



**ESCOLA BAHIANA DE MEDICINA E SAÚDE PÚBLICA**  
**CURSO DE MEDICINA**

**LUÍSA MAYAN VENTIN COVRE**

**INVISIBILIDADE DO ERRO TIPO II: PERIÓDICOS DE ALTA RELEVÂNCIA  
CIENTÍFICA GARANTEM ALTA CONFIABILIDADE DA EVIDÊNCIA?**

**TRABALHO DE CONCLUSÃO DE CURSO**

**SALVADOR - BA**

**2023**

**LUÍSA MAYAN VENTIN COVRE**

**INVISIBILIDADE DO ERRO TIPO II: PERIÓDICOS DE ALTA RELEVÂNCIA  
CIENTÍFICA GARANTEM A ALTA CONFIABILIDADE DA EVIDÊNCIA?**

Projeto de pesquisa apresentado ao Curso de Graduação em Medicina da Escola Bahiana de Medicina e Saúde Pública como requisito parcial para aprovação no componente Metodologia da Pesquisa III (MPIII).

Orientador: Diego Ribeiro Rabelo

**SALVADOR**

**2023**

## **AGRADECIMENTOS**

O presente trabalho envolveu muitas pessoas, tanto direta quanto indiretamente, e a todos declaro meu profundo agradecimento.

Marcia, Alberto e Daniel, por sempre apoiarem meus anseios e fazerem de tudo para que eles se tornem reais. Pelo exemplo de persistência e força, bem como pela incessante busca pelo saber. Agradeço por todo amor do mundo.

Rebeca, agradeço pelas reflexões, pela companhia, pelo ombro e pelo cuidado de sempre. Obrigada por compartilhar os momentos de angústia, de incerteza, mas também os abraços apertados e as boas risadas.

Caio, pela tolerância dos meus momentos de ausência para permitir a realização dos meus sonhos e a construção da profissional que almejo ser. Pelo cuidado, pelo carinho e pela segunda família.

Diego, por ter me apresentado a um mundo fantástico que me despertou muita identificação, por compartilhar sua forma de pensar e por ajudar na construção do meu pensamento científico.

São poucas as pessoas que vêm marcando a minha trajetória acadêmica, mais uma vez obrigada por fazerem isso de forma tão especial e cuidadosa.

## RESUMO

**Introdução:** Os erros sistemáticos e aleatórios interferem negativamente na validade interna dos trabalhos, interferindo na precisão das conclusões. Assim, para guiar pesquisadores e leitores, guidelines de validade interna foram feitos para auxiliar na conduta das pesquisas e na leitura de artigos. Entretanto, esses checklists abordam exigências que envolvem quase que completamente orientações para evitar os erros sistemáticos, mas nenhuma para minimizar a presença de erros aleatórios diretamente. Isso pode culminar numa grande quantidade de artigos provavelmente falso-negativos, representando uma violação da integridade científica, que pode culminar a fragilização da relação entre a ciência e a sociedade, bem como em condutas médicas equivocadas. **Objetivo:** Descrever a prevalências de erro tipo II em ensaios clínicos negativos publicados em revistas de grande relevância científica.

**Metodologia:** Trata-se de um estudo metacientífico cuja amostra são ensaios clínicos negativos publicados entre 2020-2021, em revistas de grande relevância científica (*British Medical Association, Journal of the American Medical Association, The Lancet e New England Journal of Medicine*). Os critérios de exclusão foram: ensaios clínicos não randomizados, pragmáticos, de não inferioridade, de equivalência, duplicados e de conclusão positiva. Um artigo foi considerado com alta probabilidade de erro tipo II, caso houvesse a presença de 1 ou mais das seguintes variáveis: tamanho amostral insuficiente, poder insuficiente e alta variabilidade. **Resultados:** Dentre os 607 ensaios clínicos coletados, foram excluídos 411 e selecionados 190 ensaios clínicos negativos. Desse total de artigos analisados, 98 (52%) tiveram alta probabilidade de erro do tipo II, dos quais 59 (31%) tiveram poder estatístico insuficiente, 68 (36%) tamanho amostral insuficiente e 29 artigos (30%) poder estatístico e amostra insuficientes simultaneamente. Além disso, dos 190 trabalhos analisados, 14 artigos puderam ser avaliados quanto a alta variabilidade e foram-se encontrados 8 (57%) com tal variável. **Conclusão:** Encontrou-se uma quantidade relevante de ensaios clínicos randomizados com resultados potencialmente falso-negativos. Estudos mais robustos são necessários para aprofundar a evidência.

**Palavras-chave:** Erro aleatório. Erro tipo II. Metaciência. Integridade científica.

## ABSTRACT

**Introduction:** Systematic and random errors compromise the internal validity of the article, which interferes with the accuracy of the results. Thus, to guide researchers and readers, internal validity guidelines were made to assist in conducting research and reading articles. However, these checklists address requirements that almost completely involve precepts for avoiding systematic errors, but none for minimizing the presence of random errors specifically. This can lead to a large number of probably false-negative articles, representing a violation of scientific integrity, which can culminate in the fragilization of the relationship between science and society, as well as in mistaken medical conduct. **Objective:** The present study aims to describe the prevalence of type II error in negative clinical trials published in journals of high scientific relevance. **Methodology:** This is a metascientific study whose sample is negative clinical trials published between 2020-2021, in journals with great scientific support (British Medical Association, Journal of the American Medical Association, The Lancet and New England Journal of Medicine). Exclusion criteria were: non-randomized, pragmatic, non-inferiority, equivalence, duplicate and positive conclusion clinical trials. An article was considered with a high probability of type II error, if there was the presence of 1 or more of the following variables: insufficient sample size, insufficient power and high variability. **Results:** 607 clinical trials were collected, 411 were excluded and 190 negative clinical trials were selected. From this selected articles, 98 (52%) had a high probability of type II error, which 59 (31%) had insufficient statistical power, 68 (36%) insufficient sample size and 29 articles (30%) had statistical power and insufficient sample size simultaneously. On these 190 analyzed articles, 14 could be investigated for high variability and 8 (57%) were found with this variable. **Conclusion:** A relevant number of randomized clinical trials with potentially false-negative results was found. More robust studies are needed to deepen the evidence.

**Keywords:** Random error. Type II error. Metascience. Scientific integrity.

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>6</b>
<b>2</b>	<b>OBJETIVO .....</b>	<b>9</b>
2.1	GERAL .....	9
2.2	ESPECÍFICOS.....	9
<b>3</b>	<b>REVISÃO DE LITERATURA.....</b>	<b>10</b>
3.1	QUAIS SÃO OS TIPOS DE ERROS DE UM TRABALHO CIENTÍFICO?!.....	10
3.1.1	<i>Erros sistemáticos .....</i>	<i>10</i>
3.1.2	<i>Erros aleatórios: erro do tipo 1 e do tipo 2.....</i>	<i>10</i>
3.1.2.1	<i>Teste de hipóteses .....</i>	<i>11</i>
3.1.2.2	<i>Erro tipo I – o falso positivo .....</i>	<i>12</i>
3.1.2.3	<i>Erro tipo II – o falso negativo.....</i>	<i>13</i>
3.1.2.3.1	Poder estatístico e Tamanho amostral	13
3.1.2.3.2	Magnitude do efeito	14
3.1.2.3.3	Variabilidade	14
3.2	O QUE É E QUAL A IMPORTÂNCIA DA VALIDADE DE UM TRABALHO? .....	15
3.3	POR QUAL MOTIVO O ERRO TIPO II DEVE SER ESTUDADO?!.....	16
<b>4</b>	<b>METODOLOGIA.....</b>	<b>17</b>
4.1	DESENHO DO ESTUDO .....	17
4.2	LOCAL DO ESTUDO .....	17
4.3	AMOSTRA.....	17
4.4	CRITÉRIOS DE ELEGIBILIDADE.....	17
4.4.1	<i>Critérios de inclusão .....</i>	<i>17</i>
4.4.2	<i>Critérios de exclusão .....</i>	<i>17</i>
4.5	COLETA DE DADOS .....	18
4.5.1	<i>Instrumentos utilizados.....</i>	<i>19</i>
4.6	VARIÁVEIS DE CARACTERIZAÇÃO DOS ENSAIOS CLÍNICOS .....	19
4.7	VARIÁVEIS DE DESCRIÇÃO DO ERRO TIPO II .....	20
4.8	VARIÁVEIS EXPLORATÓRIAS .....	21
4.9	ASPECTOS ÉTICOS.....	22
<b>5</b>	<b>RESULTADOS .....</b>	<b>23</b>
5.1	ANÁLISES DE CARACTERIZAÇÃO ENSAIOS CLÍNICOS NEGATIVOS.....	23
5.2	RISCO DE ENSAIOS CLÍNICOS FALSO-NEGATIVOS .....	24
5.3	ANÁLISES EXPLORATÓRIAS .....	26
<b>6</b>	<b>DISCUSSÃO .....</b>	<b>27</b>
6.1	VARIÁVEIS DE ANÁLISE DO ERRO TIPO 2 .....	27
6.2	VARIÁVEIS EXPLORATÓRIAS .....	29
6.3	PARTICULARIDADES E LIMITAÇÕES.....	30
<b>7</b>	<b>CONCLUSÃO.....</b>	<b>32</b>
	<b>REFERÊNCIAS.....</b>	<b>33</b>



