Evidence-based medicine: Friend and foe

Christian Funck-Brentano

EDITORIAL

Evidence-based medicine: friend and foe.

Christian Funck-Brentano, MD, PhD[1,2]

[1] AP-HP Sorbonne Université, Pitié-Salpêtrière University Hospital, Department of Pharmacology and Clinical Investigation Center, F-75013 Paris, France

[2] INSERM, CIC-1901 and UMR-S 1166, Sorbonne University, Faculty of Medicine, F-75013 Paris, France

Correspondence to: Professor Christian Funck-Brentano

Clinical Investigation Center, Pitié-Salpêtrière Hospital, 47–83 Boulevard de l'Hôpital, 75013 Paris, France

Email: christian.funck-brentano@aphp.fr

1 Table; 54 references

Competing interests: The author declares that he has no competing interest

Funding: No sources of funding were used to assist in the preparation of this manuscript.

Keywords: Clinical study; Conflict of interest; Data accuracy; Endpoint determination; Epidemiology; Evidence-based medicine; Research design.

Short title: Illusions of Evidence-Based Medicine

"There are three kinds of lies: lies, damned lies, and statistics"

(Attributed to Benjamin Disraeli)

There is no doubt that, after millennia of charlatanism in medicine, the advent of evidence-based medicine applied to observational and interventional studies has dramatically improved our critical appraisal of medical information. However, the development of sophisticated statistical techniques, the public demand for medical information (fueled by expanding media with selling needs), pressure from industry's sales forces, a desire for health positivism [1], together with academic pressure on researchers to publish, have resulted in biases in the conception and to spins in the reporting of so-called evidence-based studies, many denoting breaches of scientific integrity [2]. Most of those biases and spins can usually be detected with a thorough reading of articles, including supplementary files, editorials and correspondence published after articles have been made available but this requires time and skills that many physicians and the majority of journalists do not have. It is also difficult to detect whether or not authors have conflicts of interest because the format of disclosures is uninformative [3, 4]. Those conflicts of interest are not limited to financial support from private companies [5] and are often related to cognitive biases, researchers having worked all their life in a specific domain being inclined to have an arrested opinion on this domain that might limit their ability to accept contradiction [6]. All researchers know the importance of testing their hypotheses bilaterally, i.e. with equal chances to demonstrate that what they expect to be true is indeed true or to demonstrate the converse. But some studies, particularly those with commercial stakes, can be conceived to favor a desired result. This can usually be detected from the methods section of published reports, especially, among other criteria, from inclusion and non-inclusion criteria, and from a critical review of study endpoints and how they were

assessed. Ideally a researcher should have no opinion or desire on the outcome of a study he/she intends to conduct and concentrate on designing the study with equal chances to prove or to refute the tested hypotheses. This is one of the many criteria of scientific integrity [2]. Also, it has been shown that, among other financial conflicts of interest [5], the sources of funding and financial support to the principal investigator influence the results of randomized clinical trials [7, 8].

A previous article (freely accessible at https://hal.sorbonne-universite.fr/hal-03097645) discussed several limitations and deceptions of evidence-based medicine [9]. Those included: the use of a composite endpoint where a clinically inconsistent (interventional procedures, hospitalizations) or unreliably measurable (angina pectoris) criterion is combined with a "hard" endpoint (death, stroke, …) and drives a statistically significant difference with a control group; overemphasis of statistical significance when the amplitude of an effect is clinical very small; post-hoc "data-torturing" (and it should be remembered that if you torture data, it will confess [10]); reporting results as changes in relative risk and not as changes in absolute risk, sometimes without any indication of absolute numbers [11]; failing to identify the characteristics of patients who truly benefit from a therapeutic intervention or are at risk of a clinical events associated with exposure to one or several risk factors; omitting to verify or report Sir Bradford Hill's causality criteria [12] in epidemiological studies when associations between exposures and outcomes have been found; and failing to reliably assess endpoints in randomized clinical trials or to verify the quality of information in an epidemiological database. An extended list of such tricks used in evidence-based articles to embellish results is proposed in the table with references to examples of articles or commentaries illustrating or discussing the stratagems.

