Indispensable but deceptive evidence-based medicine

Christian Funck-Brentano

Indispensable but deceptive evidence-based medicine

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

[1] AP-HP Sorbonne Université, Pitié-Salpêtrière University Hospital, Department of Pharmacology and Clinical Investigation Centre, 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 Centre, Pitié-Salpêtrière Hospital, 47–83 Boulevard de l'Hôpital, 75013 Paris, France

Email: christian.funck-brentano@aphp.fr

Tel: +33 1 42 16 17 06

Competing interests: The author has no competing interests to declare.

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

Received 6 September 2020; Accepted 19 September 2020

Abstract

For millennia, medicine was empirical and dominated by beliefs, dogmas and sorcery. Evidence-based medicine then emerged over the last two centuries as a major source of progress towards enlightened clinical practice. Yet, the design of randomized clinical trials and observational studies and, even more so, their methods of data collection and analyses, and the way they are presented in the media, have become sources of important deviancies and inaccuracies. This may be explained by the extreme pressure exerted by both academic and commercial concerns on clinical research and researchers. Although there is no justification to contest the importance of evidence-based medicine, there is certainly a need to learn how to detect the ways in which medical information can be distorted to convey messages that do not reflect reality. This article discusses how and why evidence-based medicine can be deceitful, and provides examples to illustrate this point of view.

Key words (MeSH): Clinical study; Conflict of interest; Data accuracy; Endpoint determination; Epidemiology; Evidence-based medicine; Journalism; Medical; Methods; Reading; Research design

". . . I do not demand any final certainty from science . . ."

Karl Popper (Philosopher of Science)

Evidence in medicine is not as robust as it is in quantum physics and, after major breakthroughs due to the advent of evidence-based approaches developed over the past two centuries, it appears that some aspects of these approaches have gone awry. In medicine and biology, evidence is amenable to all sorts of manipulations and misinterpretations. Clinical practice requires knowing the difference between statistical evidence and clinically relevant evidence, but it still has to make decisions that rest on the best evidence available. The present article does not pretend to be an exhaustive analysis of all sources of concern, excess and overconfidence in evidence-based medicine (EBM). Indeed, critiques of EBM have already been made [1–4], and the design of studies and their statistical aspects have also been covered previously [5–8]. The intention here is to offer examples of data analyses and data presentation that may be suspected of embellishing or amplifying the results of clinical studies, whether they be randomized clinical trials (RCTs) or observational studies (OSs). Its final purpose is to emphasize the need for careful, critical reading of the scientific and medical literature in the media.

What is or may be the problem?

Sackett et al. [9] defined EBM as "the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research". This formulation emphasizes the fact, well known to practising physicians, that no result of a study performed in groups of

patients should systematically be applied to an individual patient. However, various commercial entities, scientists and the media persist in hammering home messages derived from major RCTs and OSs, with very good—but also not so good—reasons for doing so: for example, public-health protection, good clinical practices based on carefully prepared guidelines, financial profit and notoriety, among others. Many of these messages are exaggerated even by medical journals that have major impact [10], thereby leading some to consider that most published research findings are false [1], which might be a good starting point for readers, albeit certainly excessive.

The use of 'spin' in medical reporting—a way of presenting results in a 'positive' way without actually lying, but nevertheless failing to faithfully reflect the results—is a major feature of our modern times [11, 12], and strongly affects the positive interpretations of both patients and caregivers [13]. Press releases of RCTs [14] and OSs [15] are particularly prone to such deviance. As the number of studies published each year make them impossible to follow (Fig. 1), busy physicians are often tempted to rely on abstracts or media summaries, hence the difficulties they may face when digesting the literature and keeping track of medical progress [16]. Major industry-sponsored RCTs are often published in high-impact journals such as the *New England Journal of Medicine* and *The Lancet* as freely accessible articles, sometimes just minutes after they have been presented at international congresses. While there is nothing reprehensible *per se* about this practice, it should make physicians consider the commercial stakes associated with drug development, with medical publishers and how publications are used as promotional tools to foster drug prescriptions.

The horrifying concept of 'key opinion leaders' developed by the medical communication and marketing professionals says it all: industry is eager to 'improve' the communication of study results [17, 18]. Fortunately, the major medical journals that publish

influential RCTs and OSs generally provide readers with enough information to critique these studies not only through careful peer reviews that request appropriate analyses and presentation of data, but also by publishing editorials, commentaries and correspondence related to RCTs and OSs to allow their critical appraisal, provided that readers take the time to examine these additional sources of information.

Statistics largely determine our interpretation of RCTs and OSs, but may seem obscure to non-specialists. In fact, readers should be wary of multiple statistical analyses by subgroups and sophisticated statistical analytical techniques used for secondary analyses of clinical trials after publication of the main results. Yet, despite our being taught that such analyses should be avoided, our congresses and medical journals are full of these 'secondary' analyses, which can conceivably be profusely used for academic and/or commercial promotional purposes. They are also usually presented with a detailed study limitation section that is rarely recapitulated in the media coverage of these studies, including online continuing medical education (CME) programmes. In one analysis of 511 cardiovascular RCTs published by six high-impact medical journals between 1 August 2015 and 31 July 2018, it was revealed that, while multiple statistical tests were used in 58.7% of them, 71.7% of those failed to adjust their analyses to correct for multiple comparisons of the primary endpoint [19].

RCTs and OSs do not necessarily yield the same results when assessing the healthcare outcomes of treatments [20–22] although, on average, they do reach the same conclusions [23]. Indeed, various RCTs on the same topic can yield different results due to variable characteristics of their included populations and outcome assessments [24, 25], or simply because the intervention under study is too subtle to be associated with a sufficiently large effect, for example, the use of CYP2C19 genotyping to optimize clopidogrel prescription [26]. Also, systematic reviews [27] and meta-analyses [28] are subject to biases, and even those

published in the leading general medical journals have proved to be of limited generalizability [29], as have RCTs [30]. In addition, the financial ties of the principal investigators have been found to be independently associated with positive clinical trial results [31], as the source of funding does indeed influence RCT outcomes [32]. Finally, only a few recommendations from highly respected medical societies producing practice guidelines are based on high levels of evidence [33].

Thus, by and large, the above considerations tell us that the word 'evidence' in EBM should not be overstated, that reading the medical literature is a difficult exercise and that the best initial attitude to adopt when reading any article should be scepticism. Looking at the source article(s), including any supplementary file(s), and uncovering any biases in the study design or 'spin' in the presentation of results should be the rule.