## 1 INTRODUÇÃO

Durante o século XX, Michael Foucault dividiu a sua concepção sobre a verdade em dois tipos: a verdade acontecimento, considerada descontínua e particular de ocasiões muito específicas; e a verdade-demonstração, na qual é considerada global e abrangente<sup>1</sup>. Essa verdade-demonstração, além de universal é objetiva e requer a utilização de instrumentos de mensuração específicos<sup>1,2</sup>. Trazendo para o contexto científico, é possível traçar uma relação entre os resultados esperados em trabalhos científicos na área de saúde e a verdade-demonstração de Foucault, no que tange tanto a geração de um conceito universal, que pode ser generalizado entre pacientes, mas reconhecendo as suas particularidades, quanto na necessidade de se utilizar instrumentos e técnicas objetivas de mensuração dos dados. Entretanto, apesar de o objetivo da maioria desses trabalhos ser a aproximação máxima do resultado encontrado da verdade-demonstração, o que se vê é o afastamento desse por conta de erros.

Os erros em questão são os erros sistemáticos e aleatórios. Os sistemáticos são aqueles que ocorrem por conta da falha em alguma etapa do delineamento do estudo, seja durante a revisão, coleta, publicação, interpretação ou análise das informações colhidas, geram resultados e conceitos sistematicamente equivocados<sup>3</sup>, ou seja resultados enviesados em uma determinada direção<sup>4</sup>. Já os erros aleatórios são aqueles que fornecem um resultado equivocado em virtude do acaso, não havendo como prever tal evento<sup>4</sup>, além do que são intrínsecas as observações feitas nos estudos e não podem ser evitados, mas sim minimizados<sup>3</sup>.

A presença dos erros aleatórios tem influência na validade de um trabalho. A validade de um artigo permite a compreensão de perceber se determinado estudo possui inferências de alta ou baixa qualidade, e se divide em interna e externa<sup>5</sup>. A primeira concerne no quão precisos são as conclusões do trabalho<sup>6</sup>, ou seja, no grau de liberdade dos resultados em relação aos erros sistemáticos e aleatórios. Já a validade externa, é referente a capacidade do estudo de gerar um conceito, o qual possa ser extrapolado para toda a população<sup>5,7</sup>. Ademais, percebe-se a relação da validade de um estudo com o método científico hipotético-dedutivo, que permite essas generalizações<sup>8</sup>. Assim, para que muitos pesquisadores e leitores, consigam se guiar

na conduta das pesquisas e em leituras de artigo científico, respectivamente, existem diversos checklists de validade interna que os auxiliam nesse processo, como o RoB 2, utilizado em ensaios clínicos<sup>7</sup>.

Em contrapartida, o que se percebe nesses guidelines é que as exigências abordadas para que um estudo seja considerado de alta validade interna envolvem quase que completamente cuidados que evitem os erros sistemáticos, mas nenhuma delas visa minimizar a presença de erros aleatórios diretamente. Dessa forma, sabendo que a validade interna é pré-requisito para a externa<sup>7</sup>, a validade integral desses ensaios clínicos encontra-se em jogo, em virtude da possibilidade da alta frequência de erros aleatórios nos trabalhos científicos, ou seja, da existência de trabalhos que resultem em, por mais que de forma não intencional, em mentiras ou omissões. Nesse contexto, pode-se suspeitar da existência do alto risco de erro aleatório, inclusive o erro do tipo II, em ensaios clínicos negativos, incluindo os presentes em revistas de alta relevância científica.

Tal cenário evidencia a maior preocupação da comunidade científica com os erros sistemáticos, desvalorizando as importantes consequências dos erros aleatórios. Assim, em sendo a hipótese do presente trabalho verdadeira, seria possível concluir que provavelmente muitas decisões tomadas com base nos ensaios clínicos publicados em revistas de alta relevância não são tão confiáveis quanto acredita-se, afinal seus resultados podem se distanciar muito da realidade e representar violação da integridade científica, repercutindo na tomada de decisão clínica e na fragilização da relação de confiança entre a ciência e a sociedade. O enfraquecimento dessa relação, tem consequências graves, pois desde sempre as pessoas precisaram de algo para acreditar, e fazer disso, uma fonte para embasar suas decisões.

Assim, caso as pessoas deixem de confiar na ciência, essas decisões passarão a ter respaldo em argumentos de autoridade, ou seja, em um conhecimento cuja validade tem relação com a formação acadêmica prévia de quem o originou e não na consistência e comprovação do argumento em si. Isso permitiria que condutas que envolvem a vida e a saúde de outras pessoas sejam embasadas em “achismo”. Além disso, o estudo seria um alerta para as consequências desse erro e refletiria na necessidade de todo ecossistema científico, envolvendo os pesquisadores e os

profissionais, se capacitar melhor acerca dos erros aleatórios, buscando se informar e desenvolver instrumentos para a avaliação, mensuração e minimização deles.

## **2 OBJETIVO**

### **2.1 Geral**

Descrever a prevalências de erro tipo II em ensaios clínicos negativos publicados em revistas de grande relevância científica.

### **2.2 Específicos**

Qualificar os erros tipo II presentes em ensaios clínicos publicados em revistas de grande relevância científica.

### 3 REVISÃO DE LITERATURA

#### 3.1 Quais são os tipos de erros de um trabalho científico?!

Sob a ótica da ciência, os erros podem ser interpretados como o resultado equivocado de uma medição<sup>5</sup>. Nesse sentido, os artigos científicos, como a própria verdade-demonstração de Foucault, se baseiam na mensuração dos dados<sup>1,2</sup>, e como qualquer mensuração, estão sujeitos aos erros, os quais podem se manifestar por uma falha metodológica, levando a diminuição da acurácia, e/ou pela atuação do acaso e diminuindo a precisão dos resultados<sup>4,9</sup>. O primeiro erro se refere ao sistemático, já o segundo, diz respeito ao aleatório.

##### 3.1.1 Erros sistemáticos

O erro sistemático, também chamado de viés, é considerado um erro que desvia o resultado para um determinado sentido<sup>5</sup>. Ou seja, existe um afastamento sistemático dos resultados da amostra em relação ao valor real<sup>3</sup>. Isso acaba por promover uma associação equivocada entre uma exposição e um efeito numa população alvo, na qual os resultados estatísticos estimados possuem valores muito distantes aos reais<sup>3,6,10</sup>.

Na década de 70, esse tipo de erro já era discutido e foi catalogado por David L. Sackett, o qual revelou 35 tipos de vieses<sup>11</sup>. Já em 2004, Miguel Delgado-Rodrigues, com o avançar dos estudos sobre essa temática, publicou um trabalho relatando 74 tipos de vieses existentes, que hoje se dividem em diversos grupos, sendo os principais, os vieses de seleção, confusão e informação<sup>10</sup>. O viés de seleção ocorre quando a amostra não representa bem a população alvo, o de confusão quando há uma associação espúria devido a uma falha no ajuste para fatores relacionados ao fator de risco e ao resultado, e por fim, o de informação ocorre quando as medições dos desfechos não são sistematicamente feitas e padronizadas<sup>3,10,12</sup>.