Although deception in randomized clinical trials and observational studies is not the general rule – and may be unintentional – its existence should make all readers suspicious. The methodology of randomized clinical trials has considerably improved in many disciplines over the last fifty years and comprehensive recommendations have been issued on how to avoid pitfalls from design to reporting to conduct and analysis [37-40]. Randomized clinical trials are less subject to biases – and biases are easier to detect – than observational studies or so-called "real world" researches and they remain the best way to assess medical interventions [41]. However, not all medical questions can be tackled using this type of study and there are situations, e.g. surgical interventions [42-44], where methodological standards cannot be fully complied with. Therefore, epidemiological studies are indispensable.

In 1993, Peter Skrabanek argued that epidemiology led to uncertain results [45]. He suggested that epidemiology met several criteria related to the concept of "pathological science", the "science of things that aren't so", introduced by Nobel Prize winner Irving Langmuir four decades earlier [46]. Among those criteria was the small amplitude of measured effects, requiring large amounts of measurements to reach statistical significance and, despite that, claims for great accuracy. It should be emphasized that excessive confidence in statistical significance is detrimental to a sound interpretation of clinical studies, particularly observational studies which are by nature more vulnerable to biases than randomized clinical trials [47]. This has been very well illustrated with humor by John Oliver on his television show [48]. It should however not be forgotten that observational studies have brought major public health benefits, for example by identifying tobacco smoking as a risk factor for cardiovascular diseases or cancers. Current distrust of epidemiology has arisen from signals of small

amplitude [49] or contradictory studies, for example when studying the relations between cancer risk and nutrition [50] or the use of aspirin in primary cardiovascular prevention [51] and with the advent of studies using so-called “big-data” [49, 52]. One might subjectively consider that the lower and upper boundary of the 95% confidence interval for a hazard ratio of an observational study should be >1.5 or <0.66, respectively, to consider that the result might be clinically meaningful.

Over-communication of medical information is a plague of modern times which is amplified by the growing tendency to publish in the form of preprints [53] and by greedy publishers [54]. The above considerations should of course not be interpreted as a rejection of evidence-based medicine but rather as a reminder of the need to develop a critical sense of medical publications and even more so of their communication by the media.

References

[1] Benjamin DM, Hey SP, MacPherson A, Hachem Y, Smith KS, Zhang SX, et al. Principal investigators over-optimistically forecast scientific and operational outcomes for clinical trials. PLoS One. 2022;17:e0262862.

[2] Anonymous. Intégrité Scientifique. 2022. https://sante.sorbonne-universite.fr/faculte-de-medecine/integrite-scientifique [Accessed August 22nd, 2022].

[3] Demarez JP, Funck-Brentano C, Molimard M, and participants of Round Table n°4 of Giens XXVII. Conflicts of interests in the area of healthcare products and technology. Current state of affairs and recommendations. Therapie. 2012;67:289-94.

[4] Hakoum MB, Noureldine H, Habib JR, Abou-Jaoude EA, Raslan R, Jouni N, et al. Authors of clinical trials seldom reported details when declaring their individual and institutional financial conflicts of interest: a cross-sectional survey. J Clin Epidemiol. 2020;127:49-58.

[5] Jureidini J, McHenry LB. The illusion of evidence based medicine. BMJ. 2022;376:o702.

[6] Navar AM, Fonarow GC, Pencina MJ. Time to Revisit Using 10-Year Risk to Guide Statin Therapy. JAMA Cardiol. 2022;7:785-6.

[7] Ahn R, Woodbridge A, Abraham A, Saba S, Korenstein D, Madden E, et al. Financial ties of principal investigators and randomized controlled trial outcomes: cross sectional study. BMJ. 2017;356:i6770.