Making results look better than they are

Results are only results—except in the case of outright falsification—and they rest exclusively on the quality of the methods used to obtain them. Therefore, in any original article, the most important section to read is on methodology, in particular (but not only) to examine: the inclusion/exclusion criteria to determine whether they appropriately represent the general target patient population; the appropriateness of the control population and its relevance as a comparator; the primary hypothesis being tested, and how clinically valuable and appropriately powered it may be; the study endpoints (primary and secondary), their clinical relevance and dependability, and how they were validated; the study amendments, and when and why they were made; the conduct of the study and drop-outs; any adverse events to assess the benefit/risk ratio; and, of course, the data analyses. If the methods are inappropriate, the results should no longer be considered. If the methods are sound, then it is

in the presentation of results, particularly at congresses and in media reports of the article, where exaggerations may be detected. Clinical significance should prevail over statistical significance, and it should be borne in mind that the *P* value characterizes the robustness of results—how certain it can be stated that they are not due to chance—but not the amplitude of differences that can only be determined from the limits of the confidence intervals comparing different groups.

Studies, RCTs and, even more so, OSs with large numbers of participants can insist on statistically significant results that are nevertheless of little or no interest in actual practice. For example, there is endless debate over the value of polyunsaturated fats (omega-3 fatty acids, fish oils) in the prevention of cardiovascular diseases. This was the focus of a recent OS showing small, but statistically significant, reductions in cardiovascular events and mortality among nearly half a million British participants consuming fish-oil supplements compared with non-consumers [34]. As expected in an OS, the two groups had very different baseline profiles, and various models were used to adjust the analyses for these differences. The most complete model adjusted 22 baseline variables: in this model, the upper limit of the 95% confidence interval (CI) of the hazard ratio (HR) between consumers and non-consumers was consistent with a 1–11% reduction in events (all-cause mortality and cardiovascular events, among others) in consumers, and mean reductions ranged from 7% to 13%. Several interactions were found. Yet, this study, albeit performed according to high statistical standards, nonetheless tells us little about the value of fish-oil supplementation because the patients most likely to benefit from such supplementation cannot be identified from this study and, more importantly, because the amplitude of any differences appears to be too small to justify a recommendation for systematic fish-oil supplementation.

On the other hand, an RCT comparing omega-3 polyunsaturated fatty acids (PUFAs) with a placebo in the treatment of heart failure [35] could find no statistically significant differences in either mortality or cardiovascular admissions between the two groups, even though what differences there were just managed to reach the threshold of significance after adjusting for baseline differences (Fig. 2). Despite being an unusual practice in an RCT, it had been prespecified in the protocol. In both these cases the authors recognized that the benefits of fish oil supplementation or omega-3 PUFA were of limited amplitude but the fact that absolute differences between treated and untreated patients was ridiculously small, 0.1 to 2 %, was not emphasized; thus, fish oils continue to be advertised to the general public as having an impact on health. However, the problem is not that such supplements lack a possible positive impact, but that the impact, if it is present, is so small that it cannot truly justify fish-oil or omega-3 fatty acid supplements for everyone. The true clinical problem is to identify who is more likely to benefit from such supplementation.

Post-hoc analyses are also used to strengthen and rejuvenate major RCTs [36, 37], sometimes even transforming a negative study into a positive one by selecting an appropriate subgroup of responders [37]. Angiotensin receptor-neprilysin inhibitors (ARNis) have proved beneficial in patients with heart failure and reduced left ventricular ejection fractions (LVEFs) [38], but not in those with heart failure and preserved LVEFs [39]. A recent post-hoc analysis pooled these trials together with other large-scale heart-failure trials and applied an enigmatic statistical method to convey the message that ARNis are beneficial across the full range of LVEFs up to 60% [40]. One wonders what the true motivation was for these reanalyses and how useful the results are to practising physicians. Recently, emphasis has been placed on the concept of clinically meaningful benefit in medicine [41, 42], including taking patients' expectations into consideration. It can only be hoped that progress will be made in future by

decreasing the weight of statistical evidence to favour the benefit of clinically relevant evidence.

Expressing results as relative, rather than absolute, changes can also amplify the perception of large differences. For example, in an OS involving 1,249,943 person-years of follow-up and 1476 cases of confirmed sudden death due to cardiac causes, this endpoint was doubled in patients treated with erythromycin compared with those not receiving an antibiotic, and a fivefold increase was found when erythromycin was prescribed with a CYP3A inhibitor [43]. This is a rather impressive increase in risk of death. However, the absolute number of patients on which these statistically significant increases in mortality were based was 10 and 3, respectively. In view of the weaknesses of medical diagnoses in OSs of this type [44], the validity of these results is uncertain, albeit highly plausible. The authors presented their results honestly in the article, but their absolute numbers were not presented in the abstract. Two different messages can be derived from these results: (*i*) there is an increased risk of sudden cardiac death associated with erythromycin, particularly when prescribed with inhibitors of its metabolism; and (*ii*) the risk is so small that the (unresolved) problem is to characterize what makes these few people at risk of sudden death. In fact, both affirmations were already known before the study was published because erythromycin, like other macrolide antibiotics, blocks IKr potassium channels and prolongs ventricular repolarization, triggering torsade de pointes on electrocardiography and potentially resulting in sudden death in very rare cases. Therefore, this OS adds little to our present knowledge that there is a risk, but it is very small in amplitude, whereas what we need to know is who is at risk and why the vast majority of patients taking erythromycin with or without a CYP3A inhibitor do not suddenly die.

Absolute risk reduction seemingly provides a better reflection of the clinical significance of results in that its reciprocal represents the number of patients needed to treat (or to harm in cases of increased risk of adverse events). While the concept is seductive, it also has its limitations: it is influenced by the duration of follow-up and sample size; it does not account for actuarial data (time-to-events); and it is not consistent across trials, thereby preventing comparisons to be made between drugs for similar indications [45–48].

Finally, causation in epidemiological studies can be difficult, if not impossible, to establish. The criteria proposed by Hill [49] in 1965 to strengthen the validity of causal associations in epidemiological studies (Table I) are often missing in publications of OSs. Although techniques for OS analyses have been developed to mimic RCTs [50], they are unlikely to solve the causation issue.

The importance of endpoint definition and validation

The essence of clinical research is to assess differences between outcomes, measured as study endpoints, across study groups. As most OSs and RCTs primarily analyze outcomes, particular attention should therefore be paid to the nature and quality of endpoints. In OSs, including analyses of so-called 'big data', outcomes are most often extracted from large databases or registries, which are prone to heterogeneity and misclassifications [44, 51]. When the size of an OS is very large, it is thought that such heterogeneity is acceptable and can be taken into consideration. This has yet to be proved. Thus, RCTs, where feasible, remain the most appropriate approach for studying an intervention [52]. However, many issues, such as assessing the relationship between exposure to a drug and rare adverse events (as mentioned above for erythromycin-induced sudden cardiac death as a proxy for torsades de pointes), can only be analyzed by epidemiological studies. In RCTs, clinical events (or the results of

diagnostic tests) are amenable to thorough prospective acquisition, quality control and classification based on clear, reproducible prespecified definitions. Nevertheless, in both OSs and RCTs, it is essential to question the nature of the endpoints used to assess outcomes. There are now recommendations regarding the classification of cardiovascular clinical events in RCTs [6, 53–55] that were only recently agreed upon, and their extreme complexity demonstrates how difficult they are to harmonize across RCTs.

