



**BAHIANA**  
ESCOLA DE MEDICINA E SAÚDE PÚBLICA

**CURSO DE MEDICINA**

**JOÃO LUCAS CABRAL CAMPOS**

**PREVALÊNCIA DE ENSAIOS CLÍNICOS INTERROMPIDOS POR POTENCIAL  
BENEFÍCIO: UM MALEFÍCIO À CIÊNCIA**

**Salvador**

**2022**

**JOÃO LUCAS CABRAL CAMPOS**

**PREVALÊNCIA DE ENSAIOS CLÍNICOS INTERROMPIDOS POR POTENCIAL  
BENEFÍCIO: UM MALEFÍCIO À CIÊNCIA**

Trabalho de Conclusão de Curso apresentado à Escola  
Bahiana de Medicina e Saúde Pública para aprovação  
no 4º ano de Medicina.

Orientador: Dr. Luís Cláudio Lemos Correia

**Salvador**

**2022**

## RESUMO

O truncamento de ensaio clínico randomizado (ECR) ocorre quando o estudo é interrompido precocemente por se encontrar um benefício aparente em seus resultados interinos. Esses estudos encontram magnitude de efeito muito maiores do que outros com a mesma questão científica que, no entanto, não foram truncados, ou seja, os truncados encontram resultados provavelmente superestimados. Como sua prevalência na literatura tem aumentado em uma tendência linear desde 1975, e cerca de ¼ dos ECRs truncados são da área de cardiologia, esse problema se faz relevante na área. Assim, o objetivo primário do presente trabalho é descrever a prevalência de ensaios clínicos truncados na literatura cardiológica de alto impacto. Em desenho de caráter metacientífico, foram identificadas todas as publicações feitas nos anos de 2019 e 2020 nas cinco revistas de medicina geral ou de cardiologia de maior impacto: New England Journal of Medicine (NEJM), Journal of the American Association (JAMA), Circulation, Journal of the American College of Cardiology (JACC) and European Heart Journal (EHJ). Dentre essas publicações, foram incluídas as que abordassem a terapêutica cardiovascular através de um ECR. Posteriormente, foram extraídas as características de cada estudo, sendo elas: presença de financiamento da indústria farmacêutica; presença de cálculo do tamanho amostral; o tamanho amostral obtido ao fim do estudo; se alcançou o tamanho amostral planejado inicialmente; se é unicêntrico ou multicêntrico; se conduziu uma análise de superioridade ou de não-inferioridade; a positividade do estudo em relação ao desfecho primário; e se ele foi interrompido ou não. Foram incluídos 231 ECRs em nossa amostra, havendo uma prevalência de 3,1% (95% IC, 1,25% - 6,28%) de estudos truncados. Devido ao tamanho amostral que obtivemos, não foi possível inferir sobre suas características. A prevalência atual de ECRs truncados em cardiologia nas revistas de maior impacto é baixa.

**Palavras-chave:** Truncamento. Ensaio clínico randomizado. Cardiologia. Metaciência.

## ABSTRACT

Randomized clinical trial (RCT) truncation occurs when the study is stopped early based on an apparent benefit found in its interim analyses. Such studies find much higher effect sizes than others with the same scientific question that were, however, not truncated, i.e., the truncated ones find results likely to be overestimated. Since their prevalence in the literature has increased in a linear trend since 1975, and about ¼ of the truncated RCTs are from the cardiology field, this problem becomes relevant in this field. Thus, the primary objective of the present study is to describe the truncated clinical trials prevalence in the high-impact cardiology literature. In a metascientific design, we identified all published articles in the years 2019 and 2020 in the five highest impact general medicine or cardiology journals: New England Journal of Medicine (NEJM), Journal of the American Association (JAMA), Circulation, Journal of the American College of Cardiology (JACC) and European Heart Journal (EHJ). Among these publications, those that addressed cardiovascular therapy through an RCT were included. Subsequently, the characteristics of each study were extracted, namely: presence of pharmaceutical industry funding; presence of sample size calculation; sample size obtained at the end of the study; reach of the sample size initially planned; whether it is unicenter or multicenter study; whether it conducted a superiority or non-inferiority analysis; the positivity of the study regarding the primary outcome; and whether it was stopped early or not. We included 231 RCTs in our sample, with a prevalence of 3.1% (95% CI, 1.25% - 6.28%) of truncated studies. Due to the sample size we obtained, it was not possible to infer about their characteristics. The current prevalence of truncated RCTs is low in cardiology and in the highest impact journals.

**Keywords:** Stopped early for benefit. Randomized clinical trial. Cardiology. Metascience.

## SUMÁRIO

1 INTRODUÇÃO.....	5
2 OBJETIVOS .....	12
3 REVISÃO DE LITERATURA .....	13
4 MATERIAIS E MÉTODOS.....	18
5 RESULTADOS .....	21
6 DISCUSSÃO DOS RESULTADOS.....	25
7 CONCLUSÃO.....	29
REFERÊNCIAS .....	29

## 1 INTRODUÇÃO

Os ensaios clínicos randomizados (ECRs) são considerados padrão-ouro para demonstrar ou refutar benefícios de intervenções.<sup>1</sup> Para que a randomização dos tratamentos seja ética, além de um bom modo de avaliar o efeito, é preciso que a comunidade científica apresente uma genuína incerteza diante de qual tratamento é o melhor, uma vez que seria antiético oferecer intencionalmente um tratamento que é pior, ou que não funciona.<sup>2</sup> Esse é o princípio da *Equipoise*, ou melhor, o princípio da Incerteza, em que o pesquisador se abstém momentaneamente de suas tendências pessoais.<sup>2</sup> Devido a isso, muitas vezes se acredita que essa dúvida já não existe mais, quando através dos resultados preliminares das análises interinas há forte tendência de que um tratamento seja superior ao outro. Nesse contexto, alguns pesquisadores argumentam que os estudos devam ser interrompidos, uma vez que seria antiético não oferecer o melhor tratamento aos pacientes que antes estavam no braço do estudo recebendo o placebo ou o outro tratamento.<sup>3</sup>

À interrupção de um ECR por se encontrar benefício por meio de uma análise interina é dado o nome de truncamento.<sup>4-7</sup> Quando a análise interina é planejada, ela pode ocorrer conforme um intervalo de tempo definido após o início do estudo, um tempo de acompanhamento definido (i.e., em pessoa-tempo), um número de participantes atingido e um número de eventos observados.<sup>8,9</sup> Nesse processo, podem ser predefinidas algumas regras de parada, que caso sejam atingidas podem conduzir o estudo a uma interrupção, seja por benefício ou por futilidade, sendo os 3 métodos de análise mais comuns: a regra de O'Brien-Fleming que requer valores de  $p$  muito baixos nas análises interinas, fazendo-se necessário um valor de  $p$  cada vez maior, ou seja, mais próximo de 0,05, à medida que alcança o tamanho amostral planejado (por exemplo, caso sejam três análises, para ser interrompido na primeira seria necessário um valor tão baixo quanto um  $p < 0,0005$ ); a regra de Pocock que mantém um mesmo limite para todas as análises interinas, cujo exato valor de  $p$  depende do número de análises interinas planejadas, como o valor de 0,0294 quando forem duas análises interinas, e de 0,0158 quando forem cinco; a regra de Peto-Haybittle em que um valor de  $p < 0,001$  é suficiente para a parada sem depender do tamanho amostral alcançado ou de quantas análises interinas foram planejadas.<sup>3,5,10</sup> No entanto, os pesquisadores e estatísticos discordam quanto ao uso correto e ético dessas regras, devido aos

problemas de estimativa que podem decorrer desse uso. Assim, a preocupação com os estudos interrompidos tem atingido grande relevância na comunidade científica, sendo inclusive mencionada em diretrizes influentes, a exemplo da "*GRADE guidelines*" e da *CONSORT*, e também pelo então diretor do Instituto Nacional de Saúde e Excelência Clínica (NICE) do Reino Unido.<sup>1,11,12</sup>

Algumas simulações demonstraram que os estudos terminados precocemente superestimam muito o efeito do tratamento avaliado<sup>5,13,14</sup>, e o próprio intervalo de confiança não é confiável, uma vez que uma grande parte deles sequer inclui o verdadeiro valor do efeito dentro desse intervalo.<sup>14</sup> Enquanto isso, uma simulação indicou que nem sempre ocorreria, e quando ocorresse não seria tão grande assim.<sup>15</sup>

Em concordância com a maior parte das simulações, os estudos empíricos evidenciaram que cerca de 90% dos ECRs truncados apresentavam um benefício superior ao de ECRs com a mesma questão, mas que não foram descontinuados, o que pode ser causado por diversos fatores distintos.<sup>16</sup> O efeito demonstrado em um estudo sofre altas e baixas aleatórias, sofrendo posteriormente o que se chama de "retorno à verdade", quando o valor observado regressa ao valor real.<sup>17</sup> Desse modo, o benefício encontrado no truncamento pode estar enviesado porque o estudo tende a ser truncado quando a diferença entre os tratamentos estiver aleatoriamente alta.<sup>17</sup> Então, mesmo que a análise interina seja previamente planejada, o valor tende a ser superestimado.<sup>17</sup> Esse problema de se interromper em uma alta ou baixa aleatória pode ser chamado de viés condicional, que pode ser mitigado ao se utilizar métodos mais complexos para a interrupção de um estudo<sup>13,15,18,19</sup>, como através da função de gasto alfa de Lan-DeMets.<sup>20</sup> O viés condicional é a investigação da diferença entre o valor observado e o valor real, uma quantificação do viés através do estado atual de uma análise, isto é, a diferença entre a estimativa e o valor real, a qual é maior em amostras menores.<sup>18</sup> Logo, o benefício observado pode estar superestimado principalmente em amostras pequenas.<sup>17</sup>

A maioria dos estudos empíricos demonstrou de fato uma superestimação por parte de ECRs truncados. Em contrapartida, um estudo metacientífico que incluiu poucos artigos em sua análise afirmou que o efeito do tratamento de um ECR truncado permanece parecido quando comparado ao efeito observado no acompanhamento de sua amostra após a interrupção.<sup>21</sup> No entanto, uma vez que essa amostra já demonstrava uma tendência, a interrupção pode ter sido a responsável por manter um

benefício semelhante ao de quando o estudo foi descontinuado, por meio da perda do cegamento após a interrupção por exemplo, deixando a análise suscetível a outros vieses, que previamente eram pouco influentes ou não existentes. Portanto, não se caracteriza como uma evidência muito forte de que os efeitos são bem estimados no truncamento.