##### 3.1.2 Erros aleatórios: erro do tipo 1 e do tipo 2

Já no que tange ao erro aleatório, ao invés de direcionar o resultado em um único sentido ele tem a capacidade de distorcê-los em ambas as direções, em virtude do acaso<sup>4</sup>. Esse tipo de erro se dá pela variabilidade aleatória dos dados mensurados<sup>13</sup>, sendo inerente a qualquer observação, mas apesar de poder ser minimizado, nunca poderá ser evitado por completo. A principal forma de minimização desses erros é aumentar o tamanho amostral, afinal, quando maior a amostra, mais ela se assemelhará a população original<sup>14</sup>. Esses erros aleatórios se dividem em dois grupos, o erro tipo I e o erro tipo II, mas para melhor compreensão faz-se necessário entender previamente o teste de hipóteses.

### *3.1.2.1 Teste de hipóteses*

O filósofo austríaco Karl Popper enfatizava que o objetivo de um cientista não é a verificação de uma hipótese inicial, e sim a tentativa de falseá-la, afinal é impossível comprovar a veracidade de algo com a observação, mas é factível perceber que a hipótese é equivocada com a mesma atitude<sup>14</sup>. Essa ideia torna-se tangível ao se conhecer a Sátira VI do poeta romano Juvenal, escrita no século II, em que as pessoas acreditavam na regra universal de que todos os cisnes eram brancos, afinal só haviam observados cisnes dessa coloração. Entretanto, foi com a observação de um único cisne negro que essa regra se tornou falsa e foi rejeitada<sup>15</sup>.

Nesse cenário, fica mais fácil de compreender o motivo de ser indicado que os pesquisadores iniciem seus trabalhos sempre com a anulação da própria hipótese, tentando comprovar justamente o oposto, ou seja, partindo sempre da hipótese nula ( $H_0$ ). A  $H_0$  pode ser conceituada como a hipótese a ser testada<sup>16</sup>, sendo muito comum que seja a não associação de uma variável com a outra ou a ausência de diferença entre os grupos comparados em um estudo<sup>5</sup>. Isso ocorre, pois partindo do pressuposto da inexistência de uma associação na população, testes estatísticos podem auxiliar na estimativa da probabilidade de a possível associação observada ser em virtude do acaso<sup>4</sup>.

Em contrapartida, mesmo com o início da jornada sendo com a  $H_0$ , os pesquisadores visam encontrar a verdade e por consequência, acabam por rejeitá-la, para passarem

a aceitar a hipótese alternativa ( $H_1$ )<sup>16</sup>. A  $H_1$ , normalmente, é a própria hipótese da pesquisa, ou seja, se relaciona diretamente com a teoria que se quer investigar, sendo comumente a presença da diferença entre os grupos comparados no trabalho<sup>12,16</sup>. Entretanto, essa hipótese não pode ser testada diretamente, sua aceitação ocorre apenas se o teste de significância estatística rejeitar a  $H_0$ <sup>4,14</sup>.

Dessa forma, o teste de hipóteses consiste em um procedimento de testagem estatística das hipóteses<sup>17</sup>, objetivando auxiliar os pesquisadores a concluírem sobre a realidade examinando “apenas” uma amostra<sup>16</sup>. Com base nessa testagem é possível chegar em 4 possibilidades de relação entre os fatos e o resultado do trabalho: (1) aceitar a  $H_0$  quando ela realmente se aplica na sociedade, afastando a  $H_1$ ; (2) rejeitar a  $H_0$  quando ela não se aplica na sociedade, abraçando a  $H_1$ ; (3) rejeitar a  $H_0$  quando ela se aplica na sociedade, abraçando a  $H_1$ ; (4) aceitar a  $H_0$  quando ela não se aplica a sociedade, afastando a  $H_1$ . Assim, as possibilidades (1) e (2) levam a conclusões verdadeiras, mas a (3) e a (4) dizem respeito a resultados falsos, ou seja, aos erros aleatórios, sendo o terceiro o erro tipo I e o quarto o erro tipo II<sup>3,13</sup>

### 3.1.2.2 Erro tipo I – o falso positivo

A probabilidade de se cometer esse erro é denominada de  $\alpha$ , sendo ele considerado o valor limite para a significância estatística<sup>13</sup>. Dessa forma, o  $\alpha$  (alfa) é considerado o limite máximo de erro tipo 1 aceito em um estudo, normalmente de 0,05 (5%)<sup>14</sup>. Isto é, caso um estudo seja reproduzido 20 vezes, em 1 dessas repetições o resultado seria positivo, mesmo sendo a  $H_0$  verdadeira, por conta do acaso<sup>18</sup>.

Esse tipo de erro é motivado pelo acaso, mas certas condutas pelos pesquisadores podem maximizar a probabilidade do seu aparecimento. Dentre elas, as principais são o pequeno tamanho amostral e a testagem de múltiplas hipóteses, incluindo a análise de subgrupo, supervalorização de desfechos secundários e análises interinas<sup>4,19</sup>. Ademais, cabe salientar que o erro tipo 1 é considerado o pior dos erros, afinal o seu resultado faz com que uma conduta seja substituída por outra, por ser demonstrada melhor, entretanto, na realidade não seria<sup>19</sup>, ou seja, pessoas serão cuidadas da

maneira menos adequada ou de forma inefetiva, sendo submetidas, por vezes, procedimentos invasivos sem necessidade.

### 3.1.2.3 Erro tipo II – o falso negativo

O erro tipo II, também possui uma probabilidade máxima aceitável,  $\beta$  (beta), tendo, normalmente, seu valor de 0,2 (20%) <sup>14</sup>. Em outras palavras, caso um estudo seja reproduzido 20 vezes, em 4 dessas repetições o resultado seria negativo, mesmo sendo a  $H_1$  verdadeira, em virtude dos desígnios do acaso <sup>18</sup>. Assim, como o erro tipo I, o tipo II não pode ser evitado, mas pode ter sua probabilidade aumentada a depender de condutas do pesquisador. No caso do falso negativo, fatores que aumentam a sua probabilidade são: pequeno tamanho amostral, baixo poder estatístico, baixa magnitude e alta variabilidade <sup>12</sup>.

#### 3.1.2.3.1 Poder estatístico e Tamanho amostral

O poder estatístico, é considerado a capacidade de um estudo de demonstrar uma associação ou efeito que realmente existe. Matematicamente falando, Poder =  $1 - \beta$  <sup>3,12</sup>. Assim, por convenção, sendo  $\beta$ , 20% normalmente, o poder mínimo de um estudo é considerado 80%, ou seja, considera-se que a probabilidade mínima aceita para a detecção correta de uma diferença entre os grupos é 80% <sup>20</sup>. Assim, é importante perceber pela fórmula que quanto maior o poder estatístico, menor  $\beta$ , isto é, quando maior o poder, menor a probabilidade de erro tipo II. Além disso, o poder estatístico normalmente é determinado por alguns fatores, dentre eles: tamanho amostral, magnitude do efeito e a variabilidade das observações <sup>12,13</sup>.

Nesse contexto, sabe-se que o tamanho amostral tem relação tanto com o poder, quanto com o erro tipo II, afinal, quanto menor a amostra, menor é o poder estatístico e maior será a probabilidade de haver o erro tipo II <sup>14</sup>. *Mas o que é um tamanho amostral pequeno ou grande?! A amostra (quantidade de pessoas necessárias para a realização do estudo) é estabelecida antes do início do trabalho e depende do valor de  $\alpha$ , poder estatístico, natureza dos dados <sup>13</sup> e principalmente da frequência do resultado que se almeja encontrar no grupo controle, ou seja, da taxa de eventos <sup>20,21</sup>.*

Dessa forma, quanto menor a taxa de eventos, maior deverá ser o tamanho da amostra para que ela seja considerada adequada <sup>21</sup>.

#### 3.1.2.3.2 Magnitude do efeito

A magnitude é o tamanho do efeito encontrado nos resultados de um estudo, <sup>22</sup>, sendo considerado a menor diferença entre os grupos para que a intervenção seja clinicamente importante <sup>23</sup>. Esse dado é relevante, pois permite que se possa determinar o benefício e a importância do que está sendo testado no estudo <sup>18,22</sup>. O tamanho do efeito desejado para determinada intervenção é pré-estabelecido e com base nisso se decide o poder e o tamanho amostral a ser utilizado<sup>14</sup>.

