



SERIES IN INTENSIVE MEDICINE: METHODOLOGICAL UPDATE IN MEDICINA INTENSIVA

## Frequent methodological errors in clinical research<sup>☆</sup>



L.C. Silva Aycaguer

Escuela Nacional de Salud Pública de Cuba, Ciudad de La Habana, Cuba

Received 2 November 2017; accepted 25 December 2017

Available online 17 October 2018

### KEYWORDS

Clinical trials;  
Statistical  
significance tests;  
Fraud;  
Selection bias;  
Sample size;  
Predatory journals

**Abstract** Several errors that are frequently present in clinical research are listed, discussed and illustrated. A distinction is made between what can be considered an “error” arising from ignorance or neglect, from what stems from a lack of integrity of researchers, although it is recognized and documented that it is not easy to establish when we are in a case and when in another. The work does not intend to make an exhaustive inventory of such problems, but focuses on those that, while frequent, are usually less evident or less marked in the various lists that have been published with this type of problems. It has been a decision to develop in detail the examples that illustrate the problems identified, instead of making a list of errors accompanied by an epidermal description of their characteristics.

© 2018 Elsevier España, S.L.U. and SEMICYUC. All rights reserved.

### PALABRAS CLAVE

Ensayos clínicos;  
Pruebas de  
significación  
estadística;  
Fraude;  
Sesgo de selección;  
Tamaño muestral;  
Revistas  
depredadoras

### Errores metodológicos frecuentes en la investigación clínica

**Resumen** Se enumeran, discuten e ilustran diversos errores que se presentan con frecuencia en la investigación clínica. Se hace una distinción entre lo que puede considerarse un «error» surgido de la ignorancia o el descuido, de aquello que dimana de una falta de integridad de los investigadores, aunque se reconoce y documenta que no es fácil establecer cuándo estamos en un caso y cuándo en otro. El trabajo no se propone hacer un inventario exhaustivo de tales pifias, sino que se concentra en aquellas que, sin dejar de ser frecuentes, suelen ser menos evidentes o menos señaladas en los diversos listados que se han publicado con este tipo de problemas. Se ha optado por desarrollar en detalle los ejemplos que ilustran los problemas señalados en lugar de hacer una relación de errores acompañados de una descripción epidérmica de sus características.

© 2018 Elsevier España, S.L.U. y SEMICYUC. Todos los derechos reservados.

<sup>☆</sup> Please cite this article as: Silva Aycaguer LC. Errores metodológicos frecuentes en la investigación clínica. Med Intensiva. 2018;42:541–546.

E-mail address: [lcsilva@infomed.sld.cu](mailto:lcsilva@infomed.sld.cu)

## Introduction

In the opening years of this century, the epidemiologist and statistician of Greek origin and professor at Stanford University, John Ioannidis, published an article<sup>1</sup> in which he argued that contemporary research in the field of healthcare mostly comes to wrong conclusions. A claim of this magnitude inevitably stirred the scientific community and generated enormous interest. In effect, as the author himself explains in a brief interview that can be accessed on YouTube (<https://www.youtube.com/watch?v=KOZAV9AvIQE>), far over one million people have read the article. In fact, it has been the most frequently consulted publication in the prestigious open access journal *PLOS Medicine* since its creation in 2004.

Eleven years later, Ioannidis has published a complementary article in the same journal<sup>2</sup> that proves just as disturbing as the first, if not more so. The article, entitled: "Why most clinical research is not useful", states that "In general, not only are most research findings false, but even worse, most of the findings that are true are not useful".

There are many causes underlying this problem, though in principle it is obvious that neither authors nor editors wish to make mistakes or engage in futile efforts. This worrisome reality has caused the journal *MEDICINAINTENSIVA* to promote an analysis of the errors that frequently affect clinical studies or their divulgation. The present contribution seeks to examine some of these errors.

Appropriation of the findings of others, deliberate omissions, or manipulating interpretations cannot be regarded as "errors". Those are simply examples of misbehavior deserving direct repulsion. Unfortunately, the recent history of clinical research is plagued by behaviors of this kind seeking to ensure that certain procedures receive commercial success even if they are useless, if not directly harmful. The evidence in this respect is abundant and undeniable. Perhaps one of the best exponents of this is the book published by Peter Gøtzsche,<sup>3</sup> director of the *Nordic Cochrane Center*, entitled "Deadly medicines and organized crime. How big pharma has corrupted health care", which has recently been published in Spain.