O tamanho amostral influencia na probabilidade de que os resultados sejam replicados por estudos futuros, uma vez que ECRs com amostras menores são mais prováveis de serem refutados futuramente ou de o efeito visto em estudos subsequentes ser menor.<sup>22</sup> Contudo, tanto truncados com grande tamanho amostral, quanto os que apresentavam muitos eventos superestimam o benefício, sendo extremamente superestimado quando o truncado contém menos de 200 eventos observados, muito superestimado entre 200 e 500 eventos, e com mais de 500 eventos o truncado exibe uma superestimação moderada.<sup>16</sup> Então a superestimação é maior quanto mais precoce for a análise interina ou o truncamento.<sup>23</sup> Ademais, aproximadamente metade dos ECRs subsequentes apontam um resultado contrário àquele evidenciado pelo ECR truncado previamente, na maior parte por não haver diferença estatisticamente significativa.<sup>24</sup> Em suma, estudos posteriores com a mesma questão reconhecem a hipótese nula em metade das vezes, apesar de o estudo anterior ter sido interrompido porque esse benefício já seria "claro demais". Na outra metade os estudos ostentam um benefício inflacionado, principalmente quando a amostra é pequena. Esse é um bom exemplo da maldição do vencedor, ideia que sugere que o primeiro cientista a descobrir um efeito é "amaldiçoado" a ter a importância de sua descoberta diminuída à medida que estudos novos são feitos.<sup>25,26</sup> O "vencedor" acaba encontrando um efeito inflacionado porque, em estudos menores, o poder estatístico do estudo é capaz identificar apenas magnitudes de efeito grandes, e ao acaso algum estudo acaba por identificar, mesmo que o efeito real fosse menor.

Frequentemente, considera-se o valor de  $p$  como um método suficiente e infalível para testar algo, acreditando-se que um valor de  $p$  inferior a 0,05 significa que toda a metodologia do estudo foi suficiente para a avaliação de seu objetivo independentemente do poder estatístico e de outras conjecturas.<sup>27</sup> Um estudo observou quantos eventos a mais em um dos braços de um ECR seriam necessários para que o desfecho primário, previamente estatisticamente significativo, se tornasse não significativo.<sup>28</sup> E apesar de todos apresentarem inicialmente um valor de  $p$  inferior

a 0,05; mais de 25% dos ECRs apresentou um valor superior a 0,05 ao se adicionar de um a três eventos a mais no grupo controle.<sup>28</sup> Isso indica o quão frágeis são os resultados estatisticamente significantes quando as amostras não são tão grandes, sobretudo quando não ocorreram muitos desfechos.<sup>28</sup> Como geralmente um valor de  $p$  diminuto é a premissa das regras de parada, percebe-se que descontinuar um estudo a partir delas pode conduzir a valores estatísticos extremamente frágeis.

Apesar de alguns afirmarem que, devido a isso, o benefício nos truncados é “bom demais para ser verdade”<sup>17,29</sup>, e assim facilmente identificado como falso, apenas os truncados com menos de 200 eventos apresentam uma estimativa ponto do risco relativo menor que 0,5.<sup>9</sup> Isso ocorre também porque quanto maior o efeito real menos ele é superestimado, sofrendo mais inflação quando ele é menor.<sup>23</sup> Assim, por mais que estudos maiores também superestimem o efeito, o benefício costuma ser semelhante ao de outros estudos positivos, ou seja, eles podem se camuflar entre outros positivos, porque normalmente o seu benefício não é tão grande que chame a atenção.

Segundo citado anteriormente, quanto mais precocemente se realiza uma análise interina, mais os seus resultados são superestimados, mas esse não é o único problema referente a tais análises: quanto mais análises interinas forem realizadas, ou ainda, quanto mais vezes os dados forem avaliados, maior o risco de o estudo superestimar os benefícios.<sup>5</sup> Problematicamente, foi contemplado que mais da metade dos estudos interrompidos não planejam qualquer tipo de avaliação para interrupção precoce, como uma análise interina, uma regra de parada ou o uso de um DSMB (conselho de monitoramento de dados e de segurança), apesar de terem realizado essas análises para então decidir terminá-los.<sup>30</sup> Aliás, 80% dos estudos interrompidos por benefício ou futilidade foram interrompidos fora de uma análise interina formal, seja porque ela não foi planejada, ou porque foi interrompido fora de suas condições previstas.<sup>30</sup> Uma vez que os estudos truncados são corriqueiramente avaliados mais vezes do que o preestabelecido, essa má conduta se torna um acentuador da superestimação já inerente ao truncamento.

Todos esses problemas/erros de estimativa acarretam outros empecilhos à comunidade consumidora de Ciência, com consequências científicas, éticas e práticas. A consequência direta da superestimação é a imprecisão com a qual é tratado o benefício. Embora alguns defendam que em condições graves e

rapidamente progressivas, como o câncer, não seja necessário identificar com precisão qual a magnitude do efeito <sup>23</sup>, esse é um elemento fundamental para o processo de decisão médica, ainda mais nessas condições.<sup>31,32</sup> Sem se saber qual o tamanho do benefício, não se pode ponderá-lo com os riscos existentes. Um bom exemplo disso é um ECR que avaliava o efeito de um tratamento para o câncer de mama, com 13175 participantes e 368 eventos.<sup>33</sup> Ele apresentou uma diferença de incidência de câncer de mama invasivo entre o grupo controle e o tratamento que alcançou as regras de parada, levando ao truncamento do estudo.<sup>33</sup> Posteriormente, em um follow-up do estudo 7 anos após os resultados preliminares, quando os pacientes do grupo controle foram permutados para o tratamento, não se demonstrou benefício quanto à mortalidade geral ou à mortalidade relacionada ao câncer de mama, além de evidenciar que o tratamento ocasionava um maior risco de catarata, câncer endometrial invasivo e de tromboembolismo pulmonar.<sup>34</sup> Isso mostra que o benefício talvez não fosse tão bom quanto o relatado inicialmente pelo truncamento do estudo, uma vez que o benefício não depende somente da incidência de câncer invasivo, mas também da mortalidade e qualidade de vida, expondo desnecessariamente as mulheres a um risco, que previamente não teriam por estarem no grupo controle.

A geração de uma nova evidência de benefício pode ter diversos impactos na produção científica futura ou em andamento. Visto que um dos principais motivos para um estudo ser descontinuado por ter um recrutamento lento demais é a publicação de um potencial benefício durante a condução deste estudo ou anteriormente a ele na mesma questão avaliada <sup>35</sup>, muitos referem que é possível que o truncamento ocasione uma espécie de “congelamento” da produção científica na área abordada pelo estudo truncado.<sup>32</sup> Aventa-se que isso ocorra em consequência de conflitos de interesse que instituições com fins lucrativos possam apresentar em refutar os resultados iniciais apresentados pelo estudo truncado, do desconforto dos pesquisadores em conduzir um novo estudo considerando o grande efeito percebido no anterior, e das respectivas implicações éticas de se conduzir um novo estudo abordando uma questão que, para muitos, já está resolvida.<sup>32,36,37</sup> Além disso, pode ser considerado antiético continuar a conduzir os estudos que avaliam o tratamento agora considerado obsoleto, sendo pedido o seu encerramento prematuro.<sup>38</sup> O fenômeno do congelamento é particularmente difícil de se medir, pois sequer metade

dos estudos interrompidos por recrutamento lento chegam a ser publicados, conforme evidenciado em trabalhos que buscam avaliá-lo através dos protocolos de estudo.<sup>39</sup> Como a maior causa de descontinuação de um estudo é o recrutamento lento e muitos não são publicados, a própria reprodutibilidade necessária à ciência estará comprometida, passando muitos anos até que se identifique um possível erro de estimativa, ou ainda, um resultado contrário.<sup>40</sup> Apesar de essas conjecturas serem válidas e considerando a dificuldade de se avaliar o congelamento, uma simulação realizada em 2016 argumenta que o "efeito de congelamento" é mínimo<sup>15</sup>, mantendo-se indefinido o quão relevante isso pode ser na realidade.

Outro ponto é a influência do truncamento em metanálises. Muitas simulações afirmam que a inclusão de truncados na metanálise não alteraria substancialmente o seu resultado.<sup>15,41,42</sup> Isso pois teoricamente se verifica que haveria uma fonte oposta de vies: os estudos que planejam regras de parada sem jamais tê-las alcançado, ou seja, que não apresentaram um benefício satisfatoriamente grande capaz de conduzir o estudo à interrupção.<sup>5,42,43</sup> Assim, os ECRs truncados deveriam continuar nas metanálises<sup>43</sup>, pois a inclusão equilibrada e ponderada dos estudos não interrompidos, que continham regras de parada, e aqueles que foram truncados poderia contrabalançar os vieses advindos de cada um desses tipos de estudo.<sup>5,43,44</sup> Isso poderia ser feito inclusive porque a previsão feita pelas simulações estatísticas era de que os ECRs com resultados tão extremos que os levaram a ser interrompidos apresentariam muito menos peso nas metanálises que os estudos levados até o fim<sup>41</sup>, permitindo que a pequena desigualdade entre a superestimação e subestimação causada pelas regras de parada não impactassem tanto o resultado da metanálise. Toda via, não foi o que evidenciaram os resultados empíricos. Em cerca de metade das metanálises que incluíram um truncado, avaliando o mesmo desfecho que o levou à interrupção, eles apresentaram um peso maior de que 40%.<sup>45</sup> Nesse contexto, a superestimação inerente ao truncamento induz o resultado da metanálise. Essa situação torna-se ainda mais impactante ao passo que a maior parte das revisões sistemáticas (71%) não identificam quando um estudo truncado foi incluído, mesmo nos casos em que se planeja conduzir uma metanálise.<sup>45</sup> Por isso, como solução pode ser realizada uma análise de sensibilidade, isto é, uma análise incluindo e a outra excluindo os truncados.<sup>12</sup> Caso as estimativas difiram significativamente, a análise que excluiu os truncados deve ser considerada a mais confiável.<sup>12</sup>

A definição de que um novo tratamento é benéfico tem inúmeras repercussões éticas além das supracitadas. De fato, o truncamento é um processo extremamente atraente para o pesquisador inclusive neste âmbito, pois quanto mais rápido for divulgada a eficácia de um tratamento, mais pessoas poderão se beneficiar dele, ocorrendo, muitas vezes, anos antes da inscrição do último paciente do estudo.<sup>3</sup> Aqueles que focam nesta vantagem julgam que o uso das regras de parada é essencial para o desenvolvimento eficiente e ético de terapias.<sup>46</sup> Já os que consideram antiético interromper um estudo quando se espera que a continuação do estudo traga uma informação científica relevante focam na possibilidade de se estar expondo ainda mais pacientes a um tratamento cuja evidência de benefício ainda não é definitiva.<sup>36,47</sup> Nessas situações, é possível que os pacientes do próprio estudo realizem permuta para o braço do tratamento, além de que outros pacientes serão expostos para a realização de estudos mais robustos com o intuito de fornecer o resultado definitivo.<sup>3,36</sup> Para se realizar um julgamento ético a respeito do truncamento é necessário se pensar nesse segundo foco, pois caso a interrupção precoce de um estudo por benefício é ético, então seria antiético conduzir novos estudos que abordassem a mesma questão.<sup>24</sup> No entanto, metade dos truncados são seguidos por ao menos um ECR com uma questão semelhante, e por volta de 30% a 70% dos ECRs subsequentes são realizados devido ao ceticismo dos pesquisadores em relação aos resultados do truncado, seja pelo pequeno tamanho amostral, pelo elevado risco de viés, ou ainda pela inconsistência com outras evidências.<sup>24</sup> Logo, é imprescindível que cada estudo que avalie eficácia almeje prover uma resposta definitiva para a sua questão, justamente para beneficiar a maior quantidade de pessoas, não expondo desnecessariamente os pacientes aos riscos de um tratamento que talvez nem mesmo funcione.