Isso ocorre, pois, diferenças muito grandes são facilmente detectáveis em amostras, inclusive nas pequenas, entretanto diferenças menores para serem identificadas precisam de um tamanho amostral maior <sup>12,14</sup>. Assim, no caso de magnitudes pequenas, o ajuste da amostra também ajusta o poder estatístico, minimizando a presença do erro tipo II. Dessa forma, compreendendo a relação da magnitude com o poder fica evidente que quanto maior for o tamanho do efeito, maior será o poder e menor a probabilidade do erro tipo II <sup>24</sup>.

#### 3.1.2.3.3 Variabilidade

A variabilidade é a variação populacional de uma determinada variável do estudo, normalmente expressa pelo desvio padrão, no caso de variáveis contínuas <sup>25</sup> O poder de um estudo é afetado pelo grau de variação dos pacientes, afinal, quanto maior a variação entre eles em relação a característica medida, menos se pode confiar que a diferença percebida é realmente fruto da intervenção e não da variação prévia <sup>3</sup>. Assim, quanto menor a variabilidade percebida entre as observações, maior o poder estatístico e conseqüentemente menor a probabilidade do erro tipo II<sup>12</sup>. Além disso, não só a alta variabilidade, mas os outros fatores citados como, baixo poder estatístico, tamanho amostral, magnitude são determinantes da validade de um trabalho<sup>24</sup>

### 3.2 O que é e qual a importância da validade de um trabalho?

O método hipotético-dedutivo, proposto pelo filósofo Karl Popper, é um modelo que fornece um esquema com várias variações, especificamente para testagem e avaliação de hipóteses científicas <sup>26</sup>. Ele se inicia com uma lacuna/problema previamente observado, que sugere a formação de uma hipótese, na qual é testada <sup>27</sup>. Esse método ainda é amplamente utilizado, principalmente na área das ciências naturais e se diferencia de outros, como o indutivo, por permitir a generalização dos conceitos e, conseqüentemente, a construção de teorias e leis <sup>8</sup>. Essa generalização tem ampla relação com a validade de um estudo <sup>28</sup>.

A validade de um estudo científico, independentemente do seu desenho, diz respeito ao quanto o resultado encontrado nos participantes do trabalho representa achados verdadeiros em pessoas parecidas fora do estudo, ou seja, refere-se ao no quanto podemos confiá-lo <sup>5,29</sup>. Essa validade se divide em validade externa e interna. A primeira se refere ao grau que os resultados dos ensaios dão respaldo para que se generalize o conceito para outras circunstâncias clínicas<sup>5,7</sup>, ou seja, se referem a generalização e a aplicabilidade dos resultados <sup>28</sup>

No que tange a validade interna, ela concerne no grau de liberdade de erros, tanto sistemáticos quanto aleatórios, do conceito gerado <sup>5,7</sup>. As ameaças a validade interna de um trabalho são diversas, dentre elas: as mudanças fisiológicas e psicológicas dos indivíduos, uma baixa amostra ou um pequeno poder estatístico <sup>6</sup>. Dessa forma, como a baixa validade interna de um trabalho influencia de forma negativa a evidência gerada <sup>7</sup>, checklists para orientar leitores e pesquisadores, na leitura e na condução do trabalho, respectivamente, foram criados.

Os checklists para a testagem da validade interna são inúmeros, sendo cada um deles utilizado para um tipo de estudo específico. Dentre eles tem-se o Risk of Bias 2 (RoB 2) utilizado para ensaios clínicos, AMSTAR para revisões sistemáticas e o Newcastle-Ottawa Scale (NOS) para estudos de caso controle. Entretanto, o que se percebe é a inexistência de tópicos que visem avaliar/ minimizar os erros aleatórios diretamente. Assim, fica evidente que os erros aleatórios são invisibilizados, e por essa razão

acredita-se que há alta probabilidade de risco de erros aleatórios, inclusive o erro tipo II nos ensaios clínicos publicados em periódicos de alta relevância científica.

### **3.3 Por qual motivo o erro tipo II deve ser estudado?!**

O erro tipo II é considerado menos maléfico que o erro tipo I <sup>1919</sup>, por isso tem menor visibilidade, mesmo, inevitavelmente, também repercutindo na perpetuação do cuidado menos adequado ao paciente. Dentre os pontos negativos do foco no erro tipo I, tem-se o aumento dos erros tipo II e do viés de publicação de efeitos grandes e óbvios <sup>30</sup>. Ademais, caso os resultados falso negativos sejam invisibilizados, possivelmente a ciência continuará tentando encontrar soluções já existentes, mas ainda não percebidas, para questões que atormentam a saúde de muitas pessoas.

Isso ocorre, pois em vigência do falso negativo, tratamentos úteis são completamente descartados <sup>31</sup>, o que faz com que se perca a oportunidade de focar em temas que de fato não possuem solução. Outrossim, haja vista o foco da comunidade científica ser nos erros sistemáticos, e dificilmente, quando nos erros aleatórios, a preocupação ser direcionada ao erro tipo I, o erro tipo II, é pouco mencionado, principalmente nos artigos médicos <sup>32</sup>. Nesse sentido, faz-se necessário o aprofundamento da investigação sobre esse tipo de erro.

Além disso, essa temática já foi questionada em outro estudo, mas em uma área específica da medicina. Essa área é a de trauma ortopédico<sup>33</sup>, na qual concluiu um alto potencial de erro tipo II nos artigos analisados<sup>34</sup>. Em contrapartida, até hoje não foi registrado um trabalho que analisasse esse tipo de erro em revistas de grande relevância científica, uma temática de suma importância, afinal, se forem encontrados resultados positivos para essas revistas, a generalização desse conceito para outros trabalhos publicados em revistas de menor rigor técnico e metodológico poderá ser feita.

## 4 METODOLOGIA

### 4.1 Desenho do estudo

O estudo é metacientífico, cujo perfil epidemiológico é melhor descrito abaixo:

- **unidade de análise:** a unidade de origem dos dados é individuada, visto que a avaliação individual de cada artigo é relevante para os objetivos;
- **período de seguimento:** transversal, pois os dados de cada artigo são coletados uma única vez no período analisado;
- **direcionalidade temporal:** A análise descritiva é prospectiva;

### 4.2 Local do estudo

Os locais do estudo são os bancos de dados dos artigos publicados nas revistas de alta relevância científica: *The New England Journal of Medicine (NEJM)*, *The Lancet*, *British Medical Journal (BMJ)* e *The Journal of the American Medical Association (JAMA)*.

### 4.3 Amostra

Serão ensaios clínicos negativos publicados em periódicos de grande relevância científica (de acordo com o fator de impacto) nos anos de 2020 e 2021. O fator de impacto foi utilizado por ser comumente utilizado na avaliação da relevância de uma revista dentro da área que atua <sup>35</sup>. Os periódicos selecionados são: NEJM, The Lancet, BMJ e JAMA.

### 4.4 Critérios de elegibilidade

#### 4.4.1 Critérios de inclusão

- Ensaios clínicos publicados nas revistas NEJM, The Lancet, BMJ e JAMA no período de 2020-2021.

#### 4.4.2 Critérios de exclusão

- Ensaio clínico com conclusão positiva;
- Ensaio clínico não randomizado;
- Ensaio clínico pragmático;
- Ensaio clínico de equivalência;
- Ensaio clínico de não inferioridade;

#### 4.5 Coleta de dados

A coleta dos dados foi realizada de forma independente entre as duas pesquisadoras e se iniciou com a coleta manual dos ensaios clínicos presentes nas edições de 2020 e 2021 das revistas BMJ, JAMA, The Lancet e NEJM. Inicialmente as autoras acessaram os sites de todas as revistas, por meio de seus computadores pessoais, acessando respectivamente: <https://www.bmj.com/>, <https://jamanetwork.com/>, <https://www.thelancet.com/> e <https://www.nejm.org/>. Após a identificação dos ensaios clínicos, esses estudos foram previamente discutidos entre as duas pesquisadoras, analisando quais trabalhos passariam pelos critérios de exclusão. Após a primeira discussão os trabalhos foram transferidos para o Rayyan, para a aplicação desses critérios. Para aplicação desses critérios foram utilizadas medidas facilitadoras como as “tags”, ou seja, palavras previamente selecionadas que favoreceriam a exclusão ou a permanência do trabalho no estudo. Dentre as “tags” que foram utilizadas tem-se:

- Tags que favoreceram a exclusão: “pragmatic”, “non-inferiority”, “nonrandomized”, “technology”, “noninferiority”, “non-randomized” e “equivalence”.
- Tags que favoreceram a permanência do estudo: “randomized”, “randomization”, “superior” e “superiority”.

