Bias in Recent Papers on Diets and Drugs in Peer-Reviewed Medical Journals

Joel M. Kauffman, Ph.D.

ABSTRACT

Recent efforts by medical journal staffs to improve the quality of research papers have had mixed results.

Examples are given to show that randomized, placebo-controlled trials are not free from bias and that the failure to include all-cause death rates can be extremely misleading, as can the use of relative risks in the absence of absolute risks.

Other examples show how the conclusions in an abstract may not agree with the data in the body of the paper, or do not tell the whole truth. Still others use false surrogate endpoints or faulty trial protocols to favor a desired outcome. The whole picture may be seen as a breakdown of the peer-review system.

Introduction

The randomized, placebo-controlled clinical trial (RCT) is considered the pre-eminent form of scientific medical research. Recently, the Journal of the American Medical Association (JAMA) published evidence that RCTs were more likely to favor the intervention if the trial was funded by for-profit organizations.

Earlier, JAMA had published a meta-analysis of 37 trials revealing that industry-sponsored studies were significantly more likely to reach conclusions favorable to the sponsor than were non-industry-sponsored studies. The differences in outcomes observed between study centers in a number of multi-center drug RCTs are so extreme as to question their lack of bias.

In the Beta-Blocker Heart Attack Trial (BHAT) of propranolol vs. placebo on survivors of myocardial infarction, done in 32 centers in the US and Canada—which was terminated early because the all-cause mortality supposedly was reduced from 9.8 to 7.2%—examination of data from the individual centers showed increased mortality for propranolol in nine centers, and no decrease beyond the 95% confidence interval in any of the other centers.

A significant fraction of medical papers still report outcomes of epidemiological studies or trials in terms of the relative risk (RR) of a certain condition, such as cardiovascular deaths, without giving the all-cause death rates or the absolute risk. As favorable as the selective endpoint may appear to be for treatment, it is not possible to make any personal or policy decisions unless the change in the all-cause death rate with treatment is known. Since all-cause death rates are always available and not subject to medical examiner bias, there is no ethical reason not to include them. Furthermore, any reduction in the RR of all-cause death, however large, may be insignificant when the absolute risk is considered.

An example of failure to give absolute risk, cited in at least two books, was one of the reports of the West of Scotland Coronary Prevention Study Group (WOSCOPS) RCT on 6,595 men aged 45 to 64 years with initial mean cholesterol level of 272 mg/dL (7.0 mmol/L) assigned 40 mg pravastatin daily or placebo, and followed for a mean of 4.9 years. Results in the abstract included the statement, “We observed a 22% reduction in the risk of death from any cause in the pravastatin group (95% CI = 0-40%, P = 0.051).” The actual percentage of the men alive after 5 to 6 years in the placebo group was 95.9%, and in the pravastatin group 96.8%, an absolute difference of just 0.9%. This simple representation of the outcome was not seen in either the abstract or the discussion.

While it is known that the RR = 1.00 for all-cause mortality in the best trials of mammographic screening for breast cancer, physicians have told most women that their annual screening would cut their risk of dying from breast cancer by 17%, the mean of four large trials, RR = 0.83. How many women would bother with screening if told that their absolute risk was cut by 0.009%, the mean of four large trials, or even 0.09% after 10 years of screening, according to Gigerenzer? While most of the lay public expect advertisers to exaggerate the benefits of their products, physicians still may not be aware of the need for types of information that are frequently missing in medical papers, such as absolute risks, and number needed to treat (NNT) to prevent one adverse outcome.

Many interventions are justified on conveniently measured (surrogate) outcomes, such as bone density, cholesterol level, ECGs, and blood pressure. In all of these cases, examples exist in which the intervention improved the surrogate outcome, yet worsened the primary outcome of bone fracture or death.

A “primary endpoint” is supposed to indicate an important and undeniable change, such as in death rate, fracture rate, or elimination of a pathogen from the body. Many papers define primary endpoints to suit the desired outcome, such as changes in surrogate endpoints, or by mixing fatal with nonfatal events.

