What's really wrong with medical research and how to fix it.

Richard David Feinman, SUNY Downstate Medical Center, Brooklyn, NY

Introduction

Something is fundamentally wrong with the medical literature. It's not just that authors misunderstand p values. It's not that studies are under-powered. Misuse of statistics remains common enough but a general uncertainty and mistrust appear to have more global roots. Richard Horton, editor-in-chief of *The Lancet* points to "studies with small sample sizes, tiny effects, invalid exploratory analyses, and flagrant conflicts of interest." [1] Horton's critique remains only one of many [1-4], some of which make the case that half of published material may be in error [1-2]. Traditional criteria, both from common sense and established principles such as Bradford Hill's criteria [5] are known but disregarded. In my view, blame falls on editors whose failure to recognize controversial subjects allows *de facto* (or intentional) bias. Alternative points of view to published papers appear in PubPeer, in Letters-to-editors and blogs and social media from people with experience and credentials.

The diet-heart hypothesis. Is there a null hypothesis?

Nutrition research remains the most salient problem area in the medical literature. Particularly damaging, the diet-heart hypothesis, holds that fat, specifically saturated fat, raises total cholesterol or some other lipid fraction that, in turn, causes, heart disease. Numerous experimental and clinical attempts to test the idea have failed or, more accurately, have succeeded in showing that the hypothesis cannot be supported. Negative results, however, have been ignored or sometimes cited as if they were supportive. Many reviews and critiques have described a state of poor practice that can only be considered a crisis (e.g. [6-10]).
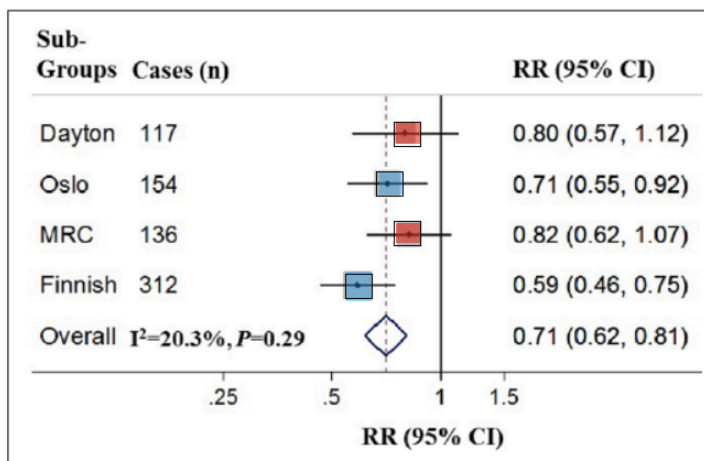
Meta-analysis to the rescue.

A recent case that pinpoints an apparent breakdown in standards or, at least, questionable peer review, the "Presidential Advisory" from the American Heart Association (AHA), claimed that if saturated fat (SF) is replaced with polyunsaturated fat (PUFA) benefit of cardiovascular risk will result [11]. Contradictory evidence was dismissed. In particular, the Advisory does not cite

two large meta-analyses [12-13] of 20 or so studies that show no significant association. Instead, the AHA advisory zeroed in on four "core" studies; the rather strange meta-analysis is shown in **Figure 1**.

Meta-analysis, in my view, remains highly questionable. Assuming that bigger n is always better and that the average of several studies, statistically significant or not, will be always be statistically significant, would need strong proof. While larger n may mathematically reduce error, common sense says that variability will increase and the persistent notion that you are trying to average several wrongs to make a right is hard to escape. The AHA paper is particularly unusual.

Remarkably, two of the four core studies showed no statistical difference between saturated and unsaturated — the statistical rule states that if the error bars (horizontal lines in the figure) cross the line for a relative risk (RR) of 1.0, this indicates no difference between experimental and controls. The two studies that did show statistical differences between SF and PUFA, the Oslo-Heart [14] and Finnish Mental Health studies [15], have been consistently criticized if not thoroughly discredited [6-10,16]. In the Finnish Mental Health study, the experimental population changed over time. So, a patient admitted in the middle of the study might have died before any substantial consumption of the diet.

Figure 1. Meta-analysis of four "core" trials from the AHA Presidential Advisory. Points to the left of RR = 1 line indicate benefit of replacing SF. Error bars that cross the line are considered not statistically significant (red symbols).

The Oslo-Heart study described a relative risk (RR) of SF *vs*. UF of 1.19, or odds of 54:46. Weak at best. (How much would you change your eating habits

over such odds?) At the same time, the RR for smokers *vs* non-smokers in the study, was found to be 2.2. The authors concluded that "a three-fold greater CHD mortality rate is demonstrable among the *hypercholesterolemic, hypertensive smokers*." (my emphasis). In other words, risk came from conflating the very weak effect of fat with the more significant impact of smoking.