[8] Vaduganathan M, Samman-Tahhan A, Patel RB, Kelkar A, Papadimitriou L, Georgiopoulou VV, et al. Association between funding sources and the scope and outcomes of cardiovascular clinical trials: A systematic review. Int J Cardiol. 2017;230:301-3.

[9] Funck-Brentano C. Indispensable but deceptive evidence-based medicine. Diabetes Metab. 2020;46:415-22.

[10] Mills JL. Data torturing. N Engl J Med. 1993;329:1196-9.

[11] Ahn HJ, Lee SR, Choi EK, Rhee TM, Kwon S, Oh S, et al. Protective effect of proton-pump inhibitor against gastrointestinal bleeding in patients receiving oral anticoagulants: A systematic review and meta-analysis. Br J Clin Pharmacol. 2022

[12] Hill AB. The Environment and Disease: Association or Causation? Proc R Soc Med. 1965;58:295-300.

[13] White HD, Schwartz GG, Szarek M, Bhatt DL, Bittner VA, Chiang CE, et al. Alirocumab after acute coronary syndrome in patients with a history of heart failure. Eur Heart J. 2022;43:1554-65.

[14] Woloshin S, Kramer BS. The Increasing Incidence of Early-Onset Colorectal Cancer. N Engl J Med. 2022;387:91-3.

[15] Tavazzi L, Maggioni AP, Marchioli R, Barlera S, Franzosi MG, Latini R, et al. Effect of n-3 polyunsaturated fatty acids in patients with chronic heart failure (the GISSI-HF trial): a randomised, double-blind, placebo-controlled trial. Lancet. 2008;372:1223-30.

[16] Khan MS, Khan MS, Ansari ZN, Siddiqi TJ, Khan SU, Riaz IB, et al. Prevalence of Multiplicity and Appropriate Adjustments Among Cardiovascular Randomized Clinical Trials Published in Major Medical Journals. JAMA Netw Open. 2020;3:e203082.

[17] Collaborators GBDCRF. The global burden of cancer attributable to risk factors, 2010-19: a systematic analysis for the Global Burden of Disease Study 2019. Lancet. 2022;400:563-91.

[18] Vaduganathan M, Jhund PS, Claggett BL, Packer M, Widimsky J, Seferovic P, et al. A putative placebo analysis of the effects of sacubitril/valsartan in heart failure across the full range of ejection fraction. Eur Heart J. 2020;41:2356-62.

[19] Huo X, Armitage J. Use of Run-in Periods in Randomized Trials. JAMA. 2020;324:188-9.

[20] Fox KM, Investigators EUtOrocewPiscAd. Efficacy of perindopril in reduction of cardiovascular events among patients with stable coronary artery disease: randomised, double-blind, placebo-controlled, multicentre trial (the EUROPA study). Lancet. 2003;362:782-8.

[21] Kristensen AMD, Pareek M, Kragholm KH, Sehested TSG, Olsen MH, Prescott EB. Unstable Angina as a Component of Primary Composite Endpoints in Clinical Cardiovascular Trials: Pros and Cons. Cardiology. 2022;147:235-47.

[22] McLeod C, Norman R, Litton E, Saville BR, Webb S, Snelling TL. Choosing primary endpoints for clinical trials of health care interventions. Contemp Clin Trials Commun. 2019;16:100486.

[23] Schuster Bruce C, Brhlikova P, Heath J, McGettigan P. The use of validated and nonvalidated surrogate endpoints in two European Medicines Agency expedited approval pathways: A cross-sectional study of products authorised 2011-2018. PLoS Med. 2019;16:e1002873.

[24] Hilal T, Sonbol MB, Prasad V. Analysis of Control Arm Quality in Randomized Clinical Trials Leading to Anticancer Drug Approval by the US Food and Drug Administration. JAMA Oncol. 2019;5:887-92.

[25] Lenzer J, Brownlee S. Should regulatory authorities approve drugs based on surrogate endpoints? BMJ. 2021;374:n2059.