According to PubMed, out of a total of 2047 scientific articles that were disqualified following publication (the so-called "retractions") over the last four decades, only one-fifth were attributable to errors.<sup>4</sup> The rest represented cases of fraud or suspected fraud.

A study published in 2015 and involving 57 controlled clinical trials carried out between 1998 and 2013 reported that the United States Food and Drug Administration (FDA) had discovered that 22 (almost 40%) provided false information; 35 (61%) contained inadequate or inexact registries; one out of every four failed in reporting adverse reactions; three out of every four contained breaches in protocol; and more than 50% did not guarantee adequate patient protection (problems referred to safety or informed consent).<sup>5</sup> The respective publications made no mention of such malpractice, and the truth is that it is difficult to know whether such shortcomings were attributable to deliberate decisions or not.

We understand that when the editors of *MEDICINAINTENSIVA* decided to dedicate journal space to address commonly made errors, they did not intend to alert to data

falsification or plagiarism, but rather aimed to draw attention to those errors that occasionally may result from out of date information or neglect. Still, it must be underscored that the dividing line between error and fraudulent behavior is thin.

In the following lines we will use the term "error" in reference to certain practices that may well carry false intentions, though not necessarily so. In any case, it would be useful for intensivists to avoid blunders that could be regarded as academic fraud, as these may generate doubts that cannot always be unequivocally refuted.<sup>6</sup> For this reason, some of the examples which the reader will find in this article may correspond to either misbehavior or to innocent consequences of ignorance or a lack of awareness of the requirements of good research practice.<sup>7</sup>

In any case, this article does not intend to offer an exhaustive inventory of the possible errors. The literature contains a series of such inventories,<sup>8-12</sup> the most exhaustive of which (describing over 50 common errors) has been published by Strasak et al.<sup>13</sup> Since these sources can be easily consulted, we will not examine the most repeated errors but rather the least evidenced and most controversial errors – though this does not mean that they are any less frequent. Furthermore, we have decided to address each problem in some detail rather than to examine many problems only superficially.

A description and comment are provided below of 6 errors that meet the aforementioned criteria and which we have chosen in view of their relevance and persistent appearance in publications.

## Clinical irrelevance and social irresponsibility

A study<sup>14</sup> based on 704 articles submitted in 2000–2001 to two top level journals (the *British Medical Journal* and the *Annals of Internal Medicine*) for possible publication showed that 514 (73%) had received the support of an expert in methodology and statistics. The rejection rate was 71% for those articles that lacked such support versus only 57% for those with such support. These data appear to confirm the advantage of having a statistician among the signing authors. Unfortunately, many health professionals appear to harbor the absurd conviction that any investigator "can" (or even worse "must") master statistics as if it were a mere question of basic grammar. However, the true mastery of statistical techniques requires highly specialized knowledge. This may explain a widespread practice in this field that deserves special mention: the wrong use of statistical significance tests (SSTs).<sup>15</sup>

The "p-value" criteria are varied and diverse, and many studies have been dedicated to establish them. Decades ago, the International Committee of Medical Journal Editors<sup>16</sup> explicitly rejected the use of "p-values" neglecting the estimation of effects. In order to gain a more in-depth perspective of the highly negative consequences of not paying attention to this recommendation, mention should be made of the current declarations of 72 outstanding investigators in different fields,<sup>17</sup> as well as of a recent detailed analysis on the subject.<sup>18</sup>

We will focus on the most serious consequence of the non-critical and unwarranted use of SSTs: the generalized and

systematic attribution of clinical (or public health) importance to therapeutic or preventive procedures that are actually very far from having any such importance.

Imagine that we are comparing recovery rates referred to a given disease following the administration of two treatments (e.g., conventional treatment versus novel therapy), and that on applying a hypothesis test we obtain a “*p*-value” of under 0.05. What we could affirm is that there is sound evidence for considering that the two treatments are not equal. No more no less. However, there is a generalized habit of interpreting this result as meaning that “the difference between the treatments is significant”. This actually constitutes a misuse of language, since we are implicitly suggesting that “the difference is relevant or substantial”.