Since Medline and other searches often yield only abstracts, it is vital for authors to include all the important findings in the abstract. When press releases based on imminent publications in medical journals are distributed to reporters, the press release is likely to include only the results in the abstract of the paper.

The use of unnamed ghostwriters and figurehead authors in papers on drug research has been well documented, along with directions from sponsors to authors about what key phrases to include, and what findings to deemphasize. At least two recent cases of biased selection of references in support of a predetermined position have been exposed.

The Journal of the American Medical Association, the British Medical Journal, Lancet, and the New England Journal of Medicine are to be commended for publishing devastating exposes of conflicts of interest, or tips on spotting misleading presentation of data. Strict disclosure of funding sources is now required. Have medical papers thus improved in the last few years?

Some examples follow of most of these types of bias, easily culled from articles in recent medical journals normally considered to have the best reputations. Rather than merely asserting that misrepresentation exists, some specific detail will be given for each of nine articles.
Lack of All-Cause or Absolute Death Rates in Abstract

The Chicago Western Electric Study followed the effects of fish consumption in 2,107 men aged 40 to 55 for 30 years. Those who ate ≥35 g daily had, for fatal coronary heart disease (CHD) or fatal myocardial infarction (MI), a RR = 0.62, compared with men who consumed none. From the abstract’s conclusions: “These data show an inverse association between fish consumption and death from CHD, especially non-sudden death from MI.” In Table 2, the age-adjusted RRs for all-cause death were: 0 g/day, 1.00; 1-17 g/day, 1.00; 18-34 g/day, 0.98; 35 g/day, 0.90. So the all-cause RR was 0.90 from most to no fish consumption, and was not significant.

The Nurses’ Health Study on 84,688 women aged 34 to 59 years, followed for 16 years for outcomes vs. fish and omega-3 fatty acid intake, had the following conclusions in the abstract: “Among women, higher consumption of fish and omega-3 fatty acids is associated with a lower risk of CHD, particularly CHD deaths.”

In the body of the paper, in text only, for all-cause mortality, the RR was 0.68 for women consuming fish five times weekly vs. once/month; RR was 0.75 for the extreme quintiles of total omega-3 intake; both were significant.

Because of concerns that mercury in fish might be damaging to health, since organic mercury compounds are associated with heart disease and neurologic disorders, it is said that Daviglus et al. and Hu et al. did not try to allay these fears of eating fish. By failing to place the favorable all-cause death rates of avid fish eaters in their abstracts, or address the mercury issue, these authors did a disservice. Pregnant women have been cautioned to restrict their intake of fish (http://www.cbc.ca/storyview/CBC/2002/10/21/Consumers/mercuryfish_021021) despite evidence that children receive most of their mercury from vaccines.21,22

The Myocardial Ischemia Reduction with Aggressive Cholesterol Lowering (MIRACL) study, an RCT on the effects of 80 mg/day of atorvastatin or placebo on 3,086 patients in hospital after angina or nonfatal MI and followed for 16 weeks, had the following conclusions in the abstract: “For patients with acute coronary syndrome, lipid-lowering therapy with atorvastatin, 80 mg/day, reduces recurrent ischemic events in the first 16 weeks, mostly recurrent symptomatic ischemia requiring rehospitalization.” Actually this is true. The unmentionable findings were that there was no change in the death rate, and no significant change in either the reinfarction rate or need for resuscitation from cardiac arrest. There was a significant drop in chest pain requiring rehospitalization. The risk-ratio plot was unusual in not having a vertical bar at the 1.00 point, making the outcomes hard to visualize from this figure. The discussion did not give any comparisons with alternate treatments, for example, that five weeks of aspirin would give significantly lower reinfarction and all-cause mortality rates in men. Lowering cholesterol levels was highlighted, despite existing knowledge that the beneficial effects of statins on CHD are independent of either the baseline or achieved levels; thus these levels were a useless surrogate endpoint.