[26] Ray WA, Murray KT, Meredith S, Narasimhulu SS, Hall K, Stein CM. Oral erythromycin and the risk of sudden death from cardiac causes. N Engl J Med. 2004;351:1089-96.

[27] Li ZH, Zhong WF, Liu S, Kraus VB, Zhang YJ, Gao X, et al. Associations of habitual fish oil supplementation with cardiovascular outcomes and all cause mortality: evidence from a large population based cohort study. BMJ. 2020;368:m456.

[28] Debras C, Chazelas E, Srour B, Druesne-Pecollo N, Esseddik Y, Szabo de Edelenyi F, et al. Artificial sweeteners and cancer risk: Results from the NutriNet-Sante population-based cohort study. PLoS Med. 2022;19:e1003950.

[29] Mehra MR, Vaduganathan M, Fu M, Ferreira JP, Anker SD, Cleland JGF, et al. A comprehensive analysis of the effects of rivaroxaban on stroke or transient ischaemic attack in patients with heart failure, coronary artery disease, and sinus rhythm: the COMMANDER HF trial. Eur Heart J. 2019;40:3593-602.

[30] Wang H, Rosner GL, Goodman SN. Quantifying over-estimation in early stopped clinical trials and the "freezing effect" on subsequent research. Clin Trials. 2016;13:621-31.

[31] Walter SD, Guyatt GH, Bassler D, Briel M, Ramsay T, Han HD. Randomised trials with provision for early stopping for benefit (or harm): The impact on the estimated treatment effect. Stat Med. 2019;38:2524-43.

[32] Valembois L, Audureau E, Takeda A, Jarzebowski W, Belmin J, Lafuente-Lafuente C. Antiarrhythmics for maintaining sinus rhythm after cardioversion of atrial fibrillation. Cochrane Database Syst Rev. 2019;9:CD005049.

[33] Bai AD, Lo CKL, Komorowski AS, Suresh M, Guo K, Garg A, et al. How generalizable are randomized controlled trials (RCTs) in Staphylococcus aureus bacteremia? A description of the mortality gap between RCTs and observational studies. Clin Infect Dis. 2022

[34] Huls H, Abdulahad S, Mackus M, van de Loo A, Roehrs T, Roth T, et al. Inclusion and Exclusion Criteria of Clinical Trials for Insomnia. J Clin Med. 2018;7

[35] Araujo CG, de Souza ESCG, Laukkanen JA, Fiatarone Singh M, Kunutsor SK, Myers J, et al. Successful 10-second one-legged stance performance predicts survival in middle-aged and older individuals. Br J Sports Med. 2022;56:975-80.

[36] Foster JA. Modulating brain function with microbiota. Science. 2022;376:936-7.

[37] Pocock SJ, McMurray JJ, Collier TJ. Making Sense of Statistics in Clinical Trial Reports: Part 1 of a 4-Part Series on Statistics for Clinical Trials. J Am Coll Cardiol. 2015;66:2536-49.

[38] Pocock SJ, McMurray JJV, Collier TJ. Statistical Controversies in Reporting of Clinical Trials: Part 2 of a 4-Part Series on Statistics for Clinical Trials. J Am Coll Cardiol. 2015;66:2648-62.

[39] Pocock SJ, Clayton TC, Stone GW. Design of Major Randomized Trials: Part 3 of a 4-Part Series on Statistics for Clinical Trials. J Am Coll Cardiol. 2015;66:2757-66.

[40] Pocock SJ, Clayton TC, Stone GW. Challenging Issues in Clinical Trial Design: Part 4 of a 4-Part Series on Statistics for Clinical Trials. J Am Coll Cardiol. 2015;66:2886-98.

[41] Collins R, Bowman L, Landray M, Peto R. The Magic of Randomization versus the Myth of Real-World Evidence. N Engl J Med. 2020;382:674-8.