In order to confirm that this is no mere problem of semantics, and to illustrate the incapacity of “*p*-values” to convert trivial effects into sound generic recommendations, we will examine a particularly expressive example in the form of a clinical trial published in *The Lancet* on the preventive usefulness of two platelet inhibitors: Plavix® (clopidogrel) and Aspirin® (acetylsalicylic acid).<sup>19</sup>

The mentioned study “demonstrates” that Plavix® (marketed by Sanofi, the company that sponsored the trial) is more effective than Aspirin® in reducing serious vascular events in patients at a high risk of suffering such events.

The study population consisted of patients with a recent history of myocardial infarction (occurring within the previous 35 days), or with a history of ischemic stroke in the 6 preceding months. Following randomized allotment, the two treatments were administered to about 20,000 individuals meeting the aforementioned criteria. The study endpoint was the occurrence of repeat ischemic stroke or myocardial infarction (whether fatal or otherwise), or any death of vascular origin.

The megastudy estimated a risk of D1 = 583 events per 10,000 subjects-year of treatment with Aspirin® and D2 = 532 events per 10,000 subjects-year of treatment with Plavix®. After applying an SST, the study triumphantly registered  $p = 0.043$  and concluded that Plavix® has greater preventive capacity than Aspirin®.

It is obvious that the overall high risk patients would spare themselves a mere D1 – D2 = 51 events per 10,000 subjects-year of preventive treatment when using Plavix® instead of Aspirin®.

Consequently, a rational decision-maker in principle would refrain from recommending Plavix® (a much more expensive drug). However, even if the mentioned difference were considered “important”, it clearly would make no sense at all to cite the *p*-value in the recommendation made. Any decision can (and must) be based on the D1 and D2 values, accompanied by other information such as the cost and adverse effects. The only role of the *p*-value in the article (and subsequently in its many citations) is to provide coverage for those who manufacture Plavix®, taking advantage of the non-critical use of the statistics and SST, along with the error (or trick<sup>20</sup>) of omitting the true significance in practical terms of the product they are recommending.

This is no isolated example. The SST “dictatorship” is still fully present in the biomedical literature. The Finnish professor Esa Läärä wrote: “The reporting of trivial and meaningless “results” that do not afford adequate quantitative information of scientific interest is a very widespread

practice”.<sup>21</sup> A couple of examples suffice to confirm this. An analysis of 71 anticancer drugs consecutively approved for the treatment of solid tumors between 2002 and 2012<sup>22</sup> found the overall survival and disease-free interval of the drugs to be 2.1 and 2.3 months, respectively.

The standards of the American Society of Clinical Oncology for considering that a drug offers “significant clinical benefit” are clearly not related to “statistical significance”. Between April 2014 and February 2016, the Food and Drug Administration approved 46 antineoplastic drugs. A recent article published in a specialized journal<sup>23</sup> showed that only 9 (19%) met these standards.

## Unacceptable generalizations

The results of a clinical trial cannot be abruptly applied to individuals that fall outside the context of the sample on which the trial is based. All studies must have a series of inclusion criteria (those specifying which subjects can be incorporated to the sample representing the study population) and exclusion criteria (those specifying the characteristics of the subjects that cannot be incorporated to the sample). For example, it may be established that a study will include non-asthmatic adults and cannot include pregnant women. If the study drug proves successful after one month of treatment of a certain illness, it should not be recommended in pregnant women or asthmatic individuals, and the drug should not be used for any period of time other than one month (no longer, no shorter).

Obviously, the way in which a given outcome is applied is not the responsibility of those who obtained the outcome. However, it is not infrequent to make the mistake of not reporting such restrictions in the abstract and conclusions of an article. Furthermore, since these are the sections of the publication that are usually read most carefully, there is an implicit invitation for the reader to apply the results or outcomes beyond what is actually permissible.