Como outra vantagem, os próprios recursos financeiros voltados para o estudo podem ser investidos em outras questões que ainda não foram investigadas.<sup>3</sup> Mas, nesse ponto, reside um grande perigo: o de se interromper por questões comerciais, pois na área de oncologia, por exemplo, cerca de 80% dos estudos são utilizados para conseguir um registro para a droga desenvolvida pela indústria farmacêutica, e comumente os ECRs truncados deixam de recrutar 1/3 dos pacientes planejados, economizando bastante tempo e dinheiro.<sup>8</sup>

Por fim, pode levar a recomendações práticas e condutas clínicas inadequadas.<sup>48</sup> O que pode ser ilustrado pelo ECR que avaliou eficácia de bisoprolol em pacientes que realizaram cirurgia vascular eletiva, que identificou precocemente um benefício estatisticamente significativo quanto à mortalidade e à incidência de infarto agudo do miocárdio, sendo, pois, interrompido.<sup>49</sup> Diretrizes de todo o mundo passaram a recomendar essa intervenção no cuidado pré-operatório de cirurgias não cardíacas, a exemplo das diretrizes da American College of Cardiology (ACC) e da American Heart Association (AHA)<sup>50</sup>, da European Society of Cardiology (ESC)<sup>51</sup>, e da Sociedade Brasileira de Cardiologia, tanto a de 2011<sup>52</sup>, quanto a de 2017.<sup>53</sup> Devido a isso, esses tratamentos foram intensamente aplicados em todo o mundo desde aquela época até o momento presente. Entretanto, ECRs subsequentes demonstraram que o benefício demonstrado não era tão grande quanto o evidenciado anteriormente, e foi indicado, por uma metanálise que os incluiu, que o uso de betabloqueador nessas condições aumenta a mortalidade geral e que havia um aumento considerável do risco de AVC<sup>54</sup>, trazendo à tona o problema vigente.

Diante de todos esses problemas, aumenta-se cada vez mais a preocupação com esta questão, já que foi reportado em 2005 que a prevalência de estudos truncados havia aumentado em uma tendência linear ao longo dos 30 anos que antecederam o estudo, especialmente em revistas de alto impacto.<sup>9</sup> Esse tipo de estudo usualmente avalia intervenções das quais os pacientes podem se beneficiar rapidamente caso surja algo novo e eficaz, como em síndromes coronarianas agudas, câncer de pulmão, entre outras áreas, abordando em 1/4 a 1/3 das vezes a área de cardiologia.<sup>9,16,24</sup> Questão que se revela ainda mais presente na área, pois, de todas as revisões sistemáticas que incluíram ao menos um ECR truncado, metade investigam intervenções cardiovasculares, fazendo dessa área a mais afetada pelo encerramento precoce de estudos por se encontrar benefício em uma análise interina.<sup>45</sup>

## **2 OBJETIVOS**

O objetivo geral desse estudo é descrever a prevalência de ensaios clínicos randomizados truncados na literatura cardiológica. Já os objetivos específicos são reportar as características associadas ao truncamento em ensaios clínicos randomizados truncados na literatura cardiológica de alto impacto, e relatar a

prevalência de outras causas de interrupção de ensaios clínicos randomizados na literatura cardiológica de alto impacto.

### 3 REVISÃO DE LITERATURA

O *overuse* é referido na literatura através de expressões como “excesso de medicina”, “cuidado de pouco valor”, “uso inapropriado”, “obsolescência” e “cuidado desnecessário”<sup>55-57</sup>, mas a definição mais geral é referida como “cuidado que pode levar a dano e consumir recursos sem adicionar valor ao paciente”.<sup>58</sup> Ou seja, o *overuse* na medicina tem sua definição associada ao risco de o dano da intervenção superar o benefício potencial<sup>59</sup>, o que inclui também os casos em que os benefícios são insignificantes, ou ainda, em que os pacientes, se fossem completamente informados, não haveriam optado pelo tratamento.<sup>60,61</sup> É um conceito diretamente relacionado ao *overdiagnosis* e ao *overtreatment*, sendo esses seus constituintes básicos.<sup>62</sup> O *overtreatment*, por estar associado a terapias medicamentosas e procedimentos indevidos, expõe o paciente desnecessariamente aos eventos adversos<sup>62</sup>, podendo causar danos ao paciente, de modo que o “tratamento” traz em si um problema que se agrega ao problema anterior, aquele que ele se propôs a resolver, porém sem êxito.<sup>63</sup> A magnitude da problemática é contemplada pela estimativa de que um terço de todos os pacientes passa por procedimentos desnecessários, inefetivos ou potencialmente danosos: erros que em sua maioria eram evitáveis.<sup>64-66</sup>

O *overuse* é um fenômeno geral, assolando inclusive a área de cardiologia, como evidenciado por um estudo de 2010 que reportou que cerca 44% de uma amostra de 1200 pacientes saudáveis, entre 40 e 60 anos, foram submetidos a testes de rastreamento para doenças cardiovasculares, que foram classificados pelo *Consumer Reports* como muito improváveis de conterem benefícios que superassem os riscos.<sup>67</sup> Além disso, a maior parte dos que foram testados não foram devidamente informados pelos profissionais de saúde, sendo que apenas 9% recebeu informações sobre acurácia do teste e 4% sobre potenciais complicações.<sup>67</sup> Assim, os pacientes cardiológicos são frequentemente submetidos sem sequer serem informados adequadamente a respeito dos elementos essenciais de seu itinerário terapêutico.

Algumas ações têm sido tomadas para a redução desses abusos, como com a reforma de responsabilidade médica, através da qual foram reduzidos testes invasivos para doença arterial coronariana, com diminuição de cerca de 24% no número de angiogramas.<sup>63</sup> No entanto, isso tem ocorrido aos poucos, pois alguns elementos relacionados ao cuidado e à saúde servem como uma resistência para avanços mais significativos e rápidos, sendo eles o próprio paciente, o provedor de saúde, as organizações de saúde, o sistema de saúde e a falta de evidência. Os pacientes, por muitas vezes chegarem ao consultório com a crença de que maior benefício é conseguido através do uso de algum procedimento, não compreendendo benefício em se evitar a realização dele <sup>68</sup>, além de uma aparente tendência dos pacientes a não se conformarem verdadeiramente com resultados negativos em testes.<sup>62</sup> Os provedores de saúde se apresentam como um obstáculo ao não possuírem conhecimento a respeito e discordarem de que determinado procedimento seja desnecessário, ou mesmo ao ceder aos anseios ingênuos de um paciente, apesar de saber que sua condição prescinde uma intervenção.<sup>68</sup> Nas organizações e no sistema de saúde, em muitas questões, não se faz presente o conhecimento ou o interesse político, financeiro e intelectual em abordar a problemática.<sup>68</sup> E, por fim, a falta de evidência de que alguns testes ou tratamentos estão sendo utilizados de modo abusivo <sup>68</sup> ajudam a perpetuar essas posturas. A falta de evidência e de confiança nos resultados dos estudos, que influencia na permanência do *overuse*, podem vir de diversos fatores, como a falta de acesso aos dados científicos <sup>68</sup>, ou a resistência em confiar nos resultados, causada pela falta de transparência na metodologia, análise de dados e clareza dos resultados, juntamente à influência do viés de publicação, *spin* e outras formas de moldar os resultados.<sup>69</sup>

Um dos maiores símbolos de evidência científica confiável são os ensaios clínicos.<sup>70</sup> No entanto, infelizmente, eles também são acometidos pelos problemas descritos anteriormente, a exemplo da pobreza de transparência e manipulação. Dentre eles, existe o truncamento, em que eles são encerrados antes do tempo previsto por ter benefício considerado tão evidente que alguns autores acreditam ser antiético continuá-lo diante dos fatos interinos.<sup>71</sup> Múltiplos problemas tendem a surgir quando um estudo é truncado <sup>70,72</sup>, estudos empíricos e simulações sugerem que eles superestimam efeitos do tratamento para o efeito estudado <sup>71</sup>, além de muitas vezes não fornecerem

informações sobre a decisão de truncar o estudo, e apresentar resultados geralmente implausíveis.<sup>9</sup> Apesar desses problemas, o truncamento é cada vez mais comum.<sup>9</sup>

Resultados expressivos devido ao acaso podem ocorrer, principalmente no início de um ECR.<sup>73</sup> Logo, ao encerrar um estudo prematuramente, os investigadores podem descrever dados superestimados.<sup>74,75</sup> Esse efeito é demonstrado no décimo segundo ensaio de leucemia mieloide aguda do *Medical Research Council* que compara a sobrevida em quatro e cinco doses de terapia. Esse estudo, mesmo apresentando resultados positivos altamente significativos a favor de uma dose adicional de tratamento em resultados preliminares em 1998, não foi truncado. Isso aconteceu pois os achados foram considerados demasiadamente significantes para ter plausibilidade clínica, desconsiderando o valor de  $p$  estatisticamente significativa.<sup>73</sup> Ao final do tempo planejado, sua conclusão desse estudo foi que não havia evidência de diferença entre as duas opções de tratamento.<sup>73</sup> Desse modo, apesar de seu resultado interino significativo advir provavelmente de variações aleatórias, em uma publicação nesse formato, uma leitura menos atenta ou menos experiente pode levar muitos profissionais de saúde a submeterem seus pacientes a um tratamento que provavelmente não é benéfico, e quiçá inclusive maléfico.