Em seguida, as duas pesquisadoras individualmente classificaram os artigos como sendo os excluídos ou não. Ao fim dessa seleção, as discrepâncias entre as pesquisadoras em relação a quais ensaios clínicos seriam realmente utilizados no trabalho, foram discutidas. Em sequência, esses foram analisados quanto a prevalência de erro do tipo II, com base nas variáveis de análise do erro do tipo II, considerando sempre o desfecho primário dos ensaios analisados. Os dados coletados foram armazenados manualmente no Excel e em contas pessoais do

Google Drive, e ao final da coleta, as duas pesquisadoras discutiram as divergências para a obtenção de um consenso. Durante o processo, quando necessário, as discordâncias foram resolvidas por um terceiro pesquisador.

#### 4.5.1 Instrumentos utilizados

- Rayyan: utilizado para a aplicação dos critérios de elegibilidade dos ensaios clínicos coletados.
- Mendeley: utilizado no auxílio na leitura de trabalhos, bem como no gerenciamento das referências utilizadas.
- Excel: utilizado para o armazenamento e gerenciamento dos dados coletados

#### 4.6 Variáveis de caracterização dos ensaios clínicos

- **Especialidade da área médica:** variável qualitativa nominal politômica, coletada com base no ramo da área médica que o trabalho abordou, sendo possível classificar os artigos nas seguintes áreas: Acupuntura, alergia e imunologia, anestesiologia, angiologia, cancerologia, cardiologia, cirurgia cardiovascular, cirurgia da mão, cirurgia de cabeça e pescoço, cirurgia do aparelho digestivo, cirurgia geral, cirurgia pediátrica, cirurgia plástica, cirurgia torácica, cirurgia vascular, clínica médica, coloproctologia, dermatologia, endocrinologia e metabologia, endoscopia, gastroenterologia, genética médica, geriatria, ginecologia e obstetrícia, hematologia e hemoterapia, homeopatia, infectologia, mastologia, medicina de família e comunidade, medicina do trabalho, medicina de trânsito, medicina esportiva, medicina física e reabilitação, medicina intensiva, medicina legal, medicina nuclear, medicina preventiva e social, nefrologia, neurocirurgia, neurologia, nutrologia, oftalmologia, ortopedia e traumatologia, otorrinolaringologia, patologia, patologia clínica/medicina laboratorial, pediatria, pneumologia, psiquiatria, radiologia e diagnóstico por imagem, radioterapia, reumatologia e urologia. Essas áreas selecionadas com base na Resolução CFM N° 1845/2008.

- **Número de pacientes utilizados (amostra):** variável quantitativa discreta, coletada através da quantidade de pessoas que compõe a amostra do ensaio clínico a partir da randomização;
- **Ser um estudo multicêntrico:** variável dicotômica, categorizada em “não” e “sim”, sendo considerado “sim” se  $> 1$ , e coletada com a base na quantidade de centros de saúde em que a pesquisa foi realizada;
- **Tipo de intervenção:** variável qualitativa nominal politômica, coletada de acordo com a conduta utilizada no grupo tratamento. Pode ser dividida em: “cirurgia”, “medicamento” e “procedimento”, sendo última categoria utilizada para todo tratamento que não é considerado cirurgia ou medicamento, incluindo mudanças alimentares, exercícios físicos, medidas educativas, controle de temperatura, vacinações e infusão venosa, por exemplo.
- **Vínculo com a indústria:** variável categórica, categorizada em “Grau 1: Disponibilização de verba para o trabalho” ou “Grau 2: Envolvimento da indústria nos métodos” e “Grau 3: Disponibilização de verba para ao menos 1 dos autores”, considerando indústria como qualquer empresa privada.
- **Ausência de desfecho primário definido:** Variável dicotômica, categorizada em “não” ou “sim”, sendo considerada “sim” quando o artigo explicitar qual o desfecho primário do trabalho. Coletada por meio da procura da declaração do desfecho primário pelos autores no trabalho.
- **Ausência de protocolo:** Variável dicotômica, categorizada em “não” ou “sim”, sendo considerada “sim”, quando o artigo explicitar o número do protocolo do trabalho. Coletada por meio da procura da declaração número do protocolo do trabalho pelos autores.

#### 4.7 Variáveis de descrição do Erro tipo II

- **Poder estatístico insuficiente:** Variável dicotômica. Categorizada como “não” ou “sim”, sendo considerado como “sim”, ou seja, insuficiente, quando o poder é  $< 80\%$ . Será extraída por meio da comparação do poder estatístico do estudo com o valor de 80%. A presença de uma amostra menor do que a planejada também indica poder estatístico insuficiente.

- **Tamanho amostral insuficiente:** Variável dicotômica. Categorizada como “não” ou “sim”, sendo considerada “sim”, ou seja, insuficiente, quando < 400 pacientes. Será coletada por meio da observação da quantidade de pacientes utilizados no ensaio clínico;
- **Alta variabilidade:** Variável dicotômica. Categorizada como “sim” ou “não”, sendo considerada “sim”, ou seja, alta quando > 20% em ao menos um dos grupos. Aferida de acordo com o coeficiente de variação, calculado por meio da fórmula:  $(\sigma / |\mu|) \times 100$ , sendo  $\sigma$  = desvio padrão;  $\mu$  = média aritmética. Quando a medida de tendência central não foi a média, e a medida dispersão não foi o desvio padrão a variabilidade não pôde calculada.

#### 4.8 Variáveis exploratórias

- **Mudança da expressão da medida de variabilidade:** Variável dicotômica classificada como “não” ou “sim”, sendo considerada “sim” quando a medida de variabilidade muda ao longo do trabalho. Coletada com base na observação da mudança da expressão da medida de variabilidade usada durante o estudo, podendo ser variância, desvio padrão ou erro padrão da média;
- **Ausência de descrição do cálculo amostral:** Variável dicotômica classificada como “não” ou “sim”, sendo considerado “sim”, quando o poder estatístico, amostra necessária e a diferença desejada entre os grupos não são descritos. Coletada com base na ausência da observação da descrição do poder, da diferença que cogitam encontrar entre os grupos e amostra que precisará compor o trabalho;
- **Presença da descrição de outros achados após a conclusão negativa:** Variável dicotômica classificada como “não” ou “sim”, sendo considerada “sim” na presença de conclusão positiva de objetivos que não são primários. Coletada com base na observação da conclusão positiva de outros achados que não condizem com o objetivo primário do trabalho.

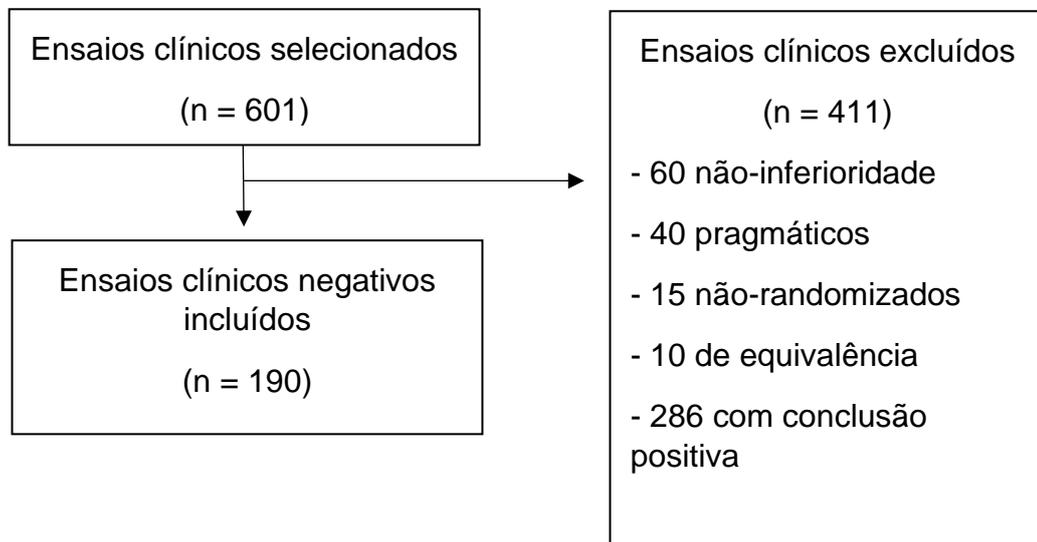
#### **4.9 Aspectos Éticos**

Segundo a resolução nº 466, de 12 de dezembro de 2012, em respeito a dignidade humana e em prol da proteção a vida, uma pesquisa envolvendo seres humanos deve atender a fundamentos éticos. Tais fundamentos envolvem, dentre outros pré-requisitos, a submissão do trabalho ao Sistema CEP/CONEP/CNS/MS, para que seja analisada e aprovada. Nesse sentido, como o trabalho em questão não envolveu seres humanos, a submissão ao sistema citado não foi necessária.