## Selection bias: the Vioxx® tragedy

Rofecoxib, a powerful analgesic marketed by Merck under the brand name Vioxx® and used by about 84 million people worldwide, appeared on the market in 1999 in over 80 countries. It was used to treat the symptoms of osteoarthritis and rheumatoid arthritis, acute pain and menstrual pain. In the year 2000, *The New England Journal of Medicine* reported on the VIGOR study, involving 8076 patients, of which one-half ( $n = 4029$ ) were randomized to 500 mg of naproxen (an old nonsteroidal anti-inflammatory drug) twice a day, while the rest ( $n = 4074$ ) received 50 mg of rofecoxib once a day.<sup>24</sup>

The endpoint was any gastrointestinal adverse effect, precisely because the purported advantage of the new drug was that it allowed the treatment of pain in chronic disorders without causing symptomatic gastric ulcers or gastrointestinal bleeding – these being undesirable effects typical of the classical analgesics. The results were strongly favorable to Vioxx®.

As became known much later, the Vice-Chairperson of clinical research at Merck, and a signing author of the mentioned study, had proposed in 1997, in an internal docu-

ment of the company, that the intended clinical trial should exclude people with prior cardiovascular problems, “to ensure that the differences in complications between those receiving Vioxx<sup>®</sup> and those receiving a classical analgesic would not be evident”. In other words, Merck was not only already aware of the possible cardiac toxicity of the new drug but also planned selection bias in order to prevent such complications from appearing during the trial.

In an interview by a British newspaper,<sup>25</sup> David Gram, an investigator of the Food and Drug Administration, the lethal influence in terms of heart attacks caused “the greatest drug-related catastrophe in the history of Medicine: between 89,000 and 139,000 American patients could have died or suffered cardiac disorders as a consequence of the drug”. In view of this situation, on 30 September 2004, Merck announced the withdrawal of its drug from the world market.

In this painful yet instructive episode, the company planned and introduced selection bias (one of the incidents related to Vioxx<sup>®</sup> and for which Merck subsequently faced a very heavy fine). However, what should be underscored here is that when such bias is made (even inadvertently), the error can have very serious consequences.

### Myths and mistakes referred to required sample size

It is essential to understand something which is usually forgotten by those who theorize about sample size and stick to orthodox and often schematic and ritual positions.<sup>26,27</sup> Ultimately, and practically regardless of the sample size, a study can be potentially useful. The information available for precisely determining a correct sample size (an inherently polemical term) is often inexistent or falls short of the required information, and in any case it is inexorably conditioned by speculation inherent to the choice of data to be used in the calculation.<sup>28</sup>

In some cases the sample size is stated to be “insufficient”. This evaluating criterion remains strongly implanted in Ethics Committees, research project evaluating agencies, and editors of scientific journals. However, the legitimacy of such statements is very questionable, since they are conflictive and non-pertinent for at least two reasons.

Firstly, because the formulas used to determine sample size are intrinsically speculative and inevitably imply subjectiveness.<sup>29</sup>

Secondly, because the term “sufficiently large sample” would only make sense if operative standards or firm conclusions could be drawn from each individual study. However, this delusion is as ingrained as it is incorrect, due to the simple reason that science does not work that way.<sup>30,31</sup>

Our scientific convictions may be more or less firm, but they are always provisional, and while our representations of reality may have a degree of credibility at a given point in time, they are open to changes and improvement as new data emerge. The consolidation of new knowledge is gradual, and all contributions are welcome. Some contributions will be more relevant and others less so, but they all add something to the process, independently of the sample size, and generally through the conduction of meta-analyses.<sup>32</sup>

Clinical research errors are not only made by those who perform the research, but also by those who disclose the results. As an example, in our opinion it is a mistake to demand authors to include the formulas used, and in general to explain their sample sizes. What authors usually do in the face of such demands is what an important article published in *The Lancet*<sup>33</sup> called “retrofitting” (i.e., choosing values that lead to a sample size that has already been chosen beforehand), and which results in a certain lack of integrity on the part of the authors. The problem is controversial, since the idea that it is necessary to explain the sample size is very ingrained, even though irrational.<sup>34</sup>

### The “negative studies” iceberg

A curious journal called *Negative Results*<sup>35</sup> has recently been created with the aim of publishing articles that refute or challenge knowledge taken to be true. The initiative is a reaction against the stigma associated with so-called “negative findings”.

So-called “publication bias”, a term first used around 1980,<sup>36</sup> is the tendency to favor the publication of certain articles conditioned to the results they obtain and not necessarily to their merits. The basic reason for this is the reluctance of many editors of medical journals to publish things that make no “novel” contribution, as if the rebuttal of a wrong conviction were not “novel”.