Ainda assim, o truncamento é realizado, pois interromper estudos precocemente pode economizar uma grande quantidade de dinheiro e evitar que pacientes sejam expostos a tratamentos ineficazes.<sup>76</sup> As regras utilizadas como limites que quando ultrapassados deflagram a interrupção do estudo são as regras de parada. Dois dos métodos mais conhecidos são as regras de O'Brien-Fleming<sup>77</sup>, de Pocock<sup>78</sup>, e de Haybittle-Petto<sup>79,80</sup>. Essas regras não são utilizadas apenas para interromper um estudo por um benefício provável, mas também quando se torna pouco provável de que algum benefício seja evidenciado ao se conduzir o estudo até o final previamente planejado. Esta é a definição de futilidade: improvável que o estudo alcance o tamanho amostral planejado ou improvável de ser identificado um benefício ao fim do estudo.<sup>3</sup>

Em dois extremos estão os estudos que são descontinuados devido a benefício e aqueles cujo motivo de interrupção foi a futilidade. Em adição aos métodos supracitados, há a análise de poder condicional para avaliar futilidade. Esse último significa a chance de se conseguir um resultado significativo estatisticamente ao final do estudo dadas as informações obtidas em uma análise interina.<sup>81</sup> Seu valor é baseado na estimativa inicial do número de desfechos e da razão de azar, em relação

ao número de desfechos e à razão de azar evidenciados no momento da análise interina.<sup>81</sup> Tal qual o truncamento apresenta certas limitações, o mesmo acontece com a interrupção por futilidade. O estudo de Jitlal et al. aborda esse último conceito na estratégia de interromper estudos que são "negativos" antecipadamente, através da análise de 5 ensaios clínicos randomizados que não mostraram benefício e 5 que mostraram benefício moderado.<sup>76</sup> Ele observou que seria problemático terminar os estudos precocemente através do poder condicional, pois mesmo já tendo a maioria dos desfechos e da quantidade de participantes, ainda havia uma baixa chance de apresentar resultados estatisticamente significantes ao final do trabalho.<sup>76</sup> Além disso, caso as análises de futilidade propostas tivessem sido aplicadas aos estudos originais, alguns daqueles que exibiram benefício ao fim do estudo apresentaram baixo poder condicional em alguma análise interina e poderiam ter sido interrompidos inadequadamente, não indicando ao fim os benefícios clínicos.<sup>76</sup> Isso revela que mesmo quando corretamente utilizados, esses métodos podem enviesar o estudo, o que sugere que os resultados de um estudo sem interrupções são mais confiáveis, sendo necessário que os clínicos sempre analisem cautelosamente os resultados dos estudos interrompidos.<sup>9</sup>

A aplicação desses critérios para o encerramento precoce tem sido generalizada, o que tem ocorrido até mesmo com estudos piloto.<sup>82</sup> Eles servem, usualmente, de fundamento para a escolha de quais fármacos devem ser pesquisados através de estudos mais robustos, como ensaios clínicos randomizado. Como exemplo dessa extensão mesmo para esses estudos iniciais, *Mehta e Cain* sugerem que eles podem ser interrompidos em duas condições: quando é extremamente provável que a droga seja eficaz, ou quando é extremamente improvável que ela seja eficaz.<sup>82</sup> No entanto, mesmo quando essas evidências preliminares sugerem um benefício "extremamente provável" de uma intervenção, elas frequentemente são vistas posteriormente como ineficientes, com um efeito menor do que o esperado, ou mesmo, com efeitos clinicamente irrelevantes.<sup>76</sup> Esse tipo de conduta diante dos pilotos, pode parecer um atalho para alcançar mais rapidamente a conclusão do estudo, porém essa execução pode custar caro, considerando os altos investimentos para se realizar um ECR. Em suma, o truncamento é relevante causa de viés, diminuindo a confiabilidade nos estudos do ecossistema científico, estabelecendo-se como um dos perpetuadores do *overuse*. O mesmo princípio se estende também para os estudos que não são pilotos,

havendo custos monetários, sociais e científicos importantes, o que afeta todos os elementos relacionados ao cuidado e à saúde.

Por parte do pesquisador e do provedor de cuidados, o truncamento e a interrupção após uma análise de futilidade são um desejo íntimo de, através de recursos estatísticos cada vez mais elaborados, beneficiar o maior número de pessoas o mais rápido possível, e utilizando a menor quantidade de "cobaias" para a descoberta do benefício. É a busca da "fórmula" que revelará antes de qualquer outra aquilo que precisamos saber para salvar uma vida a mais que seja. O modo de impedir que se prejudique uma pessoa sequer a mais do que o necessário para a "validação estatística". Então mesmo com os riscos de viés acentuados por esses métodos de interrupção, torna-se atraente o seu uso. Além disso, a interrupção pode ter uma influência comercial, como há muito tempo é aventado o problema de estudos serem encerrados por motivos outros que não motivos éticos ou rigorosamente científicos, a exemplo de interrupção de estudos por razões estratégicas das empresas financiadoras, existindo em certas ocasiões cláusulas no protocolo do estudo que permitam que isso ocorra.<sup>83</sup> Apesar de ser um problema antigo, foi reportado em 2005 que a prevalência de truncados está em ascensão, aumentando em uma tendência linear ao longo dos anos, principalmente em revistas de alto impacto.<sup>84</sup> Tal cenário é reforçado pelo viés de publicação, por meio do qual trabalhos positivos são publicados com maior frequência que trabalhos negativos.<sup>85</sup> Assim, através desse mecanismo e de muitos outros, estudos truncados tendem a ganhar espaço na literatura.

Esses estudos constituem parte de todas as áreas do panorama científico, mas atingem algumas com mais afinco. A maioria dos estudos encerrados precocemente avaliam intervenções das quais os pacientes podem obter um benefício de modo relativamente rápido em se utilizar uma nova intervenção, como em síndromes coronarianas agudas, câncer de pulmão e outras condições em que a morte é um desfecho muito frequente.<sup>84</sup> Também alcança muitas áreas em que a produção científica é elevada, principalmente aquelas que muitas vezes envolvem essas duas características, como a Cardiologia, que no estudo realizado por Montori et al. Mostrou-se que são dessa área 1/4 dos estudos encerrados precocemente identificados.<sup>84</sup>

## 4 MATERIAIS E MÉTODOS

### 4.1 DESENHO DE ESTUDO

Este trabalho representou um estudo de perfil metacientífico, utilizando, pois, outros artigos científicos como objeto de estudo. Como se objetivou nele a descrição da prevalência de ensaios clínicos truncados, tratou-se de um estudo descritivo. Tal trabalho compõe um braço do projeto maior da linha de Metaciência do grupo de pesquisa Registro de Síndromes Coronarianas Agudas (RESCA), intitulado “Descrição e Análise do Uso Excessivo e Inapropriado de Informações Científicas em Cardiologia: um estudo metacientífico”.

### 4.2 CRITÉRIOS DE ELEGIBILIDADE

Foram incluídos no projeto geral aqueles estudos que abordam tratamentos cardiológicos em humanos, independentemente do desenho de estudo, e que foram publicados entre janeiro de 2019 a dezembro de 2020 nas seguintes revistas: New England Journal of Medicine (NEJM), Journal of the American Association (JAMA), Circulation, European Heart Journal (EHJ) and Journal of the American College of Cardiology (JACC). Foram excluídos no projeto geral os estudos cujo objetivo primário não esteja relacionado à avaliação empírica de um tratamento médico, a exemplo dos seguintes: “View point”, “Research letter”, “From the heart”, “Editorial”, “Invited commentary”, “review”, “JAMA Cardiology Clinical Challenge”, “Cardiovascular images”, “Comment and response”, “Editor’s note”, “Correction”, “JAMA Masthead”. Especificamente neste trabalho, foram incluídos apenas os ensaios clínicos randomizados. A aplicação dos critérios de elegibilidade foi feita através do programa Microsoft Excel 2016.

### 4.3 TAMANHO AMOSTRAL

Conforme calculado através do programa WinPepi (versão 11.65), estimou-se que seriam necessários 139 artigos para calcular a prevalência de ensaios clínicos truncados com  $\pm 5\%$  de precisão e um alfa de 5%, em se considerando uma prevalência estimada de 10%.

#### 4.4 ESTRATÉGIA DE SELEÇÃO

Inicialmente foi realizada a identificação dos estudos publicados nos anos de 2019 e 2020 nas cinco revistas de medicina geral e de cardiologia de maior impacto, sendo elas: JAMA, NEJM, *Circulation*, EHJ e JACC

O website de cada uma das revistas científicas foi localizado através da plataforma de busca Google, sendo a coleta realizada diretamente a partir deles por dois avaliadores de forma independente. Inicialmente foram identificados todos os artigos originais publicados em cada revista de janeiro de 2019 a dezembro de 2020 e essa seleção foi realizada mediante a leitura de título e resumo de cada estudo publicado no período pré-definido. Os dados foram armazenados em planilhas independentes no programa Microsoft Excel 2016, através da coleta dois avaliadores também independentes.

#### 4.5 EXTRAÇÃO DE DADOS

Uma vez já tendo sido selecionados os artigos que compõem a amostra, foram coletadas as variáveis relacionadas às características do estudo a partir dos próprios artigos e de seus protocolos publicados no site <http://clinicaltrials.gov>, ou ainda, em outra plataforma de registros que esteja especificada no corpo do artigo. Nos casos em que o artigo retratara uma análise secundária, identificou-se a publicação da análise primária para a obtenção dos valores de algumas variáveis referentes a ela. Analogamente à seleção, a extração dos dados fora realizada por pares de avaliadores independentes que, à divergência, discutiram entre si suas respostas; caso a divergência tenha perdurado mesmo após a discussão, um terceiro avaliador foi consultado.

As variáveis que foram coletadas e suas respectivas definições neste trabalho são: título do artigo (nominal); ano de publicação (numérico); país de origem (nominal): o país onde foi realizado o estudo, ou, caso tenha sido realizado em mais de um país, o país de atuação do primeiro autor; revista da publicação (nominal); local (0 - unicêntrico; 1 - multicêntrico) - relacionada à quantidade de instituições nas quais foram coletados os desfechos; presença de financiamento da indústria farmacêutica – a espécie de apoio que o estudo recebeu da indústria farmacêutica para a sua condução; custo do estudo (numérico): valor monetário necessário para a realização

do estudo; tipo de ensaio clínico – se o desenho do estudo analisou superioridade ou não-inferioridade entre os grupos; presença de cálculo do tamanho amostral (0 - não; 1 - sim): se o artigo ou seu protocolo evidencia os valores relacionados ao cálculo do tamanho amostral, como o poder estatístico, magnitude de efeito e desvio padrão utilizados; tamanho amostral obtido (numérico) – quantidade de participantes sem levar em conta a perda de seguimento; alcance do tamanho amostral planejado (0 - não; 1 - sim): se o estudo alcançou o número de participantes que ele planejou, considerando inclusive a perda de follow-up prevista; positividade do estudo em relação ao desfecho primário (0 - não; 1 - sim) – vista a partir da aceitação da hipótese alternativa em relação a ao menos um desfecho primário; interrupção do estudo (0 - não, 1 - truncamento, 2 - por futilidade, 3 – por segurança) - processo pelo qual o estudo é encerrado previamente, antes do planejado, com base na análise interina do ECR, se tratando de truncamento quando ocorre por encontrar benefício, por futilidade quando a probabilidade de um resultado estatisticamente significativo ao fim do estudo for baixa, e por segurança quando for avaliado que o estudo não é seguro para os pacientes.