Produced for the AHA by members of the AHA administration and reviewed by an AHA committee, the Presidential Advisory appeared in *Circulation*, a journal published by the AHA. The paper was widely reported and was taken as a consensus of experts rather than a particular point of view of a private health agency.

What's wrong here.
1. While widely used, meta-analysis is highly questionable. Whether two studies are sufficiently similar that the results can be averaged and will always improve accuracy remains an untested assumption and has no agreed-upon method of testing, More significant, averaging data cannot change bad studies into good studies.
2. Small relative risk cannot be considered significant. For comparison, in Bradford Hill's classic study, the odds ratio (OR) for lung cancer was 20:1 for smokers compared to non-smokers. And, in a toxic tort case in a court of law, epidemiological data would not be accepted as evidence at less that 2:1.
3. While not universally codified, it is generally recognized that references to alternative points of view are required. Similarly, "Cherry-picking" — accepting those studies that support your conclusion and ignoring those that don't — are not acceptable (e.g. [17]).

Alice has 30 % more money than Bob. Is she rich?
Without going into the details of the statistics, in disease states or other outcomes where the incidence is low, odds ratio (OR), hazard ratio (HR) and relative risk (RR) are roughly the same. The small RR in studies like the Oslo-Heart Study speak to weak effects by themselves. However, even large relative values are almost always misleading. The principle that relative parameters hide information — Alice may have more money than Bob but they may both be on welfare — has general agreement, but the idea is more common in the breach than the observance. Absolute difference in outcome between experimental and controls should always be reported [3].

Red Meat scare.

TABLE 1
Baseline age-adjusted characteristics of participants in the 3 cohorts according to quintile (Q)

	HPFS (1986)		
Characteristics	Q1 (*n* = 7187)	Q3 (*n* = 7027)	Q5 (*n* = 7247)
Total red meat intake (servings/d)	0.24 (0.11–0.37)[2]	1.01 (0.92–1.10)	2.16 (1.95–2.57)
Age (y)	53.6 ± 9.6[3]	52.4 ± 9.5	51.9 ± 9.2
Physical activity (MET-h/wk)	27.8 ± 34.0	20.2 ± 28.3	17.7 ± 25.2
BMI (kg/m^2)	24.7 ± 3.0	25.4 ± 3.1	25.9 ± 3.4
Race, white (%)	93.0	95.4	96.0
Current smoker (%)	5.1	9.8	14.3

TABLE 2
HRs (95% CI) of type 2 diabetes risk according to quintile (Q) of red meat Health Professionals Follow-Up Study (HPFS)

	Frequency of consumption				
	Q1	Q2	Q3	Q4	Q5
Processed red meat					
Servings/d	0.02 (0, 0.05)	0.12 (0.09, 0.13)	0.21 (0.20, 0.25)	0.38 (0.33, 0.44)	0.72 (0.60, 0.97)
Cases/person-years	340/138,550	409/121,238	441/131,831	593/131,520	655/129,834
Age-adjusted model	1.00	1.38 (1.20, 1.60)	1.40 (1.22, 1.62)	1.88 (1.65, 2.15)	2.08 (1.82, 2.37)
Multivariate model 1	1.00	1.32 (1.14, 1.53)	1.32 (1.14, 1.52)	1.75 (1.51, 2.01)	1.96 (1.69, 2.27)
Multivariate model 2	1.00	1.18 (1.02, 1.37)	1.12 (0.97, 1.30)	1.44 (1.25, 1.66)	1.55 (1.33, 1.79)

Figure 2. Data from Tables 1 and 2 from Pan, *et al*. [8]. Subjects in quintile 5 (most red meat consumption) are considered to have had a "positive test" by analogy with a diagnostic test. Those subjects who developed type 2 diabetes are considered as "true positives."

Pan, *et al*. [18] studied the effect of red meat consumption on the incidence of type 2 diabetes. Three cohorts were followed, the Health Professionals Follow-Up Study, (HPFS) and two results from the Nurses' Health Studies (NHS I and NHS II).

The authors claimed that "red meat intakes were positively associated with type 2 diabetes (T2D) risk with HRs of   1.12, 1.32 and 1.14 for unprocessed, processed, and total red meat. They concluded that their "results suggest that red meat consumption, particularly processed red meat, is associated with an increased risk of T2D."

What's wrong with this picture?

Although is not immediately obvious, the data for calculating the absolute differences in risk are provided in Pan, *et al*.. Figure 2. shows data for the HPFS. Data are arranged in five groups (quintiles) according to red meat consumption. Table 1 shows that the number of high red meat eaters (Q5) was 7,247. From Table 2, we see that, of these, 655 were diagnosed with diabetes. One could think of this analysis as similar to a diagnostic test [3] and we can calculate a positive predictive value, the ratio of "true positives" (high red meat *and* diabetes) to all "positives" (everybody in Q5) = 655 / 7,247 = 0.090 or 9 %. Such a low incidence provides little support for the authors' conclusion that red meat consumption constitutes a risk for type 2 diabetes. To understand how poorly predictive this is, we can calculate the same positive predictive value for the group that consumed red meat at the low level (first quintile) = 340 / 7,187 = 0.047 or 4.7 %. The difference in risk between extremes of red meat consumption is 4.3 %.