Composite endpoints comprising major adverse cardiac events are most often used for primary endpoint assessment in cardiovascular RCTs, but it is important to examine the contribution of each component to the total effect of the composite endpoint [6]. Statistical issues regarding the analysis of multiple endpoints and whether only the first event must be considered have been discussed elsewhere [5, 56]; nevertheless, there has been insufficient debate over the nature and robustness of the endpoints used in RCTs.

Some clinical events chosen as endpoints are either not robust in their definition or their adjudication is difficult or impossible even when standard definitions are applied by adjudication committees. For example, most, if not all, recent cardiovascular RCTs have used ‘cardiovascular death’ as a component of the primary endpoint. However, cardiovascular death does not provide a clear diagnosis of the cause of death [57] and, more important, is a rather vulnerable endpoint as many deaths take place out of hospital, thereby often rendering adjudication of the cardiovascular nature of a death akin to gazing into a crystal ball. All-cause mortality is clearly much more relevant, as emphasized more than two decades ago in a commentary to a trial showing a reduction in arrhythmia deaths with no effect on all-cause mortality [58]. Likewise, the criteria for hospital admission or revascularization vary widely from country to country due to different practices, making harmonization of these endpoints

in international trials rather uncertain. Angina pectoris and even unstable angina are also extremely difficult endpoints to adjudicate.

One recent article discussed the uncertainties surrounding the definitions of clinical events in RCTs [59]. Nevertheless, adjudication committees are useful for classifying events in RCTs and have become standard in the conduct of large RCTs [55]. Although it has been suggested that estimates of treatment effect based on subjective clinical events assessed by onsite assessors *vs* adjudication committees do not differ significantly [60], other data challenge this view [61–64]. An unpublished personal example of the influence of an adjudication committee on an endpoint is demonstrated in Fig. 3. In that RCT, a statistically significant drug effect on sudden death was found according to the investigators' diagnoses. Yet, after centralized adjudication of all events in all patients participating to the trial was performed according to predetermined definitions, the statistical significance was lost even though the adjudicated differences *vs* investigators' diagnoses were small.

The lost population of 'responders'

Even when RCTs show a therapeutic benefit for indisputable clinical endpoints, it may yet be difficult to identify which patients are more likely to experience such benefit from a treatment. For example, in the BI 10773 (Empagliflozin) Cardiovascular Outcome Event Trial in Type 2 Diabetes Mellitus Patients - EMPA-REG OUTCOME, the sodium–glucose cotransporter type 2 (SGLT2) inhibitor empagliflozin achieved a surprisingly high 32% reduction in all-cause mortality, and 38% reduction in cardiovascular death, in the pooled empagliflozin 10-mg and 25-mg dose groups [65]. Total mortality is an indisputable endpoint. The primary endpoint, a composite of death due to cardiovascular causes, non-fatal myocardial infarction and/or non-fatal stroke, was decreased in the pooled empagliflozin

groups compared with a placebo, and the difference only just reached statistical significance (*P* = 0.04). The reduction in the primary endpoint was driven by cardiovascular mortality with a non-significant 13% reduction in non-fatal myocardial infarction, a non-significant 28% increase in silent myocardial infarction and a non-significant 24% increase in non-fatal stroke. There was also a 35% reduction in hospitalizations for heart failure, but only 3–4% of patients had such an event. The number of deaths from heart failure was not reported. Taken altogether, the results of the EMPA-REG OUTCOME tell us nothing about why total or cardiovascular death was reduced with empagliflozin, or which patient profile is more likely to be associated with therapeutic benefit.

What the trial findings do show, in a supplementary file, is that 3811 (33%) of 11,531 screened patients were excluded prior to randomization for failing to meet the inclusion/exclusion criteria—the same reason that 46% of patients were excluded from the recent Empagliflozin Outcome Trial in Patients with Chronic Heart Failure with Reduced Ejection Fraction (EMPEROR-Reduced) [66]—and that, in subgroup analyses, the primary endpoint was not reduced in patients aged < 65 years (representing > 50% of all included patients), in patients with glycated haemoglobin ≥ 8.5% and possibly also in those with a body mass index ≥ 30 kg/m². These later findings were obtained only in subgroup analyses and are therefore of limited clinical significance [6], but they nonetheless emphasize the difficulties for readers who seek to identify precisely what sort of patients are more likely to derive benefit if prescribed the tested drug.

Other RCTs with SGLT2 inhibitors, including the EMPEROR-Reduced [66], have failed to find any reduction in all-cause mortality, although they have shown clinical benefit for other endpoints [66, 67]. The main benefit with SGLT2 inhibitors appears to be linked to a reduction in heart failure-related events. However, only 10–14% of patients with diabetes included in

three major SGLT2 inhibitor trials [67] had a previous history of heart failure, and descriptions of heart failure-related events in those trials and whether such events were also reduced in patients with no previous history of heart failure were not reported in the main publication. In the end, it remains unknown as to which diabetes patients are more likely to justify a prescription for an SGLT2 inhibitor.

However, recent data indicate that this class of drugs might be beneficial to heart failure patients independently of diabetes [68]. The Dapagliflozin and Prevention of Adverse Outcomes in Heart Failure (DAPA-HF) trial, which showed a significant reduction in all-cause death, also found a 26% reduction in the primary composite endpoint of worsening heart failure or death due to cardiovascular causes [69]. A recent reanalysis of the DAPA-HF data according to various levels of decreased LVEF revealed that dapagliflozin significantly reduced major cardiovascular endpoints and total mortality at all levels of LVEF [70]. However, the authors acknowledged in their study limitations that LVEF was not measured consistently in all patients. This leaves readers wondering exactly what this secondary analysis adds to the primary publication and whether these data will simply be used for promotional purposes rather than enhancing patients' care.

Finally, even if there are differences among the newer antidiabetic drugs [such as SGLT2 inhibitors and glucagon-like peptide (GLP)-1 receptor agonists] in their ability to reduce different types of cardiovascular events in diabetes patients, these differences appear unlikely to have any major global impact on patients' prognoses and major adverse cardiovascular events to justify favouring the prescription of one drug class over another [71, 72] (Fig. 4). In fact, no RCT comparing the two above-mentioned drug classes head-to-head has ever been performed. However, there is no doubt that these new antidiabetic agents represent true progress in the treatment of diabetes patients. Nevertheless, it is still unclear how to select

those patients who are more likely to benefit from these treatments. Identifying responders to therapies—in other words, precision medicine that targets patients with a high probability of benefit from a given treatment while avoiding exposing others to its adverse effects—is the major challenge of the future, especially in view of the increasing costs of the newer drug therapies.

Questioning the future

The scepticism expressed in this article should not be interpreted as a disavowal of EBM, which has, in fact, rid medicine of centuries of obscurantism and witchcraft. Notwithstanding the major progress brought about by the formalization of RCTs and OSs over the last few decades, there still remains the risk of the improper use of large datasets that computer power can generate. RCTs are still the most reliable source of information [52]. Indeed, exploiting electronic health records and other large databases to test interventions, decide on health policies and approve new drugs will continue to be a challenge [51], and is now a field of active research [73–75].