#### 4.6 PLANO DE ANÁLISE

A proporção (%) e o intervalo de confiança do desfecho primário foram calculados. As variáveis independentes categóricas foram descritas utilizando proporção, ao passo que as variáveis numéricas foram expressas através de média e desvio padrão, ao se apresentarem em uma distribuição normal, e por meio de medianas e intervalos interquartis, ao se apresentarem em distribuição não normal. A normalidade das variáveis foi avaliada pelo teste de Kolmogorov-Smirnov. A concordância entre os avaliadores independentes foi calculada através do coeficiente de Kappa. Valores de Kappa menores que 0,4 foram considerados concordância pobre, valores entre 0,4 e 0,75, moderada, e maiores que 0,75, excelente. Inicialmente, a força da associação entre o truncamento e as variáveis referentes às características do estudo foi averiguada por meio do teste  $\chi^2$  e do teste T de *Student*, para variáveis categóricas e numéricas, respectivamente. Em seguida, realizou-se uma regressão logística de modelo exploratório, identificando o *odds ratio* (OR) e o intervalo de confiança de 95% do OR das variáveis consideradas preditoras independentes de truncamento. Os valores de  $p < 0,20$  e  $p < 0,05$  foram empregados para determinar, respectivamente,

se os resultados da análise univariada e da multivariada foram estatisticamente diferentes daqueles que seriam esperados caso a hipótese nula estivesse correta. Para cada uma das análises, foram aplicados testes bicaudais, e elas foram realizadas a partir do programa SPSS (Versão 25, SPSS Inc., Chicago, Illinois, USA), enquanto os dados foram armazenados no programa Microsoft Excel 2016.

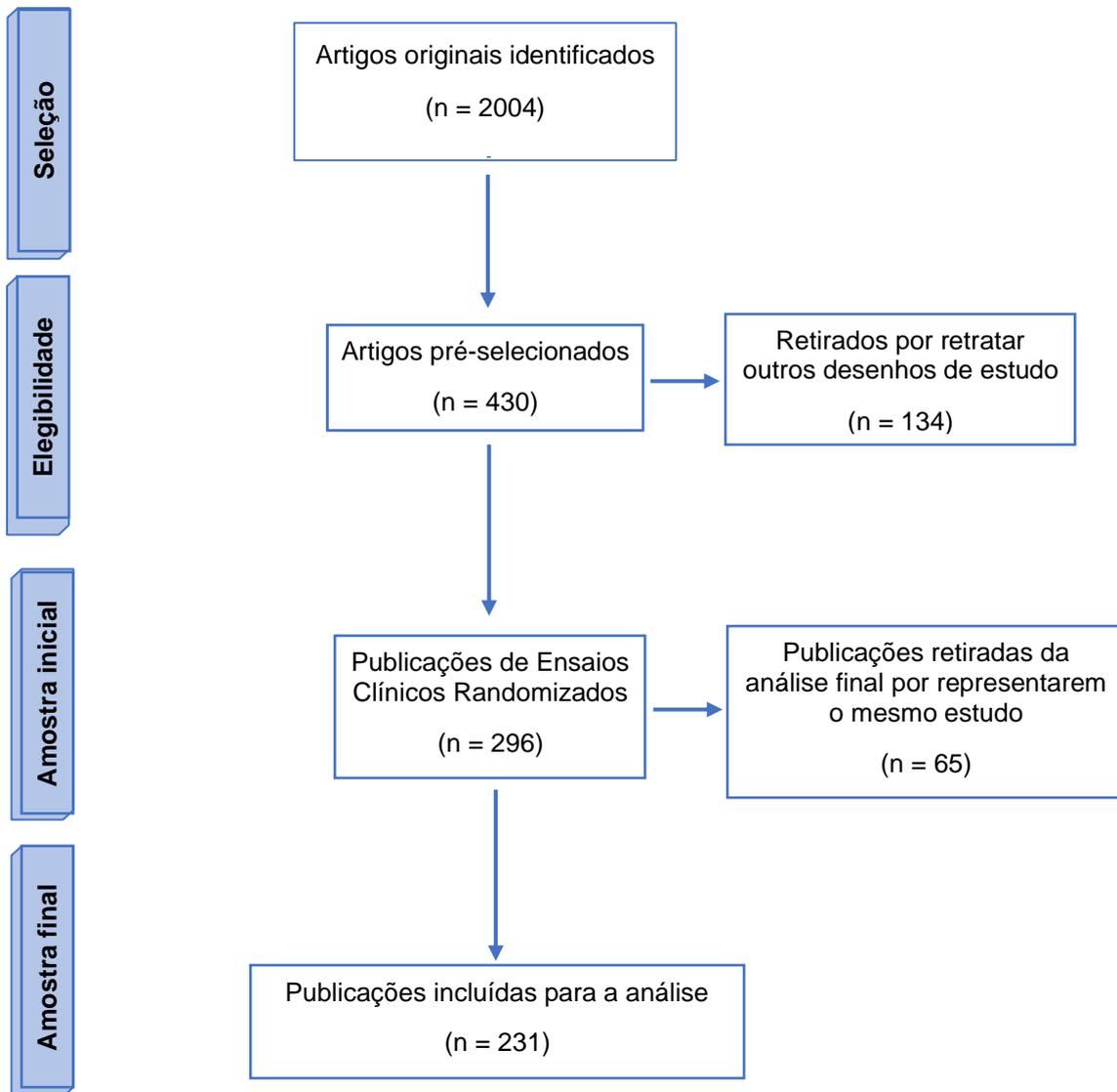
#### 4.7 SITUAÇÃO ÉTICA DO TRABALHO

Pela resolução nº 466/12, de 12 de dezembro de 2012 do Conselho Nacional de Saúde, pesquisas que utilizam dados de domínio público não necessitam ser registradas ou avaliadas pelo Comitê de Ética em Pesquisa (CEP) ou o Comitê Nacional de Ética em Pesquisa. Assim, não foi necessária a submissão do presente estudo ao CEP.

### 5 RESULTADOS

Foram selecionados inicialmente 2004 artigos para serem submetidos aos critérios de inclusão, dentre os quais 25,1% (503) foram publicados na *Circulation*; 26,1% (524) na JACC; 21,6% (432) na NEJM; 15,9% (318) na EHJ; e 11,3% (227) na JAMA. Desses artigos, 430 foram pré-selecionados por se tratar de estudos que abordam tratamentos médicos em humanos, nos quais foi realizada a coleta das variáveis. Foram retirados 134 por representarem estudos observacionais ou ensaios clínicos não randomizados, permanecendo assim 296 artigos sobre ensaios clínicos randomizados. Finalmente, apenas 231 estudos foram utilizados nas análises univariada e multivariada, pois 65 artigos foram retirados para evitar que o mesmo estudo fosse contabilizado mais de uma vez, já que no período observado alguns estudos realizaram mais de uma publicação. (**Fluxograma 1**).

Fluxograma 1 – Critérios de elegibilidade.



Fonte: dados dos próprios autores.

Considerando os 296 artigos sobre ensaios clínicos randomizados, 32,4% (96) foram publicados na *Circulation*; 25,3% (75) na *JACC*; 17,6% (52) na *NEJM*; 13,2% (39) na *EHJ*; e 11,5% (34) na *JAMA*. A maioria das publicações advinham de regiões específicas do mundo, sendo que 42,2% (125) das publicações são originárias dos Estados Unidos da América; 7,8% (23) da Alemanha; 7,1% (21) do Reino Unido; e 6,8% (20) dos Países Baixos. Além disso, os anos de 2019 e de 2020 apresentaram uma quantidade semelhante de artigos sobre ECRs, mostrando, respectivamente, 48,3% (143) e 51,7% (153) das publicações. Assim, após conferir quantos estudos os

296 artigos abordavam, nossa amostra foi de 231 ensaios clínicos randomizados dentre os quais 30,3% (70) foram publicados na *Circulation*; 22,1% (51) na JACC; 22,5% (52) na NEJM; 13% (30) na EHJ; e 12,1% (28) na JAMA, mantendo proporções semelhantes à amostra pré-selecionada e a anteriormente apresentada. Além disso, os anos de 2019 e de 2020 apresentaram, respectivamente, 48,5% (112) e 51,5% (119) das publicações, mantendo também a semelhança. (**Tabela 1**).

Tabela 1 – Características gerais da amostra de ensaios clínicos sobre terapêutica em cardiologia nas revistas de alto fator de impacto nos anos de 2019 e 2020.

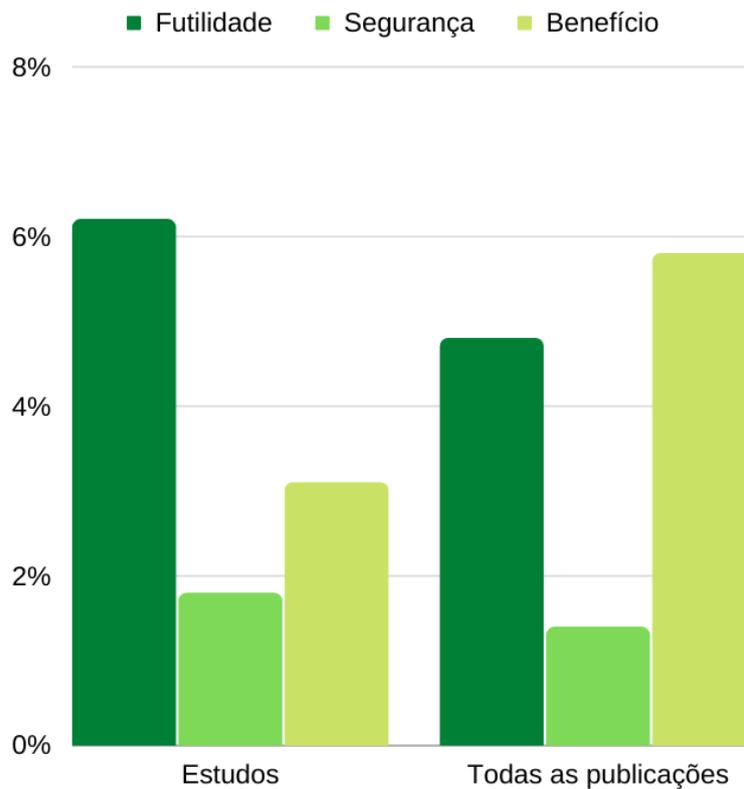
Variável	Amostra
Publicados em	
2019	112 (48,5%)
2020	119 (51,5%)
Revista de Publicação	
<i>Circulation</i>	70 (30,3%)
JACC	51 (22,1%)
NEJM	52 (22,5%)
EHJ	30 (13,0%)
JAMA	28 (12,1%)

Fonte: dados dos próprios autores. As variáveis foram expressas em números absolutos (porcentagem).