## 5 RESULTADOS

Foram incluídos no presente trabalho 607 artigos, coletados nos periódicos *British Medical Journal* (BMJ), *New England Journal of Medicine* (NEJM), *The Lancet* e *The Journal of the American Medical Association* (JAMA). Desses artigos, 125 foram excluídos, com base na aplicação dos critérios de exclusão, restando 190 ensaios clínicos com desfecho primário negativo (**figura 1**).

**Figura 1** - Seleção dos ensaios clínicos randomizado negativos.



**Fonte:** elaborado pelo autor, 2022.

### 5.1 Análises de caracterização ensaios clínicos negativos

A maioria dos artigos eram multicêntricos, sendo eles 89,47%, majoritariamente da área médica de infectologia, totalizando 25,78%, e predominantemente possuíam vínculo com a indústria grau três, sendo eles 57,30% de todos os artigos analisados. Além disso, a maior parte dos artigos tinham intervenção medicamentosa, sendo equivalentes a 57,36%. Ademais, nenhum dos estudos deixou de apresentar protocolo ou desfecho principal (**tabela 1**).

**Tabela 1** - Características dos ensaios clínicos negativos publicados na BMJ, JAMA, LANCET e NEJM entre os anos 2020 e 2021 (n = 190).

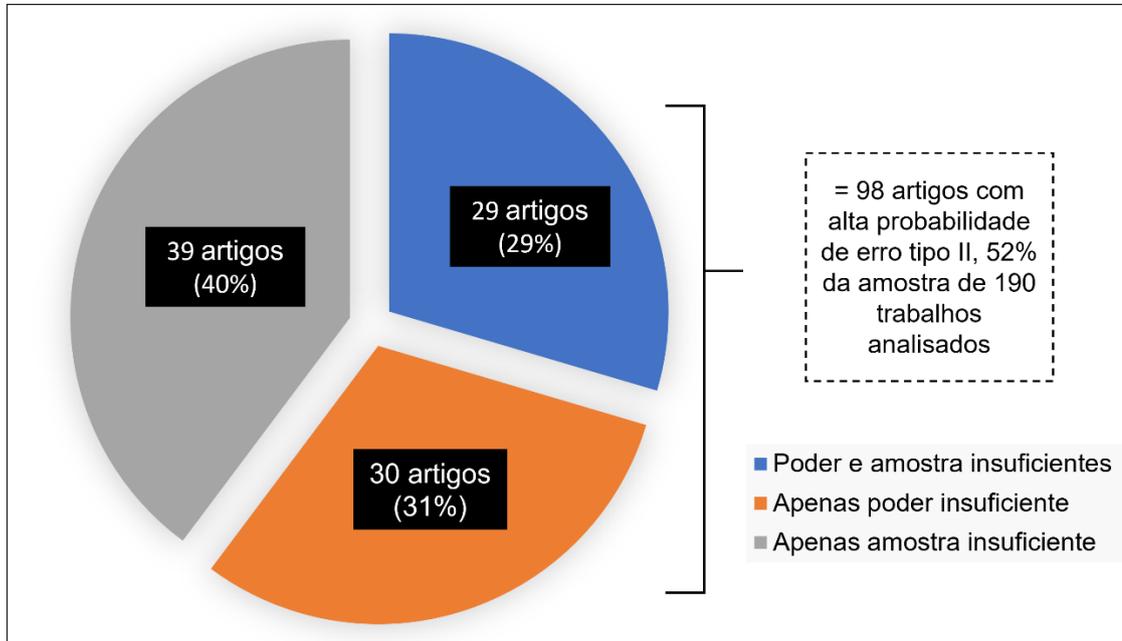
<b>Variável analisada</b>	<b>Valor encontrado</b>
<b>Especialidade</b>	
Infectologia	49 (26%)
Cardiologia	34 (18%)
Pneumologia	12 (6,3%)
Outros	95 (50%)
<b>Tipo de intervenção</b>	
Medicamento	109 (57%)
Procedimento	69 (36%)
Cirurgia	12 (6,3%)
<b>Presença de financiamento</b>	
Grau1	35 (18%)
Grau 2	0
Grau 3	109 (57%)
<b>Multicêntrico</b>	170 (90%)
<b>Ausência de protocolo</b>	0
<b>Ausência do desfecho primário</b>	0

Fonte: elaborado pelo autor, 2022.

## 5.2 Risco de ensaios clínicos falso-negativos

Dos 190 artigos com desfecho primário negativo, 98 (52%) tiveram alta probabilidade de erro do tipo II (**figura 2**), dos quais 59 (31%) tiveram poder estatístico insuficiente e 68 (36%) tamanho amostral insuficiente. Cabe salientar que 29 artigos (30%) dos 98 foram considerados com alta probabilidade de erro tipo II pela presença de ambas as variáveis descritas no período anterior, ou seja, tanto pela presença de poder estatístico quanto amostra insuficientes (**figura 2**).

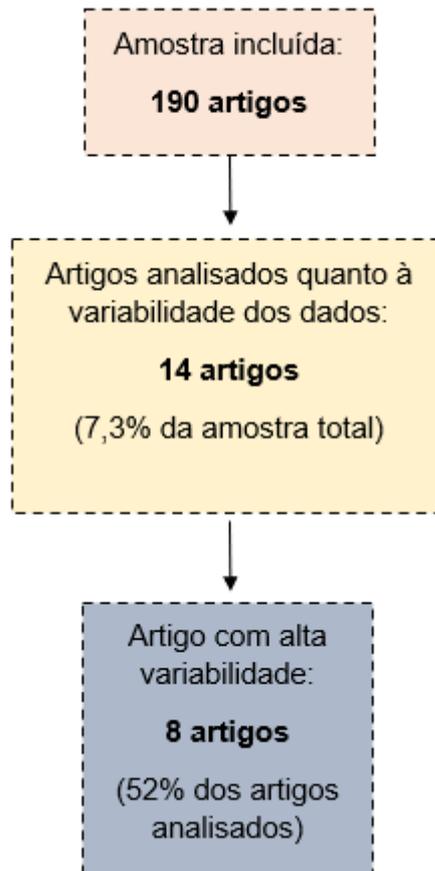
**Figura 2** - Artigos potencialmente falso negativos por amostra insuficiente e/ou poder insuficiente.



**Fonte:** elaborado pelo autor, 2022.

A alta variabilidade dos dados só conseguiu ser analisada através da avaliação do comportamento dos desfechos principais, em artigos que expressaram seus dados por média e por desvio padrão. Na amostra de 190 artigos analisados, apenas 14 estudos (7,3% da amostra total) expressaram seus dados da forma que permitisse essa análise, e desses apenas oito artigos apresentaram a variável em questão. Assim, do total de artigos que puderam ser analisados para essa variável, 57% foram considerados com alta variabilidade, ou seja, com alta probabilidade de erro do tipo II (**figura 4**).

**Figura 4** - Proporção dos artigos com alta variabilidade dos dados.



Fonte: elaborado pelo autor, 2022.

### 5.3 Análises exploratórias

Dos 190 artigos analisados, 12 artigos (6,3%) concluíram positivamente sobre desfechos que não fossem primários e 23 (12%) não possuíram a descrição do cálculo amostral. Além disso, em nenhum dos artigos essas duas variáveis coexistiram. Cabe salientar, que nenhum dos artigos analisados alterou a forma de expressar a variabilidade.

## 6 DISCUSSÃO

### 6.1 Variáveis de análise do erro tipo 2

A maioria dos ensaios clínicos tiveram alta probabilidade de erro do tipo 2, pela presença de poder estatístico insuficiente e/ou tamanho amostral insuficiente ou alta variabilidade dos dados. Assim, esse resultado revela que a maioria dos artigos de revistas relevantes tem baixa validade interna, o que aumenta a probabilidade de falso negativo<sup>5-7</sup>. Isso leva a violação da integridade científica, e conseqüentemente ao prejuízo na divulgação da ciência, associada à descredibilização da mesma, e a condutas médicas equivocadas.