Conclusion

In *Le Dictionnaire des idées reçues* (*Dictionary of Accepted Ideas*) published after his death in 1911, Gustave Flaubert defined an optimist as a "synonym for imbecile". This satirical definition reminds us that positivism (or negativism) is inappropriate in science and that neutrality should govern our judgment, as it is to be expected that most evidence-based clinical research results will be presented to emphasize the authors' particular point of view, which might be influenced by non-scientific considerations. It is therefore good practice to

initially adopt an opposing viewpoint in the interpretation of published data before forming a reasonable opinion based on the research.

Ronald H. Coase, the British economist who won the 1991 Nobel Prize in Economic Sciences, militated against complex mathematical models in the study of economics. He once wrote: “If you torture the data long enough, nature will always confess” [76]. Data torture in the analysis of OSs and RCTs has recently prospered under the pressure of commercial and academic (conflicts of) interests and now threatens the ideal of EBM. Yet, there is no alternative to evidence-based decisions in medicine. Thus, because it is unlikely that data torture, biased studies and presentations of ‘spun’ results will suddenly cease to exist, all physicians should be wary of what they read, no matter the reputation of the publishing journal, and develop a sense of critical thinking to truly appraise the nature and weight of the evidence presented.

References

[1] Ioannidis JP. Why most published research findings are false. PLoS Med 2005;2:e124. https://doi.org/10.1371/journal.pmed.0020124

[2] Greenhalgh T, Howick J, Maskrey N, Evidence Based Medicine Renaissance Group. Evidence based medicine: a movement in crisis? BMJ 2014;348:g3725. https://doi.org/10.1136/bmj.g3725

[3] Heneghan C, Goldacre B, Mahtani KR. Why clinical trial outcomes fail to translate into benefits for patients. Trials 2017;18:122. https://doi.org/10.1186/s13063-017-1870-2

[4] Ioannidis JPA. Randomized controlled trials: Often flawed, mostly useless, clearly indispensable: A commentary on Deaton and Cartwright. Soc Sci Med 2018;210:53-6. https://doi.org/10.1016/j.socscimed.2018.04.029

[5] 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. https://doi.org/10.1016/j.jacc.2015.10.014

[6] 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. https://doi.org/10.1016/j.jacc.2015.10.023

[7] 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. https://doi.org/10.1016/j.jacc.2015.10.036

[8] 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. https://doi.org/10.1016/j.jacc.2015.10.051

[9] Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996;312:71-2. https://doi.org/10.1136/bmj.312.7023.71

[10] Gyles C. Skeptical of medical science reports? Can Vet J 2015;56:1011-2.

[11] Boutron I. Spin in Scientific Publications: A Frequent Detrimental Research Practice. Ann Emerg Med 2020;75:432-4. https://doi.org/10.1016/j.annemergmed.2019.11.002

[12] McGrath TA, Bowdridge JC, Prager R, Frank RA, Treanor L, Dehmoobad Sharifabadi A, et al. Overinterpretation of Research Findings: Evaluation of "Spin" in Systematic Reviews of Diagnostic Accuracy Studies in High-Impact Factor Journals. Clin Chem 2020;66:915-24. https://doi.org/10.1093/clinchem/hvaa093

[13] Boutron I, Haneef R, Yavchitz A, Baron G, Novack J, Oransky I, et al. Three randomized controlled trials evaluating the impact of "spin" in health news stories reporting studies of pharmacologic treatments on patients'/caregivers' interpretation of treatment benefit. BMC Med 2019;17:105. https://doi.org/10.1186/s12916-019-1330-9

[14] Yavchitz A, Boutron I, Bafeta A, Marroun I, Charles P, Mantz J, et al. Misrepresentation of randomized controlled trials in press releases and news coverage: a cohort study. PLoS Med 2012;9:e1001308. https://doi.org/10.1371/journal.pmed.1001308

[15] Oliver J. Scientific Studies: Last Week Tonight with John Oliver (HBO), https://www.youtube.com/watch?v=0Rnq1NpHdmw; 2016 [accessed 18 September 2020].

[16] Shubrook JH, Pak J, Dailey G. Primary Care Physicians' Knowledge of the Cardiovascular Effects of Diabetes Medications: Findings from an Online Survey. Adv Ther 2020;37:3630-9. https://doi.org/10.1007/s12325-020-01405-y

[17] Moynihan R, Albarqouni L, Nangla C, Dunn AG, Lexchin J, Bero L. Financial ties between leaders of influential US professional medical associations and industry: cross sectional study. BMJ 2020;369:m1505. https://doi.org/10.1136/bmj.m1505

[18] Thomas K. Key opinion leaders supercharged by the internet: paid doctor and patient influencers on social media. BMJ 2019;365:l2336. https://doi.org/10.1136/bmj.l2336

[19] 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. https://doi.org/10.1001/jamanetworkopen.2020.3082

[20] Ewald H, Ioannidis JPA, Ladanie A, Mc Cord K, Bucher HC, Hemkens LG. Nonrandomized studies using causal-modeling may give different answers than RCTs: a meta-epidemiological study. J Clin Epidemiol 2020;118:29-41. https://doi.org/10.1016/j.jclinepi.2019.10.012

[21] Califf RM, Hernandez AF, Landray M. Weighing the Benefits and Risks of Proliferating Observational Treatment Assessments: Observational Cacophony, Randomized Harmony. JAMA 2020. https://doi.org/10.1001/jama.2020.13319

[22] Aguirre Davila L, Weber K, Bavendiek U, Bauersachs J, Wittes J, Yusuf S, et al. Digoxin-mortality: randomized vs. observational comparison in the DIG trial. Eur Heart J 2019;40:3336-41. https://doi.org/10.1093/eurheartj/ehz395

[23] Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. Cochrane Database Syst Rev 2014:MR000034. https://doi.org/10.1002/14651858.MR000034.pub2

[24] Gaudino M, Ruel M, Obadia JF, De Bonis M, Puskas J, Biondi-Zoccai G, et al. Methodologic considerations on four cardiovascular interventions trials with contradictory results. Ann Thorac Surg 2020. https://doi.org/10.1016/j.athoracsur.2020.04.107

[25] McCreary EK, Angus DC. Efficacy of Remdesivir in COVID-19. JAMA 2020;324:1041-2. https://doi.org/10.1001/jama.2020.16337

[26] Moliterno DJ, Smyth SS, Abdel-Latif A. CYP2C19 Genotyping to Guide Antiplatelet Therapy After Percutaneous Coronary Interventions: One Size Rarely Fits All. JAMA 2020;324:747-9. https://doi.org/10.1001/jama.2020.13094