A partir de todos os artigos incluídos no presente trabalho, foi possível avaliar a interrupção e o seu respectivo motivo em 226 deles, nos quais se observou que 88,9% (201) dos ensaios clínicos não sofreram qualquer tipo de interrupção, ao passo que 6,2% (14) foram interrompidos durante a análise interina devido a testes de futilidade, 1,8% (4) por questões de segurança, e 3,1% (7) foram interrompidos devido ao benefício encontrado, ou seja, a prevalência de ECRs truncados foi de 3,1%. Essa prevalência apresentou um intervalo de confiança de 95% que se estendeu de 1,25% a 6,28%. A relação da proporção dos tipos de estudo quanto à interrupção está disposta na **Gráfico 1**. A ocorrência de interrupção e o motivo pôde ser descrito 291 artigos, em que 88% (256) abordaram estudos que não foram encerrados prematuramente, enquanto 4,8% (14) foram interrompidos devido à futilidade, 1,4%

(4) pela segurança dos pacientes, e 5,8% (17) foram truncados, com um intervalo de confiança de 95% de 3,2% a 8,5%.

Gráfico 1 – Proporção entre as diferentes interrupções em nossa amostra de estudos e de publicações.



Fonte: dados dos próprios autores.

As associações medidas através do teste  $X^2$  não podem ser utilizadas para inferência estatística, uma vez que não se alcançou a premissa básica do teste de possuir menos de 20% das caselas com um valor esperado inferior a 5. Juntamente a isso, o teste T de Student também não apresentou uma diferença com veracidade estatística ( $p = 0,11$ ) entre as médias do tamanho amostral obtido pelos ECRs truncados e não truncados, 11642 ( $\pm 12010$ ) e 4294 ( $\pm 7016$ ), respectivamente. Os resultados dos testes univariados supracitados estão dispostos na **tabela 2**.

Tabela 2 – Associação de truncamento com características que possuem plausibilidade de serem preditoras.

Característica	Ensaio Clínico Truncado	Ensaio Clínico Não Truncado	Teste de hipótese
Local			$p = 0,15$
<i>Unicêntrico</i>	2 (28,6%)	24 (11,0%)	
<i>Multicêntrico</i>	5 (71,4%)	195 (89,0%)	
Relação com a indústria farmacêutica			$p = 0,24$
<b>Sem relação</b>	1 (14,3%)	60 (27,4%)	
<i>Financiado</i>	3 (42,9%)	107 (48,9%)	
<i>Conduzido</i>	3 (42,9%)	52 (23,7%)	
Análise de			$p = 0,74$
<i>superioridade</i>	6 (85,7%)	196 (79,7%)	
<i>não-inferioridade</i>	1 (14,3%)	26 (12,3%)	
<i>ambas</i>	0	17 (8,0%)	
Presença de cálculo do tamanho amostral	7 (100%)	184 (92,5%)	$p = 0,45$
Tamanho amostral obtido	7141 ( $\pm$ 9426)	4236 ( $\pm$ 6984)	$p = 0,29$
Alcance do tamanho amostral planejado	90% ( $\pm$ 16,5%)	96% ( $\pm$ 10,2%)	$p = 0,37$
Positividade em relação ao desfecho primário	4 (66,7%)	138 (63,6%)	$p = 0,88$

Fonte: dados dos próprios autores. As variáveis qualitativas foram expressas em números absolutos (porcentagem), e as quantitativas em média ( $\pm$  desvio padrão).

Mediante o não cumprimento dos critérios de adequação da regressão logística e seus resultados não estatisticamente significantes, a inferência a partir dela se faz inapropriada.

## 6 DISCUSSÃO DOS RESULTADOS

O presente estudo indica que nos últimos anos uma pequena proporção dos ensaios clínicos randomizados publicados em revistas de alto impacto são truncados. Não obstante, se essa pequena proporção atual representa um progresso ou um retrocesso da comunidade científica é algo que requer discussão. Neste trabalho, optamos por basear nossa amostra nas cinco revistas de medicina geral e de cardiologia de maior fator de impacto, com o intuito de adquirir uma amostra de

conveniência dos estudos mais relevantes para a comunidade médica e científica, em suas pesquisas, recomendações práticas e condutas clínicas.

Como posto anteriormente, essa prevalência numericamente pequena pode ter significados um tanto distintos. Inicialmente de que não é uma prática tão frequente no ecossistema científico, o que é pertinente considerando os potenciais impactos da publicação desse tipo de estudo citados anteriormente. Para mais, esse valor também pode refletir que a prevalência de estudos truncados na literatura continua a crescer em conformidade com trabalhos anteriores, ou que finalmente foi refreada, estagnando ou, até mesmo, encolhendo.

Observa-se que de um modo geral a literatura vinha apresentando um aumento progressivo na quantidade absoluta <sup>8</sup> e na prevalência de estudos truncados <sup>9</sup>, até meados de 2007. Apesar de não termos conhecimento de pesquisas que investigassem o aumento da prevalência de estudos truncados posteriormente a esse período até o presente momento, pode-se observar a prevalência descrita em diferentes épocas e estudos na literatura de alto impacto, como a de 1,2% entre 2000 e 2004 <sup>9</sup>, e a de 1,7% (95%IC, 0,8 – 3,3) dentre os ECRs publicados em qualquer revista entre os anos de 2000 e 2013.<sup>39</sup> Neste, quase todos os ECRs foram publicados após o ano de 2003, uma vez que ele os selecionou a partir dos protocolos publicados entre 2000 e 2003.<sup>39</sup> Juntamente ao nosso resultado de 3,1% (95% IC, 1,25% – 6,28%), pode-se sugerir que a prevalência de estudos truncados provavelmente continua a crescer na literatura, apesar de toda a discussão sobre o tema na comunidade científica. Já em relação ao local onde se coleta a prevalência também pode se perceber uma discrepância, havendo uma prevalência maior em revistas de alto impacto que quando consideradas todas as revistas na mesma época, como, respectivamente, 1,2% e 0,1% no período entre 2000 e 2004 <sup>9</sup>, além de uma prevalência de 1% (95%IC, 0,5 – 2,0) contra uma de 1,7% (95%IC, 0,8 – 3,3), quando a prevalência foi avaliada dentre todos os trabalhos desenvolvidos pela comunidade científica ou dentre apenas os que foram publicados.<sup>39</sup> Isso indica um viés de publicação relacionado ao truncamento, presente principalmente em revistas de alto impacto. Afinal, metade dos estudos interrompidos por futilidade, recrutamento lento e problemas de segurança foram publicados, ao passo que todos os truncados foram publicados na amostra de Kasenda et al. <sup>39</sup>, e os truncados eram proporcionalmente mais publicados em revistas de alto impacto que outros estudos que exploravam a

mesma questão, mas que não foram interrompidos (68% vs 30%,  $p < 0,001$ ).<sup>16</sup> Alguns indícios disso também se manifestam em nossa amostra, pois pode-se perceber que diferentemente dos interrompidos por futilidade e por segurança, que tiveram apenas uma publicação, alguns truncados tiveram mais publicações. Aliás algo há muito tempo demonstrado é que resultados positivos influenciam diretamente na probabilidade de um estudo ser submetido à publicação e também de ser publicado, fenômeno chamado de viés de publicação.<sup>86</sup>

Especificamente no meio dos ECRs, os positivos têm 4 vezes mais chance de serem publicados que um ensaio clínico negativo, além de os positivos tenderem a publicar os resultados em 4 ou 5 anos, enquanto os negativos demoram de 6 a 8 anos geralmente.<sup>87</sup> Considerando a existência do viés de publicação, em que estudos negativos são menos frequentemente submetidos a uma revista, e menos frequentemente ainda aceitos para a publicação, estudos positivos se tornam quase a “regra”.<sup>86</sup> Assim, os pesquisadores são, inclusive, mais incentivados a ter resultados positivos que necessariamente verdadeiros, precisos e confiáveis. Para isso, o truncamento pode ser uma excelente ferramenta.

Tendo tudo isso em vista, tais estudos apresentam menos probabilidade de o benefício demonstrado refletir o efeito real do tratamento, ou seja, contêm menos valor preditivo positivo. Soma-se aos pontos recém descritos que quanto menor o tamanho amostral, menor o valor preditivo positivo.<sup>25,88</sup> Em consonância, ECRs com baixo poder estatístico possuem um valor preditivo positivo normalmente inferior a 50%, mesmo quando bem conduzidos e com boa probabilidade prévia de que o tratamento funcione, ou seja, é quase tão provável que o resultado positivo seja verdadeiro, como é provável que seja falso.<sup>88</sup> Ainda que o poder estatístico esteja adequado, a presença de vieses no estudo ou na literatura da qual ele faz parte, como o viés de publicação, e uma baixa probabilidade prévia de que o tratamento fosse eficaz levam a um baixo valor preditivo positivo.<sup>88</sup>

Essa prevalência do truncamento que se mantém crescente, provavelmente tem relação com a crença na lei dos pequenos números, que é a convicção de que pequenas amostras terão características essenciais semelhantes a amostras maiores, sendo representativas de toda a população.<sup>89</sup> Assim, há exagerada confiança em resultados originados de amostras menores, superestimando o efeito e o poder e subestimando a amplitude de intervalos de confiança<sup>89</sup>, o que é visto reiteradamente

nesse tipo de estudo.<sup>16</sup> Juntamente a isso, o viés de confirmação pode estar manifestando grande influência na predileção pelo uso das regras de parada. Afinal, os investimentos financeiros e temporais relacionados a um estudo somente são realizados quando o patrocinador e os investigadores acreditam que o tratamento possa ser benéfico, o que implica em uma tendência maior de ver resultados interinos promissores como definitivos, pois são interpretados sob a visão de alguém que já esperava que o benefício fosse revelado.<sup>90</sup> Entretanto, nessa situação, o próprio princípio da *Equipose* estaria violado, sendo evidência de que os pesquisadores não foram capazes de colocar suas tendências pessoais de lado.<sup>2</sup>

Como os truncados possuem uma maior probabilidade de apresentar resultados falsos, eles podem guiar erroneamente as mãos de muitos profissionais que acabem por basear suas atitudes neles, inclusive após evidências contrárias, como visto no caso do bisoprolol <sup>54</sup>, o que pode ser consequência do fenômeno da dissonância reduzida. Nesse fenômeno, há uma espécie de apego a escolhas realizadas no passado, tanto no campo material quanto no campo das ideias, em que se realça as virtudes e se minimiza os defeitos daquilo que já foi incorporado, por exemplo, realçando o benefício descrito e minimizando os vieses e efeitos colaterais.<sup>91</sup> A partir desse processo, torna-se difícil ponderar e julgar as evidências contrárias existentes, mantendo resultados menos prováveis de serem verdadeiros como referência em determinada área, atrasando a possível correção desse engano.<sup>90,92</sup>

Ademais, em nossa amostra 6,2% foram interrompidos por futilidade, 3,1% por benefício e 1,8% por segurança. Em outros estudos sobre a publicação de ECRs independentemente do fator de impacto da revista, foi observada uma prevalência de estudos encerrados por futilidade de 11%, por segurança de 2,3% e de 1,7% por benefício.<sup>39</sup> Essa diferença entre os dois estudos pode ser uma variação aleatória, um indicativo de que os truncados estão sendo mais publicados ultimamente, ou ainda, que são mais publicados em revistas de alto impacto, ao passo que a prevalência de estudos publicados após serem encerrados por futilidade ou segurança reduziu e a de interrompidos por benefício aumentou.