Since then there has been growing awareness of the problem, known as the “file drawer problem”<sup>37</sup>: journals publish a small fraction of the studies that are made (whether useful or not), while most of the remaining studies (those with negative results, i.e., studies failing to yield *significance*) are confined to the laboratory drawers. As commented in the presentation of the aforementioned journal: “scientific findings are like an iceberg, with the 10% of published findings floating on the 90% of studies with negative results”.

The most serious problem associated with this type of bias, denounced time and time again,<sup>38</sup> is that it distorts the view we have of reality and therefore compromises the capacity to make correct decisions. Each study may be objective, but our global view no longer remains objective if part of the evidence is suppressed. Naturally, one of the areas most seriously limited by this problem is meta-analysis, the declared purpose of which is to formalize the identification of consensus. Perhaps the most spectacular example in recent times is that of the management of the influenza A pandemic of 2009. As reported in an excellent study that details this episode,<sup>39</sup> in response to the threat, the use of two antiviral agents, Tamiflu<sup>®</sup> (oseltamivir) and Relenza<sup>®</sup> (zanamivir), was recommended. Many governments spent millions of euros on these products.

The decision was warranted by a review of antiviral drugs carried out in 2006 and fundamentally based on a study from 2003, grouping 10 studies (all financed by Roche, the company that markets Tamiflu<sup>®</sup>), and which *demonstrated* their effectiveness and safety.<sup>40</sup> However, many doubts were raised regarding these supposed properties, and scientific and political pressure finally forced the drug companies to share the raw data corresponding to the clinical trials with antivirals, which had been kept under key until then with the argument that they formed part of commercial secrecy. First

Glaxo (in relation to Relenza<sup>®</sup>) and a few months later Roche (Tamiflu<sup>®</sup>) submitted the mentioned data to the authors of a new review in 2013. The enormous task of examining about 160,000 pages led to the conclusion that antiviral agents in fact have very modest efficacy in alleviating the symptoms, and no impact in terms of a decrease in complications or deaths. Furthermore, such drugs were seen to produce important adverse effects and proved unable to modify the mechanisms of contagion.<sup>41</sup> Likewise, it was found that many negative results had been concealed, the disclosure of desired findings was favored, and negative results were hidden – thereby leading to biased conclusions from the start.

In relation to the problem of suppressing negative results, we firmly believe our readers will enjoy the conference given by Dr. Ben Goldacre, an investigator at the London School of Hygiene and Tropical Medicine, which can be accessed on YouTube ([http://www.ted.com/talks/ben\\_goldacre\\_battling\\_bad\\_science?language=en](http://www.ted.com/talks/ben_goldacre_battling_bad_science?language=en)), in view of its clearness and amenity.

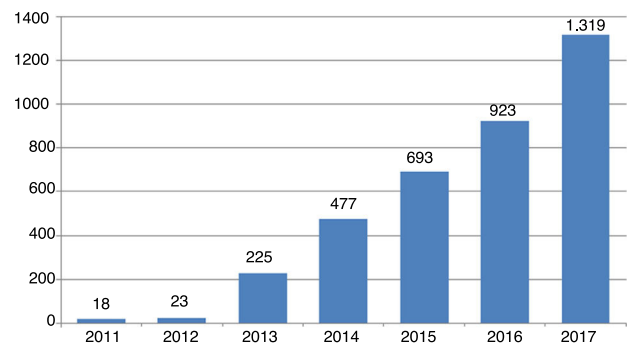
### Predatory editorial practices

As investigators, we routinely receive invitations to submit articles to certain journals that are presumably very interested in our scientific contributions. These are very often so-called “predatory journals”: a lucrative and fraudulent phenomenon characterized by features such as promising the rapid publication of articles, the offering of tempting prices, commitment to ensure strict peer reviews which in fact never occur or are extremely weak, the self-attribution of false impact factors, or the inclusion of inexistent academicians on their editorial boards.

In this regard, mention must be made of the work of Jeffrey Beall, a librarian at the University of Colorado, who is famous for having coined the abovementioned term and for having published lists of “potential, possible or probable predatory open access journals and printing houses” in the past. Curiously, in January 2017 Beall canceled the website where he regularly updated his lists.<sup>42</sup> His only explanation was that he had been threatened. However, a team that preserves its anonymity, precisely in order to avoid such problems, has continued his task, and keeps the list up to date (<https://predatoryjournals.com/>) for consultation by anyone who wishes to do so.