[27] Page MJ, McKenzie JE, Kirkham J, Dwan K, Kramer S, Green S, et al. Bias due to selective inclusion and reporting of outcomes and analyses in systematic reviews of randomised trials of healthcare interventions. Cochrane Database Syst Rev 2014:MR000035. https://doi.org/10.1002/14651858.MR000035.pub2

[28] Kicinski M, Springate DA, Kontopantelis E. Publication bias in meta-analyses from the Cochrane Database of Systematic Reviews. Stat Med 2015;34:2781-93. https://doi.org/10.1002/sim.6525

[29] Malmivaara A. Generalizability of findings from systematic reviews and meta-analyses in the Leading General Medical Journals. J Rehabil Med 2020;52:jrm00031. https://doi.org/10.2340/16501977-2659

[30] Malmivaara A. Generalizability of findings from randomized controlled trials is limited in the leading general medical journals. J Clin Epidemiol 2019;107:36-41. https://doi.org/10.1016/j.jclinepi.2018.11.014

[31] 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. https://doi.org/10.1136/bmj.i6770

[32] 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. https://doi.org/10.1016/j.ijcard.2016.12.119

[33] Fanaroff AC, Califf RM, Windecker S, Smith SC, Jr., Lopes RD. Levels of Evidence Supporting American College of Cardiology/American Heart Association and European Society of Cardiology Guidelines, 2008-2018. JAMA 2019;321:1069-80. https://doi.org/10.1001/jama.2019.1122

[34] 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. https://doi.org/10.1136/bmj.m456

[35] 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. https://doi.org/10.1016/S0140-6736(08)61239-8

[36] Fitchett D, Inzucchi SE, Wanner C, Mattheus M, George JT, Vedin O, et al. Relationship between hypoglycaemia, cardiovascular outcomes, and empagliflozin treatment in the EMPA-REG OUTCOME(R) trial. Eur Heart J 2020;41:209-17. https://doi.org/10.1093/eurheartj/ehz621

[37] 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. https://doi.org/10.1093/eurheartj/ehz427

[38] McMurray JJ, Packer M, Desai AS, Gong J, Lefkowitz MP, Rizkala AR, et al. Angiotensin-neprilysin inhibition versus enalapril in heart failure. N Engl J Med 2014;371:993-1004. https://doi.org/10.1056/NEJMoa1409077

[39] Solomon SD, McMurray JJV, Anand IS, Ge J, Lam CSP, Maggioni AP, et al. Angiotensin-Neprilysin Inhibition in Heart Failure with Preserved Ejection Fraction. N Engl J Med 2019;381:1609-20. https://doi.org/10.1056/NEJMoa1908655

[40] 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. https://doi.org/10.1093/eurheartj/ehaa184

[41] Weinfurt KP. Clarifying the Meaning of Clinically Meaningful Benefit in Clinical Research: Noticeable Change vs Valuable Change. JAMA 2019. https://doi.org/10.1001/jama.2019.18496

[42] Leopold SS, Porcher R. Editorial: The Minimum Clinically Important Difference-The Least We Can Do. Clin Orthop Relat Res 2017;475:929-32. https://doi.org/10.1007/s11999-017-5253-5

[43] 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. https://doi.org/10.1056/NEJMoa040582

[44] Desai RJ, Levin R, Lin KJ, Patorno E. Bias Implications of Outcome Misclassification in Observational Studies Evaluating Association Between Treatments and All-Cause or Cardiovascular Mortality Using Administrative Claims. J Am Heart Assoc 2020:e016906. https://doi.org/10.1161/JAHA.120.016906

[45] Monnier L, Colette C, Bonnet F, Owens D. Number Needed-to-Treat (NNT): Is it a necessary marker of therapeutic efficiency? Diabetes Metab 2020;46:261-4. https://doi.org/10.1016/j.diabet.2020.01.004

[46] Ludwig L, Darmon P, Guerci B. Computing and interpreting the Number Needed to Treat for Cardiovascular Outcomes Trials : Perspective on GLP-1 RA and SGLT-2i therapies. Cardiovasc Diabetol 2020;19:65. https://doi.org/10.1186/s12933-020-01034-3

[47] Wald NJ, Morris JK. Two under-recognized limitations of number needed to treat. Int J Epidemiol 2020;49:359-60. https://doi.org/10.1093/ije/dyz267

[48] Mendes D, Alves C, Batel-Marques F. Number needed to treat (NNT) in clinical literature: an appraisal. BMC Med 2017;15:112. https://doi.org/10.1186/s12916-017-0875-8

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

[50] Thomas L, Li F, Pencina M. Using Propensity Score Methods to Create Target Populations in Observational Clinical Research. JAMA 2020;323:466-7. https://doi.org/10.1001/jama.2019.21558

[51] Cave A, Brun NC, Sweeney F, Rasi G, Senderovitz T, Taskforce H-EJBD. Big Data - How to Realize the Promise. Clin Pharmacol Ther 2020;107:753-61. https://doi.org/10.1002/cpt.1736

[52] 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. https://doi.org/10.1056/NEJMsb1901642

[53] Hicks KA, Tcheng JE, Bozkurt B, Chaitman BR, Cutlip DE, Farb A, et al. 2014 ACC/AHA Key Data Elements and Definitions for Cardiovascular Endpoint Events in Clinical Trials: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Data Standards (Writing Committee to Develop Cardiovascular Endpoints Data Standards). J Am Coll Cardiol 2015;66:403-69. https://doi.org/10.1016/j.jacc.2014.12.018

[54] Hicks KA, Mahaffey KW, Mehran R, Nissen SE, Wiviott SD, Dunn B, et al. 2017 Cardiovascular and Stroke Endpoint Definitions for Clinical Trials. J Am Coll Cardiol 2018;71:1021-34. https://doi.org/10.1016/j.jacc.2017.12.048

[55] Cefalu WT, Kaul S, Gerstein HC, Holman RR, Zinman B, Skyler JS, et al. Cardiovascular Outcomes Trials in Type 2 Diabetes: Where Do We Go From Here? Reflections From a Diabetes Care Editors' Expert Forum. Diabetes Care 2018;41:14-31. https://doi.org/10.2337/dci17-0057

[56] Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. Eur Heart J 2012;33:176-82. https://doi.org/10.1093/eurheartj/ehr352

[57] Ghosal S, Sinha B. Redefining Cardiovascular (CV) Death as a primary end point component in Cardiovascular Outcome Trials. Curr Diabetes Rev 2020. https://doi.org/10.2174/1573399816666200206113549

[58] Gottlieb SS. Dead is dead--artificial definitions are no substitute. Lancet 1997;349:662-3. https://doi.org/10.1016/S0140-6736(97)22010-6

[59] Drexel H, Rosano GMC, Lewis BS, Huber K, Vonbank A, Dopheide JF, et al. The age of randomized clinical trials: three important aspects of randomized clinical trials in cardiovascular pharmacotherapy with examples from lipid and diabetes trials. Eur Heart J Cardiovasc Pharmacother 2020;6:97-103. https://doi.org/10.1093/ehjcvp/pvz029