Já as associações avaliadas em nosso estudo, que não serviram para uma adequada inferência estatística, também foram testadas em outros estudos, como o financiamento e o alcance do tamanho amostral planejado. Foi reportado por Montori et al. em 2005 que os truncados apresentaram uma proporção maior de estudos

financiados por organizações com fins lucrativos e também por organizações sem fins lucrativos que os estudos não interrompidos.<sup>9</sup> Além disso, a proporção de truncados financiados por instituições com fins lucrativos foi maior que a de truncados o financiamento de instituições sem fins lucrativos.<sup>9</sup> Entretanto, esses resultados não foram postos por esses pesquisadores em um teste de hipótese, nem apresentaram um intervalo de confiança que permitisse uma inferência estatística de que essas diferenças são reais. O alcance do tamanho amostral planejado no estudo de Montori et al. foi em média de pouco mais da metade em relação ao planejado.<sup>9</sup> Esse resultado é evidência de como eles costumam ser menores e, portanto, mais enviesados, fazendo-se necessário investigar futuramente se essa característica se mantém nos ECRs truncados mais recentes.

Nosso estudo deve ser visto à luz de sua vantagem em avaliar os estudos truncados onde eles podem manifestar uma maior influência e se fazem mais presentes, isto é, na literatura de alto impacto da área de cardiologia; bem como em apresentar uma amostra recente, o que o distingue dos outros estudos que descreveram a prevalência de truncados, que avaliaram longos períodos, concentrando-se principalmente no início do século. Em contrapartida, o nosso trabalho teve algumas limitações: o custo do estudo não foi encontrado nas revistas, nos artigos, ou mesmo em seus protocolos; o desfecho primário avaliado quanto à positividade foi o do artigo presente em nossa amostra de conveniência, não sendo necessariamente o da análise primária, responsável por conduzir o ECR ao truncamento; e poucos ECRs truncados foram identificados na amostra de conveniência escolhida, o que impediu a inferência estatística de quais as características de um estudo que podem estar relacionadas ao truncamento.

## **7 CONCLUSÃO**

A prevalência de ensaios clínicos randomizados truncados nas revistas cardiológicas de alto impacto é baixa.

## **REFERÊNCIAS**

1. Rawlins M. De testimonio: on the evidence for decisions about the use of therapeutic interventions. *Lancet* [Internet]. 2008;372(9656):2152–61.

- Disponível em: [http://dx.doi.org/10.1016/S0140-6736\(08\)61930-3](http://dx.doi.org/10.1016/S0140-6736(08)61930-3).
2. Freedman B. Equipoise and the Ethics of Clinical Research. *N Engl J Med* [Internet]. 1987 Jul 16;317(3):141–5. Disponível em: <http://www.nejm.org/doi/abs/10.1056/NEJM198707163170304>.
  3. Casazza G, Casella F. Can we trust in trials stopped early for benefit? *Intern Emerg Med*. 2012;7(6):559–61.
  4. Guyatt GH, Briel M, Glasziou P, Bassler D, Montori VM. Problems of stopping trials early. *BMJ*. 2012;344(7865):9–11.
  5. Walter SD, Guyatt GH, Bassler D, Briel M, Ramsay T, Han HD. Randomised trials with provision for early stopping for benefit (or harm): The impact on the estimated treatment effect. *Stat Med*. 2019;38(14):2524–43.
  6. Walter SD, Han H, Briel M, Guyatt GH. Quantifying the bias in the estimated treatment effect in randomized trials having interim analyses and a rule for early stopping for futility. *Stat Med*. 2017;36(9):1506–18.
  7. Prutsky GJ, Domecq JP, Erwin PJ, Briel M, Montori VM, Akl EA, et al. Initiation and continuation of randomized trials after the publication of a trial stopped early for benefit asking the same study question: STOPIT-3 study design. *Trials*. 2013;14(1):2–9.
  8. Trotta F, Apolone G, Garattini S, Tafuri G. Stopping a trial early in oncology: For patients or for industry? *Ann Oncol*. 2008;19(7):1347–53.
  9. Montori VM, Devereaux PJ, Adhikari NKJ, Burns KEA, Eggert CH, Briel M, et al. Randomized Trials Stopped Early for Benefit. *Jama*. 2005;294(17):2203.
  10. Pocock SJ. When (Not) to Stop a Clinical Trial for Benefit. *JAMA* [Internet]. 2005 Nov 2;294(17):2228. Disponível em: <http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.294.17.2228>.
  11. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *Int J Surg*. 2011;9(8):672–7.
  12. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence - Study limitations (risk of bias). *J Clin Epidemiol*. 2011;64(4):407–15.
  13. Pocock SJ, Hughes MD. Practical problems in interim analyses, with particular regard to estimation. *Control Clin Trials*. 1989;10(4 SUPPL. 1):209–21.
  14. Wang Z. Randomised trials stopped early for benefit: How often should they occur? In: Heard JLA, editor. *Population Health Congress 2008*. Brisbane; 2008. p. 56–60.
  15. Wang H, Rosner GL, Goodman SN. Quantifying over-estimation in early stopped clinical trials and the “freezing effect” on subsequent research. *Clin Trials*. 2016;13(6):621–31.
  16. Bassler D, Briel M, Montori VM. Stopping Randomized Trials Early for Benefit and Estimation of Treatment Effects: Systematic Review and Meta-regression Analysis. *JAMA*. 2010 Mar 24;303(12):1180.

17. Pocock S, White I. Trials stopped early: Too good to be true? *Lancet*. 1999;353(9157):943–4.
18. Fan X, DeMets DL, Lan KKG. Conditional bias of point estimates following a group sequential test. *J Biopharm Stat*. 2004;14(2):505–30.
19. Goodman SN. Stopping trials for efficacy: An almost unbiased view. *Clin Trials*. 2009;6(2):133–5.
20. Demets DL, Lan KKG. Interim analysis: The alpha spending function approach. *Stat Med*. 1994 Jul 15;13(13–14):1341–52.
21. Korn EL, Freidlin B, Mooney M. Stopping or Reporting Early for Positive Results in Randomized Clinical Trials: The National Cancer Institute Cooperative Group Experience From 1990 to 2005. *J Clin Oncol* [Internet]. 2009 Abr 1;27(10):1712–21. Disponível em: <http://ascopubs.org/doi/10.1200/JCO.2008.19.5339>.
22. Ioannidis JPA. Contradicted and Initially Stronger Effects in Highly Cited Clinical Research. *JAMA* [Internet]. 2005 Jul 13;294(2):218. Disponível em: <http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.294.2.218>.
23. Freidlin B, Korn EL. Stopping clinical trials early for benefit: Impact on estimation. *Clin Trials*. 2009;6(2):119–25.
24. Murad MH, Guyatt GH, Domecq JP, Vernooij RWM, Erwin PJ, Meerpohl JJ, et al. Randomized trials addressing a similar question are commonly published after a trial stopped early for benefit. *J Clin Epidemiol* [Internet]. 2017;82:12–9. Disponível em: <http://dx.doi.org/10.1016/j.jclinepi.2016.10.006>.
25. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, et al. Power failure: Why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci*. 2013;14(5):365–76.
26. Zöllner S, Pritchard JK. Overcoming the Winner’s Curse: Estimating Penetrance Parameters from Case-Control Data. *Am J Hum Genet* [Internet]. 2007 Abr;80(4):605–15. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0002929707610970>.
27. Greenwald AG. Consequences of prejudice against the null hypothesis. *Psychol Bull*. 1975;82(1):1–20.
28. Walsh M, Srinathan SK, McAuley DF, Mrkobrada M, Levine O, Ribic C, et al. The statistical significance of randomized controlled trial results is frequently fragile: a case for a Fragility Index. *J Clin Epidemiol* [Internet]. 2014 Jun;67(6):622–8. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0895435613004666>.
29. The Lancet. Stopping trials early for benefit: too good to be true. *Lancet* [Internet]. 2008 Abr;371(9621):1310. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0140673608605693>.
30. Stegert M, Kasenda B, Von Elm E, You JJ, Blümle A, Tomonaga Y, et al. An analysis of protocols and publications suggested that most discontinuations of clinical trials were not based on preplanned interim analyses or stopping rules. *J Clin Epidemiol* [Internet]. 2016;69:152–60. Disponível em:

- <http://dx.doi.org/10.1016/j.jclinepi.2015.05.023>.
31. Atkins CD. Stopping trials early for positive results: The need to know how much. *J Clin Oncol*. 2009;27(21):26–7.
  32. Guyatt GH. Bias and Trials Stopped Early for Benefit—Reply. *JAMA* [Internet]. 2010 Jul 14;304(2):156. Disponível em: <http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2010.933>.
  33. Fisher B, Costantino JP, Wickerham DL, Redmond CK, Kavanah M, Cronin WM, et al. Tamoxifen for Prevention of Breast Cancer: Report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *JNCI J Natl Cancer Inst* [Internet]. 1998 Set 16;90(18):1371–88. Disponível em: <https://academic.oup.com/jnci/article/90/18/1371/897928>.
  34. Fisher B, Costantino JP, Wickerham DL, Cecchini RS, Cronin WM, Robidoux A, et al. Tamoxifen for the Prevention of Breast Cancer: Current Status of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *JNCI J Natl Cancer Inst* [Internet]. 2005 Nov 16;97(22):1652–62. Disponível em: <http://academic.oup.com/jnci/article/97/22/1652/2521468/Tamoxifen-for-the-Prevention-of-Breast-Cancer>.
  35. Briel M, Speich B, Von Elm E, Gloy V. Comparison of randomized controlled trials discontinued or revised for poor recruitment and completed trials with the same research question: A matched qualitative study. *Trials*. 2019;20(1).
  36. Bassler D, Ferreira-Gonzalez I, Briel M, Cook DJ, Devereaux PJ, Heels-Ansdell D, et al. Author reply: the master-question: meta-analyses and stopped-early randomized trials—a problem? *J Clin Epidemiol* [Internet]. 2008 Jan;61(1):96–8. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0895435607002971>.
  37. Briel M, Bassler D, Wang AT, Guyatt GH, Montori VM. The dangers of stopping a trial too early. *J Bone Jt Surg - Ser A*. 2012;94(SUPPL. 1):56–60.
  38. Leandro G. How to interpret a randomized controlled study stopped early. *Intensive Care Med*. 2013;39(9):1642–3.
  39. Kasenda B, Von Elm E, You J, Blümle A, Tomonaga Y, Saccilotto R, et al. Prevalence, characteristics, and publication of discontinued randomized trials. *JAMA - J Am Med Assoc*. 2014;311(10):1045–51.
  40. Pashler H, Harris CR. Is the Replicability Crisis Overblown? Three Arguments Examined. *Perspect Psychol Sci*. 2012;7(6):531–6.
  41. Hughes MD, Freedman LS, Pocock SJ. The Impact of Stopping Rules on Heterogeneity of Results in Overviews of Clinical Trials. *Biometrics*. 1992;48(1):41.
  42. Marschner IC, Schou IM. Underestimation of treatment effects in sequentially monitored clinical trials that did not stop early for benefit. *Stat Methods Med Res*. 2019;28(10–11):3027–41.
  43. Marschner IC, Askie LM, Schou IM. Sensitivity analyses assessing the impact of early stopping on systematic reviews: Recommendations for interpreting guidelines. *Res Synth Methods*. 2020;11(2):287–300.