[42] McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. BMJ. 2002;324:1448-51.

[43] Wartolowska K, Collins GS, Hopewell S, Judge A, Dean BJ, Rombach I, et al. Feasibility of surgical randomised controlled trials with a placebo arm: a systematic review. BMJ Open. 2016;6:e010194.

[44] Wallis CJD, Detsky AS, Fan E. Establishing the Effectiveness of Procedural Interventions: The Limited Role of Randomized Trials. JAMA. 2018;320:2421-2.

[45] Skrabanek P. The epidemiology of errors. Lancet. 1993;342:1502.

[46] Langmuir I. Characteristic Symptoms of Pathological Science. 1953. Conference transcript available at https://www.cs.princeton.edu/~ken/Langmuir/langB.htm#Characteristic%20Symptoms [Accessed August 22nd, 2022].

[47] Smith GD. Reflections on the limitations to epidemiology. J Clin Epidemiol. 2001;54:325-31.

[48] Oliver J. Scientific Studies: Last Week Tonight with John Oliver (HBO). 2016. https://www.youtube.com/watch?v=0Rnq1NpHdmw [Accessed August 22nd, 2022].

[49] Leibovici L, Turjeman A, Paul M. The temptation of large numbers. Clin Microbiol Infect. 2018;24:931-2.

[50] Schoenfeld JD, Ioannidis JP. Is everything we eat associated with cancer? A systematic cookbook review. Am J Clin Nutr. 2013;97:127-34.

[51] Shufelt CL, Mora S, Manson JE. Aspirin for the Primary Prevention of Atherosclerotic Cardiovascular Disease in Women. JAMA. 2022;328:672-3.

[52] Mehra MR, Desai SS, Ruschitzka F, Patel AN. RETRACTED: Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis. Lancet. 2020

[53] Flanagin A, Fontanarosa PB, Bauchner H. Preprints Involving Medical Research-Do the Benefits Outweigh the Challenges? JAMA. 2020;324:1840-3.

[54] Siler K, Vincent-Lamarre P, Sugimoto CR, Lariviere V. Predatory publishers' latest scam: bootlegged and rebranded papers. Nature. 2021;598:563-5.

Table: Examples of tricks, spins and data torturing used to embellish the results of clinical researches.

Modifying the ordinate scale to amplify differences between survival curves [13, 14].
Adjusting for baseline group differences in a randomized clinical trial which did not find statistically significant differences with unadjusted data [15].
Performing multiple analyses without adjustment of the p values [16].
Using complex statistical analyses that only a few can understand [17, 18].
Excluding patients from randomization after an open run-in phase is performed to assess patients' tolerance to a new treatment [19, 20].
Using a composite endpoint that includes conditions of uncertain precision (e.g. angina pectoris) or of highly variable clinical significance (e.g. death combined with coronary interventions) [21].
Using a primary endpoint that is irrelevant or cannot be ascertained (e.g. cardiovascular death, angina pectoris) [22, 23].
Choosing a suboptimal control group in a randomized clinical trial (e.g. inappropriate dose; less rigorous assessment) [24].
Using a surrogate endpoint which has not been appropriately validated (i.e. the surrogate endpoint parallels the clinical endpoint and any intervention that changes the surrogate endpoint changes the clinical endpoint proportionately) [25].
Studying a very large population in an observational study leading to a statistically significant but clinically irrelevant difference [26-28].
Reanalyzing one or several pooled clinical trials using a "favorable" subgroup of responders revealed by the individual trial(s) [29].
Inappropriate early stopping of clinical trials [30, 31]
Using small old studies in systematic reviews or meta-analyses which do not meet current quality standards or used heterogeneous definitions of the assessed endpoints [32].
Excluding patients from a randomized clinical trial when their characteristics are likely to yield negative results [33, 34].
Impossibility to ascertain that confounding variables have been accounted for [27, 28, 35].
Excessive or premature extrapolation of experimental data to the clinic [36].