No atual contexto científico, em que os guidelines de validade interna priorizam os erros sistemáticos, a possibilidade da presença dos erros aleatórios acaba por gerar poucas preocupações a comunidade científica, que tende a minimizar as suas conseqüências. Porém, mesmo que esses erros sejam como um todo preteridos, a preocupação com os falsos positivos ainda é maior do que com os falsos negativos, haja vista esse ser considerado o pior dos erros<sup>19</sup>. Logo, pode-se inferir que um dos motivos para o elevado número de publicações provavelmente falso negativas, seja pelo simples fato de tais erros serem ignorados.

A negligência desse erro e o conseqüente resultado do presente trabalho podem ser interpretados como uma violação da integridade científica. Essa integridade diz respeito aos deveres éticos que os cientistas, instituições e participantes das pesquisas se comprometem a cumprir ao realizarem trabalhos científicos<sup>35</sup>. O que visa assegurar que os conceitos gerados no trabalho foram baseados em confiança e responsabilidade<sup>36</sup>. Desse modo, a conclusão deste trabalho revela uma má prática em pesquisa, haja vista o descuido em relação a potenciais modificadores dos resultados, transformando-os em conceitos errôneos, mesmo que de forma não intencional.<sup>3536</sup>

Essa presença da alta probabilidade de falso negativo em artigos de revistas de grande relevância ratifica a afirmação do renomado cientista, John Ioannidis, de que

a maior parte dos trabalhos publicados têm resultados falsos <sup>3737</sup>. Esses trabalhos, segundo ele podem ser falsos por diversos motivos, dentre eles pela presença de erros sistemáticos, múltiplas testagens de uma mesma hipótese e/ou por poder estatístico e tamanho amostral insuficientes <sup>37</sup>, sendo parte desses fatores, inclusive, presentes na maior parte dos artigos analisados no vigente estudo. As consequências disso são alarmantes e interferem negativamente na vida das pessoas e na imagem da ciência.

Cerca de 1/3 dos brasileiros, entre 2018 e 2019, considerava os cientistas, e conseqüentemente a ciência, a terceira fonte mais confiável de informação, ultrapassados apenas por jornalistas e médicos<sup>38</sup>. Dedutivamente, sendo a maior parte dos autores dos artigos analisados médicos e cientistas essa confiabilidade tende a ser ainda maior. Nesse contexto, a maior parte dos cidadãos brasileiros reconhece a utilidade e a importância da pesquisa, fazendo com que a imagem pública da ciência seja boa e denote confiabilidade<sup>38</sup>, sobretudo em um contexto pós-pandêmico, na qual vacinas foram desenvolvidas em tempo recorde. Mas como seria para as pessoas saber que a maior parte do produto desses pesquisadores é falso <sup>37</sup> e incompatível com a realidade?!

A ciência vive de sua credibilidade e do seu potencial em fazer diferença na vida das pessoas, haja vista poder orientar racionalmente as suas ações<sup>39</sup>. Assim, ao perceber que a maior parte dos artigos tem alta probabilidade de serem falsos, a relação de confiança entre a ciência e a sociedade pode se fragilizar, fazendo com que a divulgação científica verdadeira seja prejudicada e preterida em relação a fontes alternativas de conhecimento, como os argumentos de autoridade. Por consequência a tomada de decisão tenderá a se embasar em um conhecimento cuja validade não tem relação empírica, mas sim com a formação acadêmica prévia de quem o originou, isto é, condutas serão embasadas no que é aparentemente correto (ou até mesmo, apenas no que é mais cômodo) e não necessariamente no provável.

Ainda nesse contexto, sendo a tomada de decisão médica respaldada em artigos científicos, tem-se que boa parte da tomada de decisão atual pode estar embasada em falsos conceitos, levando a condutas equivocadas. Tanto falsos positivos, quanto falsos negativos tem consequências perigosas. O primeiro revela consequências

consideradas mais graves, porque condutas que não deveriam ser feitas, passam a ser realizadas em ampla escala<sup>19</sup>. Já no segundo, condutas que deveriam ser realizadas, não são, por aparentarem ser equivocadas, ou seja, os falsos negativos acabam por invisibilizarem soluções.

Assim, por não se saber que determinada conduta seria correta e principalmente em um contexto de emergência, o não desempenho de ações tem consequências graves, afinal, essas condutas poderiam melhorar a qualidade de vida, minimizar os riscos, ou até mesmo salvar um paciente. Ademais, faz com que a ciência continue tentando encontrar respostas já existentes, mas ainda não percebidas, para questões que importunam a saúde de muitas pessoas. Dessa forma, tratamentos úteis são completamente descartados<sup>31</sup>, o que faz com que se perca tempo, investimento financeiro e sobretudo a oportunidade de focar em questionamentos que, de fato, ainda não possuem solução.

Como opção para minimizar esse cenário, além da promoção da cultura de valorização da integridade científica, existe uma necessidade de o ecossistema científico se capacitar e entender as consequências do erro aleatório e como evitá-lo. Isso pode ser feito por meio do desenvolvimento de instrumentos para a avaliação, mensuração e, conseqüentemente, minimização dos erros aleatórios como um todo. Esse instrumento de mensuração poderia ser um checklist que englobasse como variáveis os prováveis preditores de falso negativo, como avaliação do tamanho amostral, poder estatístico, magnitude e variabilidade dos dados, auxiliando os próprios pesquisadores no momento de planejamento de seus trabalhos.

## **6.2 Variáveis exploratórias**

Uma considerável quantidade de artigos não descreveu o cálculo amostral utilizado ou concluiu positivamente sobre desfechos secundários. Esse primeiro denota falta de clareza, suprimindo do leitor a possibilidade de reproduzi-lo<sup>40</sup>, de julgar por completo a sua metodologia, e por consequência confiar ou não nele. Já o segundo pode indicar a supervalorização de trabalhos positivos, culminando na positivação de um objetivo que não pode ser positivado, pela estrutura do artigo não ter sido formatada para tal. Ambas as situações também infringem os princípios da integridade

científica ao não prezarem pela transparência e ao realizarem múltiplas comparações, repercutindo novamente na imagem da ciência e nas condutas médicas.

Ainda nesse contexto, percebe-se que o cenário científico atual tem predileção por resultados positivos<sup>41</sup>, subestimando a utilidade dos negativos<sup>37</sup>. Assim, como é raro que periódicos relevantes publiquem artigos negativos, os pesquisadores têm pouco incentivo para assumi-los<sup>42</sup>. Com isso, foram criadas maneiras de gerar trabalhos positivos, mesmo que o preço pago seja a integridade científica.

Uma dessas maneiras é a múltipla comparação, exemplificada pela conclusão positiva de desfechos secundários. Entretanto, essa prática aumenta a probabilidade de se encontrar um resultado positivo pelo mero efeito do acaso, sendo inclusive uma forma de spin, por moldar a apresentação dos resultados, ao reduzir impactos negativos e/ou aumentar os positivos<sup>43</sup> com base nos interesses dos autores. Logo, essa prática acaba por levar a conclusões provavelmente falsas<sup>44</sup>. Por fim, o resultado do presente estudo pode ratificar esse cenário, já que uma parcela dos artigos, mesmo sendo negativos, geraram conclusões positivas, mesmo que provavelmente errôneas.

### **6.3 Particularidades e limitações**

Outro artigo já questionou sobre a alta probabilidade do erro do tipo II na área médica, sendo ele uma revisão sistemática de ensaios clínicos randomizados. Esse trabalho focou em um nicho específico da área médica, abordando a ocorrência desses erros na literatura sobre traumas ortopédicos<sup>33</sup>, sendo que também evidenciou alta probabilidade de resultados falso negativos. Além disso, o artigo possuiu uma amostra de estudos analisados menor, de respectivamente 117 trabalhos em relação ao atual, com 190.

O presente artigo também se diferencia dos demais por ter como foco, artigos publicados nas melhores revistas médicas do mundo, segundo o Journal Citation Reports (JCR) pela Clarivate em 2022, sendo essas as de maior rigor técnico quanto a qualidade dos artigos que publica. Assim, isso acaba por aumentar o poder estatístico do presente trabalho. Além disso, se as mais relevantes revistas médicas do mundo, publicam artigos com altos índices de provável erro do tipo 2,

provavelmente outras revistas, tendem a ter trabalhos com ainda mais chances de serem falsos.

