Coortes Observacionais Tendem ao Falso Positivo

Isto ocorre quando a indicação da conduta se associa a pacientes de melhor prognóstico. No caso de hábitos de vida, a escolha por um melhor hábito marca pacientes com outras características que fazem bem à saúde. Assim, hábitos como exercício, alimentação, uso de suplementos (omega-3, vitaminas) e até uma elegante taça de vinho ao dia se associam sistematicamente a menor risco de qualquer coisa. Sim, de qualquer coisa, pois essas pessoas são muito mais saudáveis.

Quanto à indicação do médico, esta se associa a melhor prognóstico quando o médico tem medo (*risk aversion*) da sua indicação causar dano ao paciente. Desta forma, ele escolhe pacientes menos doentes para adotar o tratamento, e isso carrega um melhor prognóstico para o grupo tratado, o que pode não decorrer da conduta médica *per si*.

Este fenômeno de *risk aversion* é o responsável pelo paradoxo risco-tratamento, onde pacientes de maior risco são menos tratados. Um exemplo é a terapia de anticoagulação para prevenção de AVC em pacientes com fibrilação atrial, especialmente na era da warfarina, quando havia mais relutância pela complexidade do tratamento.

Quando uma coorte observacional tende ao falso negativo

Isto ocorre em tratamentos que os médicos julgam efetivos e não temem as consequências negativas. Ou seja, quando a decisão do médico é mais norteada pelo benefício da conduta do que pelo receio de complicações. Nestes casos, quanto mais graves os pacientes, mais os médicos indicam o tratamento. Isto causa um viés negativo, ou seja, a gravidade dos pacientes aumenta o risco do desfecho e atenuará a percepção de um efeito supostamente benéfico do tratamento.

O melhor exemplo é a decisão por internar ou não internar um paciente com pneumonia. O médico internará os mais graves e por serem mais graves estes pacientes têm maior mortalidade, podendo anular a associação de internamento com desfecho favorável. Pode até parecer que o internamento cause morte, pois os mais graves, internados, têm maior mortalidade.

Trabalho publicado por Ioannidis no [JAMA 2001](#) comparou resultados de estudos observacionais com ensaios clínicos randomizados em diferentes situações. Notamos que o mais frequente é a superestimativa do efeito, ou seja, uma tendência a mais estudos falso positivos do que falso negativos. Isto talvez esteja de acordo com a impressão da maioria dos leitores de artigos: estudos falso positivos são mais frequentes do que falso negativos. Provavelmente não, na verdade esta impressão vem do viés de publicação. As evidências positivas tendem a ser mais publicadas do que as negativas. Sendo assim, devemos considerar que em potencial, os dois mecanismos de falso positivo e negativo coexistem em semelhante proporção.

Conclusão

A definição de qual das três categorias acima se aplica a um estudo de coorte não está contida nos algoritmos tradicionais de *critical appraisal* de evidências observacionais, que focam em vieses de amostragem, observação, qualidade do ajuste estatístico que foi feito para variáveis de confusão (modelos multivariados, escores de propensão). Esta análise tradicional é insuficiente, pois um estudo observacional de metodologia perfeita não deve ser necessariamente considerado como suficiente para gerar notícia médica. Precisamos fazer a avaliação aqui proposta e sugiro que esta venha antes do *critical appraisal* tradicional.

Meus alunos muitas vezes me perguntam: “estudos observacionais não deveriam existir?”. Respondo sempre com a valorização destes desenhos de estudo para as indicações de modelos preditores, efetividade ou causalidade como o caso do tabagismo. Porém o *insight* da pergunta inocente dos alunos está correto: muitos destes estudos não deveriam existir. Estes muitos estudos são representados por aqueles com alto potencial de viés de confusão negativo ou positivo.

O mundo a olho nu é uma fábrica de ilusões e o viés de confusão é o principal fenômeno ilusório. Mesmo assim, em meio a estudos observacionais podemos considerar alguns fenômenos verdadeiros. Desde que a análise seja criteriosa e feita como uma boa dose de ceticismo. Como solução, proponho o algoritmo abaixo.

Hipótese de Causalidade