The Anglo-Scandinavian Cardiac Outcomes Trial (ASCOT) study, specifically on atorvastatin, did not mention in its abstract that: (1) women were worse off with treatment, the same as with aspirin;27 (2) that after 3.5 years there was no significant change in the all-cause death rate, marking atorvastatin as even less effective than pravastatin in the WOSCOPS trial; and (3) that total cardiovascular events and procedures were 2% lower with atorvastatin than with placebo after 3.5 years–instead they only gave the RR of 0.79. This is a poorer performance than that of Bufferin, for which RR = 0.31 for nonfatal MI in men. The control group had more previous stroke, TIA, diabetes, and other CHD, showing poor randomization.

A meta-analysis of 44 trials of atorvastatin intended to highlight its safety neglected to mention in its abstract that: (1) the median treatment period was only one year; (2) the all-cause death rate of 1% did not differ from that of placebo, and thus was not even as good as that of pravastatin in the WOSCOPS trial; (3) 65% of the treatment group vs. 45% of controls experienced an adverse event; (4) the total unadjusted withdrawal rate was 4% vs. 1% for placebo for all adverse events; (5) 10% of patients suffered serious adverse events vs 8% for placebo and (6) reduction in treatment-associated adverse cardiovascular events (a judgmental determination) was 1% absolute (from 2% on placebo to 1% on atorvastatin, exactly as in the ASCOT trial, a poorer performance than that of Bufferin in men).

The authors were thoughtful enough to provide the information, in the Methods section, that from April 1, 1998, the FDA allowed the exclusion of cancer and overdose from drug side-effects, both of which might have been significant with this drug based on the results of the Prospective Study of Pravastin in the Elderly at Risk (PROSPER) and Cholesterol and Recurrent Events (CARE) trials.

Misleading Abstract and Faulty Trial Protocol

In a study partially funded by the AMA, a prodigious literature review on articles in English on the efficacy and safety of low-carbohydrate diets was performed by use of Medline and other searches for those articles published between Jan. 1, 1966, and Feb. 15, 2003. All 2,609 potentially relevant articles were perused. All but 107 articles on 94 studies on 3,268 subjects receiving 0-901 g/day of carbohydrates for four to 365 days were excluded, but the reasons for exclusion of so many of the trials were obscure.

Only five studies–which were non-randomized and had no control groups–lasted more than 90 days. “...These [low-carbohydrate] diets have not been adequately evaluated for use longer than 90 days, or for individuals aged 53 years or older, or for use by participants with hyperlipidemia, hypertension, or diabetes.” Conclusions in the abstract, verbatim, are that “There is insufficient evidence to make recommendations for or against the use of low-carbohydrate diets, particularly among participants older than age 50, for use longer than 90 days, or for diets of 20 g/d or less of carbohydrates. Among the published studies, participant weight loss while using low-carbohydrate diets was principally associated with decreased caloric intake and increased diet duration, but not with reduced carbohydrate content.”

Of the two main “low-carbohydrate” groups into which the trials were divided, the ≤60 g/d groups’ mean intake of carbohydrate was 29 g/d, and total energy intake of all foods was 1,446 kcal/day. In the >60 g/d group the mean intake of carbohydrate was 236 g/d, and total energy intake of all foods was 1,913 kcal/day (their Table 3). In all 11 books on low-carbohydrate diets examined by this writer, any intake exceeding about 150-200 g/d of total carbohydrate would not be considered low-carbohydrate.

In the true low-carbohydrate group the mean weight loss in trials was 17 kg, while in the higher-carbohydrate group it was 2 kg (their Table 5). The authors do not consider this significant and attribute the result to the lower total caloric intake. This view has
been falsified in several studies. For example, controlled trials in hospitals have shown that a diet of just 1,000 kcal/day that is 90% carbohydrate led to weight gain, and intakes of 1,000 to 2,600 kcal/day with a very low carbohydrate content led to weight loss. Thus the conclusions should have been that low-carbohydrate diets are both safe and effective. Only by intermingling trials of low to medium and high-carbohydrate diets could the authors reach the conclusions quoted above.