[60] Ndounga Diakou LA, Trinquart L, Hrobjartsson A, Barnes C, Yavchitz A, Ravaud P, et al. Comparison of central adjudication of outcomes and onsite outcome assessment on treatment effect

estimates. Cochrane Database Syst Rev 2016;3:MR000043. https://doi.org/10.1002/14651858.MR000043.pub2

[61] Heagerty A, Deverly A, Palmer C, Kaplinsky E, Salvetti A, Wahlgren NG, et al. The role of the critical event committee in a major cardiovascular outcome study. Blood Press 2002;11:339-44. https://doi.org/10.1080/080370502321095294

[62] Mahaffey KW, Roe MT, Dyke CK, Newby LK, Kleiman NS, Connolly P, et al. Misreporting of myocardial infarction end points: results of adjudication by a central clinical events committee in the PARAGON-B trial. Second Platelet IIb/IIIa Antagonist for the Reduction of Acute Coronary Syndrome Events in a Global Organization Network Trial. Am Heart J 2002;143:242-8. https://doi.org/10.1067/mhj.2002.120145

[63] Quinlan DJ, Eikelboom JW, Goodman SG, Welsh RC, Fitchett DH, Theroux P, et al. Implications of variability in definition and reporting of major bleeding in randomized trials of oral P2Y12 inhibitors for acute coronary syndromes. Eur Heart J 2011;32:2256-65. https://doi.org/10.1093/eurheartj/ehr143

[64] Seltzer JH, Turner JR, Geiger MJ, Rosano G, Mahaffey KW, White WB, et al. Centralized adjudication of cardiovascular end points in cardiovascular and noncardiovascular pharmacologic trials: a report from the Cardiac Safety Research Consortium. Am Heart J 2015;169:197-204. https://doi.org/10.1016/j.ahj.2014.11.003

[65] Zinman B, Wanner C, Lachin JM, Fitchett D, Bluhmki E, Hantel S, et al. Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. N Engl J Med 2015;373:2117-28. https://doi.org/10.1056/NEJMoa1504720

[66] Packer M, Anker SD, Butler J, Filippatos G, Pocock SJ, Carson P, et al. Cardiovascular and Renal Outcomes with Empagliflozin in Heart Failure. N Engl J Med 2020. https://doi.org/10.1056/NEJMoa2022190

[67] Cosentino F, Grant PJ, Aboyans V, Bailey CJ, Ceriello A, Delgado V, et al. 2019 ESC Guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD. Eur Heart J 2020;41:255-323. https://doi.org/10.1093/eurheartj/ehz486

[68] Jarcho JA. More Evidence for SGLT2 Inhibitors in Heart Failure. N Engl J Med 2020. https://doi.org/10.1056/NEJMe2027915

[69] McMurray JJV, Solomon SD, Inzucchi SE, Kober L, Kosiborod MN, Martinez FA, et al. Dapagliflozin in Patients with Heart Failure and Reduced Ejection Fraction. N Engl J Med 2019;381:1995-2008. https://doi.org/10.1056/NEJMoa1911303

[70] Dewan P, Solomon SD, Jhund PS, Inzucchi SE, Kober L, Kosiborod MN, et al. Efficacy and safety of sodium-glucose co-transporter 2 inhibition according to left ventricular ejection fraction in DAPA-HF. Eur J Heart Fail 2020. https://doi.org/10.1002/ejhf.1867

[71] Pineda ED, Liao IC, Godley PJ, Michel JB, Rascati KL. Cardiovascular Outcomes Among Patients with Type 2 Diabetes Newly Initiated on Sodium-Glucose Cotransporter-2 Inhibitors, Glucagon-Like Peptide-1 Receptor Agonists, and Other Antidiabetic Medications. J Manag Care Spec Pharm 2020;26:610-8. https://doi.org/10.18553/jmcp.2020.26.5.610

[72] Fei Y, Tsoi MF, Cheung BMY. Cardiovascular outcomes in trials of new antidiabetic drug classes: a network meta-analysis. Cardiovasc Diabetol 2019;18:112. https://doi.org/10.1186/s12933-019-0916-z

[73] Eichler HG, Koenig F, Arlett P, Enzmann H, Humphreys A, Petavy F, et al. Are Novel, Nonrandomized Analytic Methods Fit for Decision Making? The Need for Prospective, Controlled, and Transparent Validation. Clin Pharmacol Ther 2020;107:773-9. https://doi.org/10.1002/cpt.1638

[74] Schmidli H, Haring DA, Thomas M, Cassidy A, Weber S, Bretz F. Beyond Randomized Clinical Trials: Use of External Controls. Clin Pharmacol Ther 2020;107:806-16. https://doi.org/10.1002/cpt.1723

[75] Franklin JM, Pawar A, Martin D, Glynn RJ, Levenson M, Temple R, et al. Nonrandomized Real-World Evidence to Support Regulatory Decision Making: Process for a Randomized Trial Replication Project. Clin Pharmacol Ther 2020;107:817-26. https://doi.org/10.1002/cpt.1633

[76] Coase RH. Essays on economics and economists. Chicago: The University of Chicago Press; 1994.

Figure legends

Fig. 1. Graph showing the number of randomized clinical trials and observational studies published in the decade since 2009.

Fig. 2. Randomized clinical trial comparing omega-3 polyunsaturated fatty acids (n-3 PUFA) and placebo in heart failure [35]: there were no statistically significant differences in either mortality or cardiovascular admissions, although differences did reach the significance threshold after adjusting for baseline differences. *Estimates were calculated with a Cox proportional hazards model, with adjustment for admission to hospital for heart failure in the previous year, previous pacemaker, and aortic stenosis. *Reprinted and adapted from The Lancet, 2008, ref [35] with permission from Elsevier.*

Fig. 3. In this randomized clinical trial, the statistically significant drug effect on sudden death according to the investigators' diagnoses (left panel) was lost after centralized adjudication of all events in all patients (right panel).

Fig. 4. Comparisons and odds ratios (ORs) of the newer antidiabetic drug classes *vs* placebo regarding their ability to reduce different types of cardiovascular events in diabetes patients. GLP-1, glucagon-like peptide 1; SGLT2, sodium–glucose cotransporter type 2; DPP-4, dipeptidyl peptidase 4. **a.** MACE (major adverse cardiovascular events). **b.** Nonfatal myocardial infarction. **c.** Nonfatal stroke. **d.** Cardiovascular mortality. **e.** All-cause mortality. **f.** Hospitalisation for heart failure. **g.** Renal composite outcome. *Reprinted from Cardiovascular Diabetology, ref [72] under the Creative Commons CC BY license (https://creativecommons.org/licenses/).*

Table 1: Sir Austin Bradford Hill's criteria for causation in epidemiological studies.