44. Schou IM, Marschner IC. Meta-analysis of clinical trials with early stopping: An investigation of potential bias. *Stat Med*. 2013;32(28):4859–74.
45. Bassler D, Ferreira-Gonzalez I, Briel M, Cook DJ, Devereaux PJ, Heels-Ansdell D, et al. Systematic reviewers neglect bias that results from trials stopped early for benefit. *J Clin Epidemiol*. 2007;60(9):869–73.
46. Goodman S. Bias and Trials Stopped Early for Benefit. *JAMA [Internet]*. 2010 Jul 14;304(2):156. Disponível em: <http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2010.931>.
47. Iltis AS. Stopping trials early for commercial reasons: The risk-benefit relationship as a moral compass. *J Med Ethics*. 2005;31(7):410–4.
48. Mueller PS, Montori VM, Bassler D, Koenig BA, Guyatt GH. Ethical Issues in Stopping Randomized Trials Early Because of Apparent Benefit. *Ann Intern Med [Internet]*. 2007 Jun 19;146(12):878. Disponível em: <http://annals.org/article.aspx?doi=10.7326/0003-4819-146-12-200706190-00009>.
49. Poldermans D, Boersma E, Bax JJ, Thomson IR, van de Ven LLM, Blankensteijn JD, et al. The Effect of Bisoprolol on Perioperative Mortality and Myocardial Infarction in High-Risk Patients Undergoing Vascular Surgery. *N Engl J Med [Internet]*. 1999 Dez 9;341(24):1789–94. Disponível em: <http://www.nejm.org/doi/abs/10.1056/NEJM199912093412402>.
50. Eagle KA, Berger PB, Calkins H, Chaitman BR, Ewy GA, Fleischmann KE, et al. ACC/AHA guideline update for perioperative cardiovascular evaluation for noncardiac surgery - Executive summary: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Update the 1996 Guid. *J Am Coll Cardiol [Internet]*. 2002;39(3):542–53. Disponível em: [http://dx.doi.org/10.1016/S0735-1097\(01\)01788-0](http://dx.doi.org/10.1016/S0735-1097(01)01788-0).
51. Poldermans D, Bax JJ, Boersma E, De Hert S, Eeckhout E, Fowkes G, et al. Guidelines for pre-operative cardiac risk assessment and perioperative cardiac management in non-cardiac surgery. *Eur Heart J*. 2009;30(22):2769–812.
52. Feitosa ACR, Marques AC, Caramelli B, Ayub B, Polanczyk CA, Jardim C, et al. II Diretriz de Avaliação Perioperatória da Sociedade Brasileira de Cardiologia. *Arq Bras Cardiol [Internet]*. 2011;96(3):1–68. Disponível em: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0066-782X2011000800001&lng=pt&nrm=iso&tlng=pt](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0066-782X2011000800001&lng=pt&nrm=iso&tlng=pt).
53. Gualandro D, Yu P, Caramelli B, Marques A, Fornari L, Pinho C, et al. 3ª Diretriz de Avaliação Cardiovascular Perioperatória da Sociedade Brasileira de Cardiologia. *Arq Bras Cardiol*. 2017;109(3-supl.1):1–115.
54. Bouri S, Shun-Shin MJ, Cole GD, Mayet J, Francis DP. Meta-analysis of secure randomised controlled trials of  $\beta$ -blockade to prevent perioperative death in non-cardiac surgery. *Heart*. 2014;100(6):456–64.
55. Garner S, Docherty M, Somner J, Sharma T, Choudhury M, Clarke M, et al. Reducing ineffective practice: Challenges in identifying low-value health care using cochrane systematic reviews. *J Heal Serv Res Policy*. 2013;18(1):6–12.

56. Moynihan R, Smith R. Too much medicine? Almost certainly. *Perspect Complement Altern Med A Read*. 2019;324(Abril):33–6.
57. Elshaug AG. Building the Evidence Base for Disinvestment From Ineffective Health Care Practices : 2007;(Outubro).
58. Levinson W, Kallewaard M, Bhatia RS, Wolfson D, Shortt S, Kerr EA. “Choosing Wisely”: A growing international campaign. *BMJ Qual Saf*. 2015;24(2):167–74.
59. Chassin MR. Is Health Care Ready for Six Sigma Quality? *Milbank Q*. 1998;76(4):565–91.
60. Mulley AG, Trimble C, Elwyn G. Stop the silent misdiagnosis: Patients’ preferences matter. *BMJ*. 2012;345(7883):6–11.
61. Institute of Medicine (US) Committee on Quality of Health Care in America. A new health system for the 21st century - Crossing the quality chasm. *Pharm J*. 2003;271(7266):340–1.
62. Morgan DJ, Wright SM, Dhruva S. Update on medical overuse. *JAMA Intern Med*. 2015;175(1):120–4.
63. Morgan DJ, Dhruva SS, Coon ER, Wright SM, Korenstein D. 2019 Update on Medical Overuse: A Review. *JAMA Intern Med*. 2019;179(11):1568–74.
64. Care M. Successes and Failures in the Implementation of Evidence-Based Guidelines for Clinical Practice. *Med Care*. 2001;39(8):II:46--II:54.
65. Smith M, Saunders R, Stuckhardt L, Mcginnis JM. Best Care at Lower Cost. *Best Care Low Cost*. 2013;(Cdc).
66. Cassel CK, Guest JA. Choosing wisely: Helping physicians and patients make smart decisions about their care. *JAMA - J Am Med Assoc*. 2012;307(17):1801–2.
67. Reports C. 44 Percent of Healthy Adults Getting Unneeded Heart Screenings. 2011;
68. Ellen ME, Wilson MG, Vélez M, Shach R, Lavis JN, Grimshaw JM, et al. Addressing overuse of health services in health systems: A critical interpretive synthesis. *Heal Res Policy Syst*. 2018;16(1):1–14.
69. Hardwicke TE, Serghiou S, Janiaud P, Danchev V, Crüwell S, Goodman SN, et al. Calibrating the scientific ecosystem through meta-research. *Annu Rev Stat Its Appl*. 2020;7:11–37.
70. Psaty BM, Rennie D. Stopping Medical Research to Save Money. *Jama*. 2003;289(16).
71. Jitlal M, Khan I, Lee SM, Hackshaw A. Stopping clinical trials early for futility: Retrospective analysis of several randomised clinical studies. *Br J Cancer*. 2012;107(6):910–7.
72. Pocock SJ, Hughes MD. Practical problems in interim analyses, with particular regard to estimation. *Control Clin Trials*. 1989;10(4 SUPPL. 1):209–21.

73. Wheatley K, Clayton D. Be skeptical about unexpected large apparent treatment effects: The case of an MRC AML12 randomization. *Control Clin Trials*. 2003;24(1):66–70.
74. Schulz KF, Grimes DA. Multiplicity in randomised trials II: Subgroup and interim analyses. *Lancet*. 2005;365(9471):1657–61.
75. Pocock S, White I. Trials stopped early: Too good to be true? *Lancet*. 1999;353(9157):943–4.
76. Jitlal M, Khan I, Lee SM, Hackshaw A. Stopping clinical trials early for futility: Retrospective analysis of several randomised clinical studies. *Br J Cancer*. 2012;107(6):910–7.
77. O'Brien PC, Fleming TR. A Multiple Testing Procedure for Clinical Trials. *Biometrics*. 1979;35(3):549.
78. Pocock SJ. Group sequential methods in the design and analysis of clinical trials. *Biometrika*. 1977;64(2):191–9.
79. Haybittle JL. Repeated assessment of results in clinical trials of cancer treatment. *Br J Radiol*. 1971;44(526):793–7.
80. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard S V., et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Br J Cancer*. 1976;34(6):585–612.
81. Lachin JM. A review of methods for futility stopping based on conditional power. *Stat Med*. 2005;24(18):2747–64.
82. Mehta CR, Cain KC. Charts for the early stopping of pilot studies. *J Clin Oncol*. 1984;2(6):676–82.
83. Hopf G. Early stopping of trials. Vol. 350, *Lancet*. 1997. p. 891.
84. Montori VM, Devereaux PJ, Adhikari NKJ, Burns KEA, Eggert CH, Briel M, et al. Randomized Trials Stopped Early for Benefit. *Jama*. 2005;294(17):2203.
85. Song F, Hooper L, Loke YK. Publication bias: What is it? How do we measure it? How do we avoid it? *Open Access J Clin Trials*. 2012;5(1):71–81.
86. Coursol A, Wagner EE. Effect of Positive Findings on Submission and Acceptance Rates. A Note on Meta-Analysis Bias. *Prof Psychol Res Pract*. 1986;17(2):136–7.
87. Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database Syst Rev*. 2009;(1).
88. Ioannidis JPA. Why Most Published Research Findings Are False. *PLoS Med* [Internet]. 2005 Ago 30;2(8):e124. Disponível em: <https://dx.plos.org/10.1371/journal.pmed.0020124>.
89. Tversky A, Kahneman D. Belief in the law of small numbers. *Psychol Bull*. 1971;76(2):105–10.
90. Raymond S. Nickerson. Confirmation bias: a ubiquitous phenomenon in many

- guises. *Rev Gen Psychol.* 1998;2(2):175–220.
91. Festinger L. Cognitive Dissonance. *Sci Am [Internet]*. 1962 Oct;207(4):93–106. Disponível em: <https://www.scientificamerican.com/article/cognitive-dissonance>.
  92. Ioannidis JPA. Why Science Is Not Necessarily Self-Correcting. *Perspect Psychol Sci.* 2012;7(6):645–54.