A recent one-year diet trial supposedly designed to evaluate the Atkins diet examined 63 subjects, of whom the 33 assigned to the Atkins diet were given a copy of Atkins’s 2002 book and instructed to follow it, including no restriction on the amount of fat and protein. The 30 assigned to the low-fat diet – 60% carbohydrates, 25% fat, 15% protein by fuel values – were restricted to 1,200-1,500 kcal/day for women and 1,500-1,800 kcal/day for men, definitely a slimming diet. In conclusions in the abstract, verbatim, are: “The low-carbohydrate diet produced a greater weight loss (absolute difference, approximately 4%) than did the conventional diet for the first six months, but the differences were not significant at one year...” All subjects met with a registered dietician four times. Since registered dieticians are indoctrinated by the American Dietetic Association to promote high-carbohydrate diets, this variable was not properly controlled, since the controls would have had reinforcement (placebo effect) and low-carbohydrate subjects would not (nocebo effect).

In addition, subjects were excluded if they were ill, had non-insulin dependent diabetes (NIDDM), were taking lipid-lowering medications or ones that affect body weight, or were pregnant or lactating. In other words, many subjects who would have benefited the most from the Atkins diet were excluded; this was the most serious fault in the trial design. Nevertheless, there was more weight loss among subjects on the Atkins diet, highly significant at 3 and 6 months, but claimed not to be significant at 12 months using all participants, including those who did not complete the study, but whose values were extrapolated to 12 months so as to show non-significance.

The absolute weight loss difference was actually 3% between groups, favoring the low-carbohydrate group, at 12 months, among those actually completing the study, and this was shown as significant in their Fig. 1B. Low-carbohydrate dieters had increased high-density lipoproteins (HDL) and decreased triglycerides (TG). Adherence was poor and attrition high in both groups, but attrition was less in the low-carbohydrate group. More trials were recom-mended, and all the usual discredited shibboleths about low-carbohydrate diets were resurrected – kidney and liver damage, higher cholesterol intake – including the unfounded concerns about saturated fat consumption.

Relative vs. Absolute Risk

A prospective analysis of the relation between dietary fat intake and breast cancer risk among 90,655 premenopausal women in the Nurses’ Health Study II, followed for eight years, showed, according to the results in the abstract: “…Relative to women in the lowest quintile of fat intake, women in the highest quintile of intake had a slight increased risk of breast cancer (RR = 1.25, 95% CI = 0.98 to 1.59; P trend = 0.06). The increase was associated with intake of animal fat but not vegetable fat; the [multi-variables-adjusted] RRs for the increasing quintiles of animal fat intake were 1.00 (referent), 1.28, 1.37, 1.54, and 1.33 (95% CI = 1.02 to 1.73; P trend = 0.002)…” Conclusions in the abstract are: “Intake of animal fat, mainly from red meat and high-fat dairy foods during pre-menopausal years, is associated with an increased risk of breast cancer.”

The absence of absolute risk levels is misleading. Their Tables 1 and 2 allow us to calculate what the unadjusted chances not to get breast cancer are: Median energy percentage intake as animal fat/percentage who did not get breast cancer for each quintile: 12/99.3, 15/99.2, 17/99.2, 20/99.1, 23/99.3. Many of the adjustments used were of doubtful soundness, such as body-mass index and alcohol intake, and there was no trend.

Surrogate Endpoint (Hypertension) and Lack of Placebo

The Anti-Hypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT), in its antihypertensive arm, sought to compare a diuretic (chlorthalidone) with a calcium channel blocker (amlodipine) and an angiotensin converting enzyme (ACE) inhibitor (lisinopril) for the control of blood pressure (the surrogate endpoint) in high-risk patients. No placebos were used in this study that followed 33,357 subjects for a mean of 4.9 years. Primary outcomes were defined as fatal CHD or nonfatal MI, and were observed in 2,956 subjects. There was no difference in frequency found among the three treatments.