Hill's Criteria	Meaning
1. Strength of the association	Associations of large size are more likely to be causal than those of small size.
2. Consistency	Has the association been repeatedly observed by different persons, in different places, circumstances and times?
3. Specificity	There is no apparent other explanation for the association.
4. Temporality	The effect occurs after the cause (taking into account a plausible delay in some cases).
5. Biological gradient	Greater exposure is associated with greater risk (or reduction of risk), i.e. there is a "dose"-effect relationship.
6. Plausibility	A mechanism relating effect and cause can be found (often not possible, still supporting).
7. Coherence	The association is coherent with several factors, including biological, and doe not seriously conflict with the generally known facts of the natural history and biology of the disease.
8. Experiment	If possible, some form of interventional study can be undertaken to support the association. For example, suppression of the cause suppresses the effect.
9. Analogy	The association is more likely to be causal if it has already been observed in similar settings.

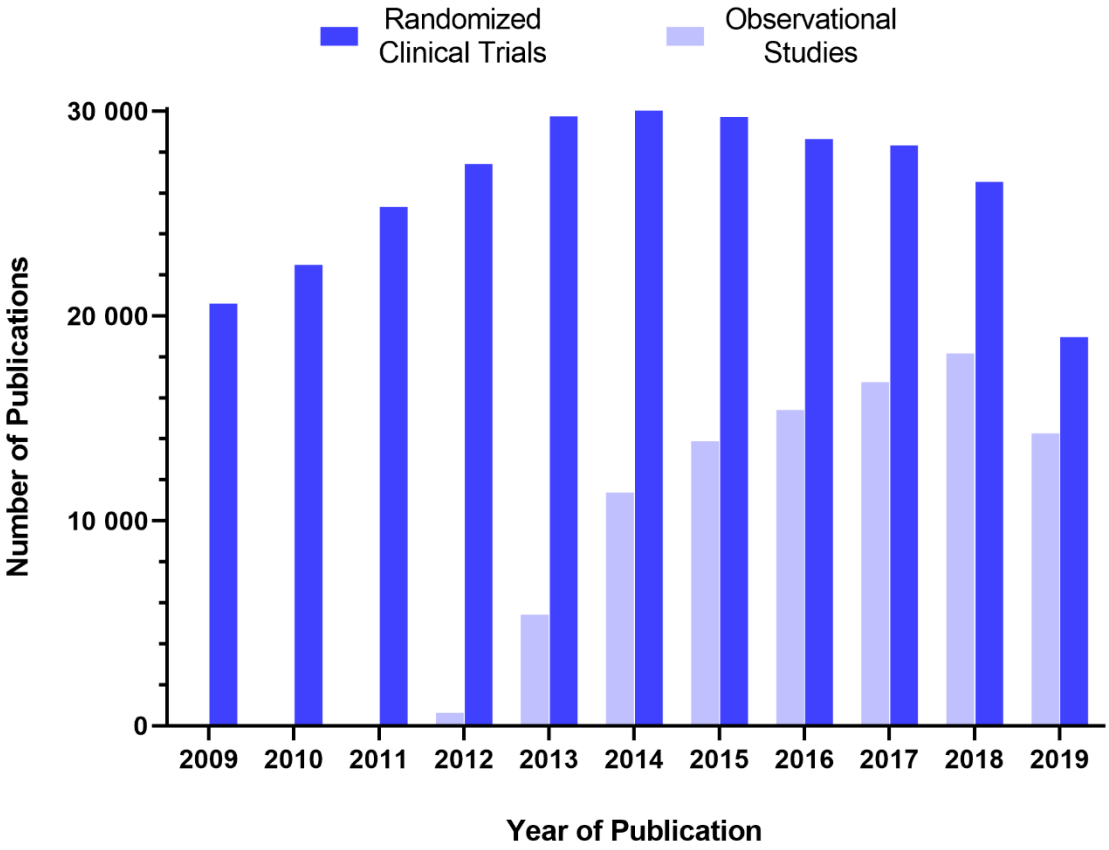

Figure 1: Evolution of the number of publications indexed as Randomized Clinical Trials or Observational Studies in PubMed from 2009 to 2019. The proportion of observational studies (term included in PubMed in 2014) is increasing.

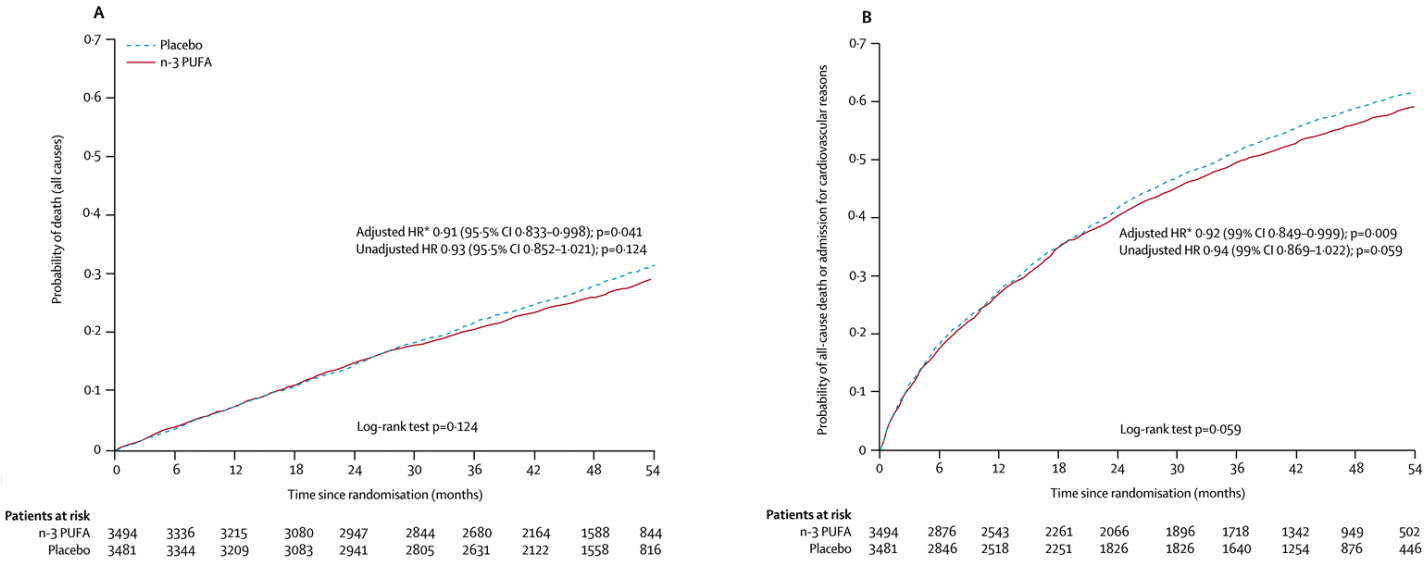

Figure 2: Kaplan-Meier curves for time to all-cause death (A) and for time to all-cause death or admission to hospital for cardiovascular reasons (B)

PUFA=polyunsaturated fatty acids. *Estimates were calculated with a Cox proportional hazards model, with adjustment for admission to hospital for heart failure in the previous year, previous pacemaker, and aortic stenosis. *Reprinted and adapted from The Lancet, 2008, ref [34] with permission from Elsevier.*