Deve-se reconhecer as limitações do presente estudo no que concerne à utilização de uma amostra de conveniência. Assim, diante dessa limitação, para melhorar a qualidade dos dados obtidos e ampliar o conceito gerado, é necessário que outros estudos sejam realizados.

## **7 CONCLUSÃO**

O estudo metacientífico foi realizado para analisar o risco de falso-negativo em ensaios clínicos negativos publicados em revistas de grande relevância científica, no período de 2020-2021. Encontrou-se uma quantidade relevante de ensaios clínicos randomizados com resultados potencialmente falso-negativos. Estudos mais robustos na área meta-científica são necessários para aprofundar a evidência nesse âmbito.

## REFERÊNCIAS

1. Galantin DV. Verdade-acontecimento e alteração no pensamento de Michel Foucault. *Revista de Filosofia Aurora* [Internet]. 2019 May 6;31(52):167–90. Available from: <https://periodicos.pucpr.br/index.php/aurora/article/view/24675>
2. Candiotto C. Verdade e diferença no pensamento de Michel Foucault. *Kriterion: Revista de Filosofia* [Internet]. 2007;48(115):203–17. Available from: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-512X2007000100012&lng=pt&tlng=pt](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-512X2007000100012&lng=pt&tlng=pt)
3. Fletcher RHSWF. *Epidemiologia Clínica: Elementos Essenciais*. 5ª edição. 2014.
4. Stephen B. Hulley, Steven R. Cummings, Warren S. Browner, Deborah G. Grady, Thomas B. Newman, Michael Schmidt Duncan AGI. *Delineando a Pesquisa Clínica*. 4ª edição. Artmed; 4ª edição (12 dezembro 2014); 2014. 400 p.
5. Last JM. A dictionary of epidemiology. *Int J Epidemiol*. 1986;15(2):277.
6. Fregni F, Illigens BMW. *Critical Thinking in Clinical Research: Applied Theory and Practice Using Case Studies*. 2018. 537 p.
7. Akobeng AK. Assessing the validity of clinical trials. *J Pediatr Gastroenterol Nutr*. 2008;47(3):277–82.
8. Andrade MM De. *INTRODUÇÃO À METODOLOGIA DO TRABALHO CIENTÍFICO*.
9. Luis F, Moncayo G. *Principles and Practice of Clinical Trial Medicine*.
10. Delgado-Rodríguez M, Llorca J. Bias. *J Epidemiol Community Health* (1978). 2004;58(8):635–41.
11. Sackett DL. Bias in analytic research. *J Chronic Dis*. 1979;32(1–2):51–63.
12. Petrie A, Sabin C. *Medical Statistics at a Glance*. Medical Statistics at a Glance. 2010. 79–81 p.
13. Akobeng AK. Understanding type I and type II errors, statistical power and sample size. *Acta Paediatrica, International Journal of Paediatrics*. 2016;105(6):605–9.
14. Banerjee A, Chitnis U, Jadhav S, Bhawalkar J, Chaudhury S. Hypothesis testing, type I and type II errors. *Ind Psychiatry J*. 2009;18(2):127.
15. Hakan T. Philosophy of science and black swan. *Child's Nervous System*. 2021;(January).
16. Daniel WW. *BIostatistics Analysis, A Foundation for in the Health Sciences*. JOHN WILEY & SONS, INC; 2009. 192 p.
17. Yiannoutsos C, Ph D. *Principles of Biostatistics Class notes to accompany the textbook by Pagano and Gauvreau*.
18. Ward MM. Primer: Measuring the effects of treatment in clinical trials. *Nat Clin Pract Rheumatol*. 2007;3(5):291–7.
19. Petraitis PS. Intuitive Biostatistics. Harvey Motulsky . Vol. 72, *The Quarterly Review of Biology*. 1997. 239–240 p.
20. Sedgwick P. Sample size: How many participants are needed in a trial? *BMJ (Online)*. 2013;346(7896):1–2.
21. Beck RW. Sample Size for a Clinical Trial: Why Do Some Trials Need Only 100 Patients and Others 1000 Patients or More? *Ophthalmology*. 2006;113(5):721–2.
22. Sousa MDR de. *Principais medidas de magnitude do efeito utilizadas na comparação de dois grupos*. 2018;

23. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS; 1988.
24. Cohen J. *Statistical Power Analysis*. *Curr Dir Psychol Sci*. 1992;1(3):98–101.
25. Noordzij M, Tripepi G, Dekker FW, Zoccali C, Tanck MW, Jager KJ. Sample size calculations: Basic principles and common pitfalls. *Nephrology Dialysis Transplantation*. 2010;25(5):1388–93.
26. Bielik L. *Methodology of Science: An Introduction*. 1st editio. Comenius university in bratislava ; 2019. 232 p.
27. Provdanov CC, Freitas EC De. *Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico* [Internet]. Novo Hamburgo: Feevale. 2013. 276 p. Available from: <http://www.feevale.br/Comum/midias/8807f05a-14d0-4d5b-b1ad-1538f3aef538/E-book Metodologia do Trabalho Cientifico.pdf>
28. Sun YZ, Wang WN, Sun A, Cui PC, Chen WX, Li GZ, et al. External validity of randomised controlled trials: “To whom do the results of this trial apply?” *Chinese Journal of Clinical Rehabilitation*. 2004;8(1):94–5.
29. Patino C, Ferreira J. Internal and external validity: can you apply research study results to your patients? *J Bras Pneumol*. 2018;44(3):183.
30. Lieberman MD, Cunningham WA. Type I and Type II error concerns in fMRI research: Rebalancing the scale. *Soc Cogn Affect Neurosci*. 2009;4(4):423–8.
31. Rogatko A, Litwin S. Phase II Studies : Which Is Worse , False Positive or False Negative ? When comparing treatments in phase Are Time or Intensity Factors Important to the Definition of Metastases of Unknown Origin ? *Notes. Cancer*. 1996;88(7):20892–20892.
32. Brown GW. Errors, Types I and II. *American Journal of Diseases of Children*. 1983;137(6):586–91.
33. Lochner H V., Bhandari M, Tornetta P. Type-II error rates (beta errors) of randomized trials in orthopaedic trauma. *Journal of Bone and Joint Surgery - Series A*. 2001;83(11):1650–5.
34. Bailey CS, Fisher CG, Dvorak MF. Type II Error in the Spine Surgical Literature. *Spine (Phila Pa 1976)* [Internet]. 2004 May;29(10):1146–9. Available from: <http://journals.lww.com/00007632-200405150-00018>
35. Luiz Henrique Lopes dos Santos. *Sobre a integridade ética da pesquisa* [Internet]. FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo. 2011 [cited 2023 Mar 8]. Available from: <https://fapesp.br/6566/sobre-a-integridade-etica-da-pesquisa>
36. Fraga J, Pró-Reitor De Pesquisa T, Roesler R, Pró-Reitor V, Bruno P, Neto C, et al. *Guia para Integridade em Pesquisa Científica*.
37. Ioannidis JPA. Why most published research findings are false. In: *Getting to Good: Research Integrity in the Biomedical Sciences*. Springer International Publishing; 2018. p. 2–8.
38. *Percepção Pública da C&T no Brasil-2019* [Internet]. Available from: <http://www.cgee.org.br>
39. Henrique Lopes dos Santos L. *soBrE A intEgridAdE éticA dA pEsquisA \**.
40. Kretser A, Murphy D, Bertuzzi S, Abraham T, Allison DB, Boor KJ, et al. *Scientific Integrity Principles and Best Practices: Recommendations from a Scientific Integrity Consortium*. *Sci Eng Ethics*. 2019 Apr 15;25(2):327–55.
41. Rabelo DR. *PREDILEÇÃO POR RESULTADOS POSITIVOS: “O ELEFANTE NA SALA” DA INTEGRIDADE CIENTÍFICA EM ESTUDOS MÉDICOS TESE DE DOUTORADO SALVADOR 2021*. 2021.
42. Simmons JP, Nelson LD, Simonsohn U. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol Sci*. 2011;22(11):1359–66.

43. Boutron I, Ravaud P. Misrepresentation and distortion of research in biomedical literature. *Proc Natl Acad Sci U S A*. 2018;115(11):2613–9.