There were minor differences in stroke rates (1% absolute, amlodipine best) and in hospitalizations for “heart failure” (2% absolute, amlodipine best). The abstract stated: “Anti-hypertensive therapy is well established to reduce hypertension-related morbidity and mortality, but the optimal first-step therapy is unknown.” The conclusions were: “Thiazide-type diuretics are superior in preventing one or more major forms of CVD and are less expensive. They should be preferred for first-step anti-hypertensive therapy.”

It should be noted that chlorthalidone is not a thiazide. It had the greatest effect on systolic blood pressure, yet did not have the most favorable effect on any outcome (their Fig. 4). The World Health Organization does not think older adults should use this drug because the risk of serious side effects is so high.

An older trial, also supported by drug companies, on 17,354 subjects with worse hypertension than in ALLHAT and with six years of follow-up, compared placebo, a beta blocker (propranolol), and a thiazide diuretic (bendroflouazide). There was no difference in the all-cause death rate and minor reductions in the rates of strokes and all cardiac events in the treatment groups. It was interesting that placebo reduced blood pressure significantly, but there was no correlation of the amount of reduction with death rates, another indication of a useless surrogate endpoint.

Another trial examined 484 randomly selected hypertensive men, except that they were all aged 68 at the beginning. Many of them were taking a wide spectrum of anti-hypertensive drugs, mainly from the thiazide diuretic and beta-blocker classes. They were followed for 10 years and had a cardiac-event-free survival of 65%, while those men not on medication had an 82% survival. Those with diastolic pressure at baseline <90mm Hg had a RR = 4 with treatment, and even those with >90mm Hg fared worse with treatment. Funding was mostly from non-drug-company foundations and a Swedish government agency.

The older trials show that it is not certain that antihypertensive drugs lower morbidity or mortality. It is obvious that the ALLHAT conclusions should have been: No standard treatment with prescription antihypertensive agents is worthwhile.
Conclusions

Failure to compare trial results with earlier work is bad science. This is compounded when reports of trial results do not make comparisons with over-the-counter drugs or supplements, as shown above.

Expensive RCTs may have flaws in design or be subject to the impossibility of double blinding, as in the case of anticholesterol and anticancer drugs, because of their unmistakable side effects, which provide a strong placebo effect. Even lack of placebo controls is sometimes falsely justified by a perverted medical ethic, and assertion that the standard treatment is actually beneficial can lead to misleading results, as shown above. Sometimes clinical observations depict outcomes in chronic diseases more accurately than RCTs.

In an effort to expose financially based conflicts of interest, major medical journals have adopted lengthy forms to reveal all funding sources and authors’ ties to commercial sponsors of research. This is certainly worthwhile, but this has not stopped bias in the resulting papers.

Examples of failure to use absolute risk, of failure to provide all-cause death rates or NNT, of faulty trial protocols, of misleading abstracts (mostly by omission), of worthless surrogate endpoints, and of lack of placebo were easily found among well-funded trials in major medical journals, demonstrating bias or even scientific misconduct. The breakdown of the peer-review system that allows publication of so many flawed articles is of great concern.

Joel M. Kaufman, Ph.D. is Professor of Chemistry Emeritus at the University of the Sciences, 600 S. 43rd St., Philadelphia, PA 19104-4495. E-mail: kauffman@hsrc.org.

Acknowledgements: Jane M. Orient, M.D., and Uffe Ravnskov, M.D., Ph.D., provided encouragement and advice.

Potential conflicts of interest: The author has neither financial interest in any drug, diet, or treatment mentioned, nor in any alternate treatments.

REFERENCES
37. ALLHAT. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs. diuretic. JAMA 2002;288:2981-2997.