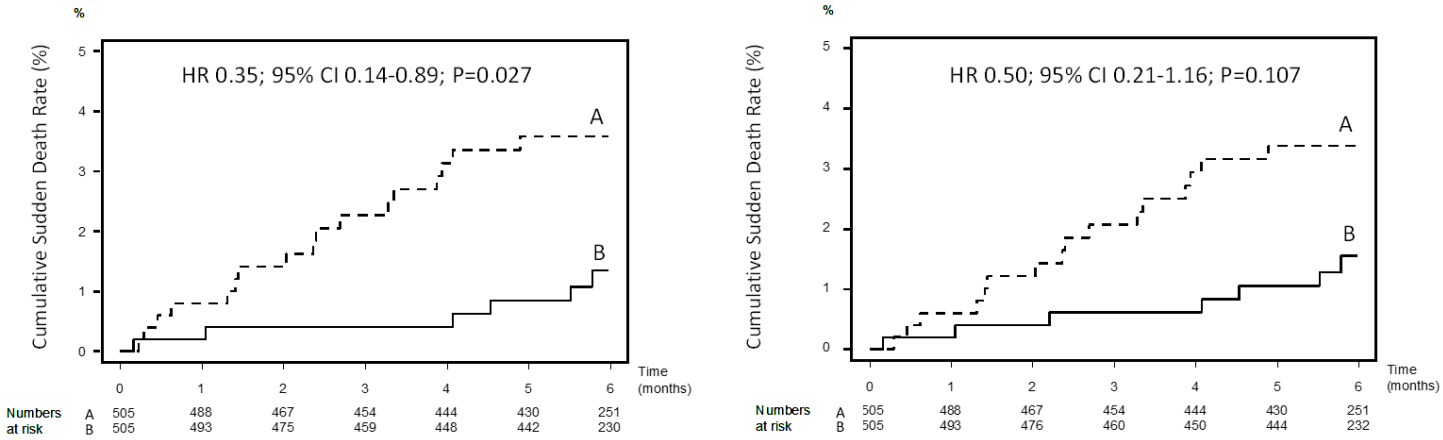

Figure 3: Influence of an adjudication committee on an outcome assessment. Sudden death was first analyzed in a randomized clinical trial using investigators declarations and showed a statistically significant reduction with treatment B (left panel). The statistical significance (but not the clinical significance) disappeared after the adjudication committee blindly adjudicated each case based on predefined criteria and supportive documents (right panel). CI: confidence interval; HR: hazard ratio. *Personal unpublished data.*

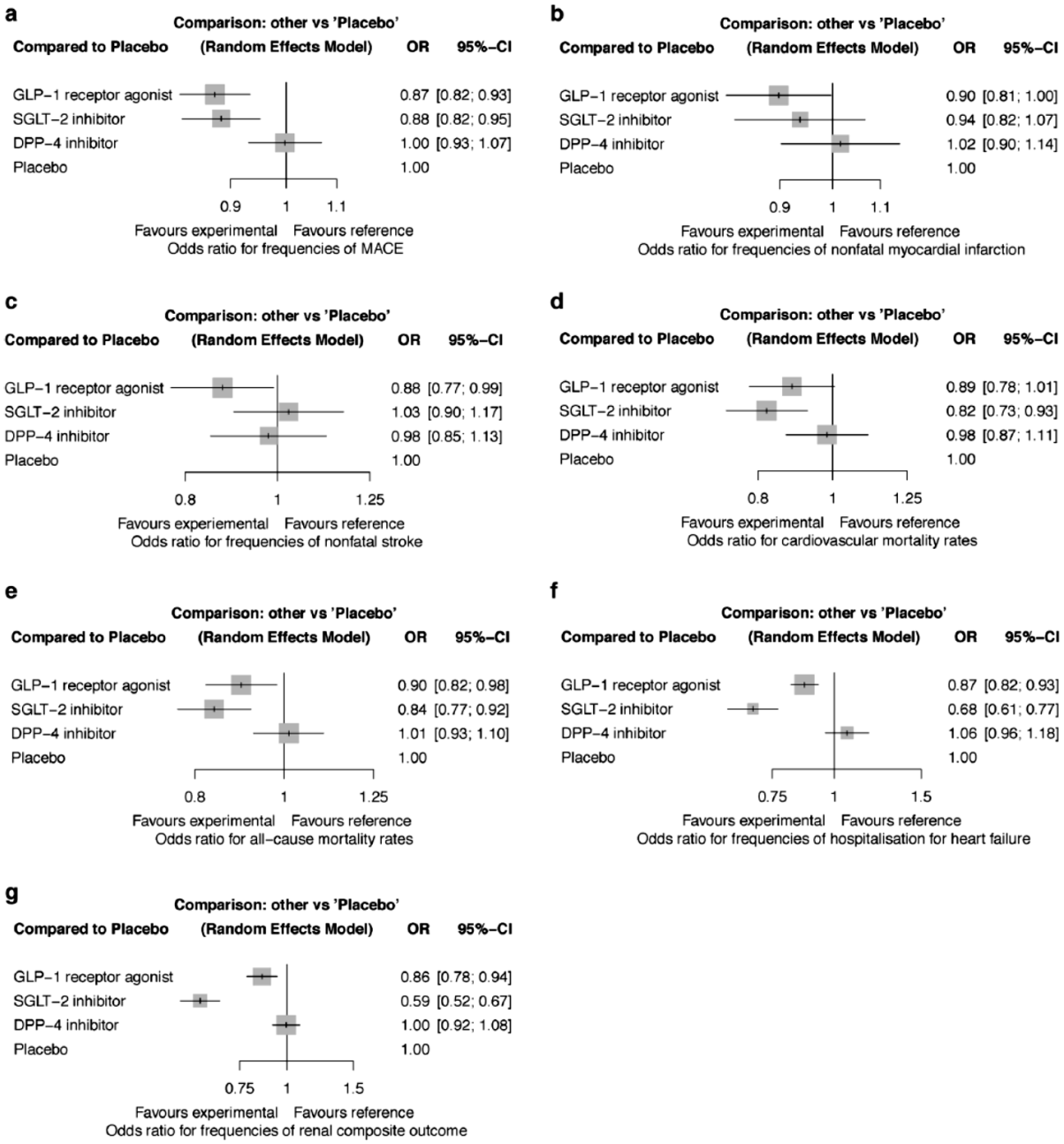

Figure 4: Risk of outcomes with different antidiabetic drug classes compared to placebo.

a. MACE (major adverse cardiovascular events). **b.** Nonfatal myocardial infarction. **c.** Nonfatal stroke. **d.** Cardiovascular mortality. **e.** All-cause mortality. **f.** Hospitalisation for heart failure. **g.** Renal composite outcome. *Reprinted from Cardiovascular Diabetology, ref [66] under the Creative Commons CC BY license (https://creativecommons.org/licenses/).*