An idea of the magnitude of the phenomenon can be gained from the trend in number of new journals of this kind since their first appearance in 2011: from 18 journals that year to 23 one year later, and since then the rise has been exponential, reaching 1319 journals in October 2017 (Fig. 1).

The great volume of polluted information this academic plague generates is evident. However, the reason for mentioning it here is that we consider it a mistake (and not a minor one at that) to submit articles to journals of this kind. Doing so not only contributes to further false editorial practices but also has an impact upon the personal prestige of those who are foolishly trapped by such journals.<sup>43</sup>



**Figure 1** Evolution of the number of predatory journals between 2011 and 2017.

### Conclusions

There are many and diverse possible methodological errors in clinical research. Our study examines a small group of them in the hope of alerting intensivists to mistakes that are not always repaired. Improper statistical processing, the omission of crucial elements, inertial behaviors and the lack of criterion in publishing have been described using examples. While not always intentional, such errors are equally damaging to medical science, which has been experiencing moments of crisis ever since John Ioannidis shook the scientific community with his penetrating contributions cited at the start of this paper.

### Conflicts of interest

The author declares that he has no conflicts of interest.

### References

- Ioannidis JPA. Why most published research findings are false. *PLoS Med.* 2005;2:e124.
- Ioannidis JPA. Why most clinical research is not useful. *PLoS Med.* 2005;13:e1002049.
- Gøtzsche PC. *Deadly medicines and organized crime. How big pharma has corrupted health care.* London: Radcliffe Publishing Ltd.; 2013.
- Fang F, Steen G, Casadevall A. Misconduct accounts for the majority of retracted scientific publications. *Proc Natl Acad Sci U S A.* 2012;109:17028–33.
- Mayor S. Research misconduct found by FDA inspections of clinical trials is rarely reported in journal studies. *BMJ.* 2015;350:h747.
- Nath SB, Marcus SC, Druss BG. Retractions in the research literature: misconduct or mistakes? *Med J Aust.* 2006;185:152–4.
- Steen RG. Retractions in the scientific literature: do authors deliberately commit research fraud? *J Med Ethics.* 2011;37:113–7.
- Freemantle N, Rait G. Catalogue of errors in papers reporting clinical trials. *BMJ.* 2015;351:h4843, <http://dx.doi.org/10.1136/bmj.h4843>.
- Clark GT, Mulligan R. Fifteen common mistakes encountered in clinical research. *J Prosthodont Res.* 2011;55:1–6.
- Kim MY, Goldberg JD. The effects of outcome misclassification and measurement error on the design and analysis of therapeutic equivalence trials. *Stat Med.* 2000;20:2065–78.

11. Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. *Designing clinical research*. 3rd ed. Philadelphia: Lippincott, Williams and Wilkins; 2007. p. 51–63.
12. Henderson WG. Mistakes in clinical trials. In: Itani K, Reda D, editors. *Clinical trials design in operative and non operative invasive procedures*. Cham: Springer; 2017.
13. Strasak AM, Zaman Q, Pfeiffer KP, Göbel G, Ulmer H. Statistical errors in medical research – a review of common pitfalls. *Swiss Med Wkly*. 2007;137:44–9.
14. Altman PG, Goodman NS, Schroter S. How statistical expertise is used in medical research. *JAMA*. 2002;287:2817–20.
15. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, *P* values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol*. 2016;31:337–50.
16. ICMJE. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals [Internet]; 2016. Available from: <http://www.icmje.org/icmje-recommendations.pdf> [accessed October 2017].
17. Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers EJ, Berk R, et al. Redefine statistical significance. *Nat Hum Behav*. 2018;2:6–10, <http://dx.doi.org/10.1038/s41562-017-0189-z>.
18. Silva-Aycaguer LC. Las pruebas de significación estadística: seis décadas de fuegos artificiales. *Rev Fac Nac Salud Publica*. 2016;34:205–12. Available from: [http://lcsilva.sbhac.net/Articulos/75.Las\\_PSE\\_6\\_decadas\\_de\\_fuegos\\_artificiales\\_RFSAP.pdf](http://lcsilva.sbhac.net/Articulos/75.Las_PSE_6_decadas_de_fuegos_artificiales_RFSAP.pdf) [accessed October 2017].
19. CAPRIE Steering Committee. A randomised, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events. *Lancet*. 1996;348:1329–39.
20. Smith R. The trouble with medical journals. *J R Soc Med*. 2006;99:115–9.
21. Läärä E. Statistics: reasoning on uncertainty, and the insignificance of testing null. *Ann Zool Fennici*. 2009;46:138–57.
22. Fojo T, Mailankody S, Lo A. Unintended consequences of expensive cancer therapeutics—the pursuit of marginal indications and a me-too mentality that stifles innovation and creativity: the John Conley Lecture. *JAMA Otolaryngol Head Neck Surg*. 2014;140:1225–36.
23. Kumar H, Fojo T, Mailankody S. Guidelines for oncology clinical trials. *JAMA Oncol*. 2016;2:1238–40.
24. Bombardier C, Laine L, Reicin A, Shapiro D, Burgos R, Davis B, et al., VIGOR Study Group. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. *N Engl J Med*. 2000;343:1520–8.
25. Ortín A. El científico que recomendó retirar Vioxx advierte sobre miles de muertes. *El País*; 2005.
26. Thompson B. In praise of brilliance: where that praise really belongs. *Am Psychol*. 1998;53:799–800.
27. Bacchetti P. Current simple size conventions: flaws, harms and alternatives. *BMC Med*. 2010;8:17.
28. Fernandes-Taylor S, Hyun JK, Reeder RN, Harris AH. Common statistical and research design problems in manuscripts submitted to high-impact medical journals. *BMC Res Notes*. 2011;4:304.
29. Silva LC, Benavides A. Apuntes sobre subjetividad y estadística en la investigación. *Rev Cubana Salud Publica*. 2003;29:170–3.
30. Goodman SN, Altman DG, George SL. Statistical reviewing policies of medical journals: caveat lector? *J Gen Intern Med*. 1998;13:753–6.
31. Marks HM. Rigorous uncertainty: why RA Fisher is important. *Int J Epidemiol*. 2003;32:932–7.
32. Sackett DL, Cook DJ. Can we learn anything from small trials? *Ann N Y Acad Sci*. 1993;703:25–31.
33. Grimes DA, Schulz KF. Descriptive studies: what they can and cannot do. *Lancet*. 2002;359:145–9.
34. Silva LC, Alonso P. Explicación del tamaño muestral empleado: una exigencia irracional de las revistas biomédicas. *Gac Sanit*. 2013;27:53–7.
35. Negative Results [Internet]. Available from: <http://www.negative-results.org/index.php/about-our-company/> [accessed October 2017].
36. Smith ML. Publication bias and meta-analysis. *Eval Educ*. 1980;4:22–4.
37. Rosenthal R. The ‘file drawer problem’ and tolerance for null results. *Psychol Bull*. 1979;86:638–41.
38. Ioannidis JP. Journals should publish all “null” results and should sparingly publish “positive” results. *Cancer Epidemiol Biomark Prev*. 2006;15:186.
39. Gervas J. El asunto Tamiflú/Relenza, la salud pública y algunas lecciones para la decisión y la ética. *Rev Cubana Salud Publica*. 2014;40:334–48.
40. Cohen D, Carter P. WHO and the pandemic flu “conspiracies”. *BMJ*. 2010;340:e2912.
41. Heneghan CJ, Onakpoya I, Thompson M, Spencer EA, Jones M, Jefferson T. Zanamivir for treatment and prevention of influenza in adults and children: systematic review of clinical study reports. *BMJ*. 2014;348:g2547.
42. Straumsheim C. No more ‘Beall’s List’. *Inside Higher Ed*; 2017. Available from: <https://www.insidehighered.com/news/2017/01/18/librarians-list-predatory-journals-reportedly-removed-due-threats-and-politics> [accessed October 2017].
43. Silva LC. Las editoriales depredadoras: una plaga académica. *Rev Cub Inform Cien Sal*. 2016;27. Available from: <http://www.acimed.sld.cu/index.php/acimed/article/view/996/608> [accessed October 2017].