Now the difference in absolute risk is very small but it is not zero. It is often claimed that if this difference were scaled up to a large population, "thousands of lives would be saved." The assumption is that 4.3 % represents a small, but valid number. That would be true if the independent variable, the assignment to different quintiles, were reliable, that is, if we have accurate dietary records. In fact, "Diet was assessed by validated food-frequency questionnaires, and data were updated every 4 y." Food-questionnaires are, in fact, notably unreliable, especially if a point is taken only every four years — and, in the current case, "validation" does not mean certification of no error but rather statement of the inaccuracy which, as one would expect, is never in the single digit range [19].

The most serious fault of the medical literature rests with acceptance of such questionable statistical analysis in the face of contradictory evidence from experiment or even common sense. In the thirty years from 1980 to 2010, per capita red consumption *decreased* from 127 lbs. to 106 lbs, while type 2 diabetes increased from 5.5 million to 21 million and is even higher now. Such population data has to be considered as at least as epidemiologically sound as the cohort studies. While also subject to error, the changes are large while the cohort studies provide much smaller effects.

What's wrong here.
1. Statistical analysis does not provide information about biological or clinical relevance. Experimenters must defend their interpretation based on established science and previous work, that is real outcomes [5].
2. Relative values hide information. Differences in absolute risk should always be shown in addition to OR, HR or RR [3].

What can be done?

1. The publication model of (physical) scientific literature is not appropriate for medical studies where there is controversy and where results are more tentative. Medical reports require post-publication review or periods of tentative acceptance subject to continued public comment [1].

2. External universally respected, neutral organizations are needed to codify good practice and to adjudicate conflicts. In the mean time…

3. Editors must recognize controversial MSS and appoint reviewers from both sides of the controversy. Failure to do so constitutes *de facto* (or intentional bias).

References

1. Horton R: **Offline: What is medicine's 5 sigma?** *The Lancet* 2015, **385**:1380.

2. Ioannidis JP: **Why most published research findings are false**. *PLoS Med* 2005, **2**(8):e124.

3. Gigerenzer G: **Misleading communication of risk. Editors should enforce transparent reporting in abstracts**. *BMJ* 2010, **341**(c4830).

4. Feinman RD, Keough SM: **Ethics in Medical Research and the Low-Fat Diet-Heart Hypothesis**. *Ethics in Biology, Engineering and Medicine* 2014, **5**(2):149-159.

5. Hill AB: **The Environment and Disease: Association or Causation?** *Proceedings of the Royal Society of Medicine* 1965, **58**:295-300.

6. Ravnskov U, Diamond D, Karatay MC, Miller DW, Okuyama H: **No scientific support for linking dietary saturated fat to CHD**. *Br J Nutr* 2012, **107**(3):455-457; author reply 458-460.

7. Ravnskov U: **The Cholesterol Myths: Exposing the Fallacy that Cholesterol and Saturated Fat Cause Heart Disease**. Washington, DC: NewTrends Publishing, Inc.; 2000.

8. Elliott J: **Flaws, Fallacies and Facts: Reviewing the Early History of the Lipid and Diet/Heart Hypotheses**. *Food and Nutrition Sciences* 2014, **5**:1886-1903.

9. Taubes G: *Good Calories, Bad Calories*. New York: Alfred A. Knopf; 2007.

10. Teicholz N: *The Big Fat Surprise. Why Butter, Meat & Cheese Belong in a Health Die**t***. New York: Simon & Schuster; 2014. *http://www.ravnskov.nu/cm/*

11. Sacks FM, Lichtenstein AH, Wu JHY, Appel LJ, Creager MA, Kris-Etherton PM, Miller M, Rimm EB, Rudel LL, Robinson JG *et al*: **Dietary Fats and Cardiovascular Disease: A Presidential Advisory From the American Heart Association**. *Circulation* 2017, **136(3):e1-e23.**

12. Siri-Tarino PW, Sun Q, Hu FB, Krauss RM: **Meta-analysis of prospective cohort studies evaluating the association of saturated fat with cardiovascular diseas**e. *Am J Clin Nutr* 2010, 91(3):535-546.

13. Jakobsen MU, O'Reilly EJ, Heitmann BL, Pereira MA, Balter K, Fraser GE, Goldbourt U, Hallmans G, Knekt P, Liu S *et al*: **Major types of dietary fat and risk of coronary heart disease: a pooled analysis of 11 cohort studies**. *Am J Clin Nutr* 2009, **89**(5):1425-1432.

14. Leren P: The Oslo diet-heart study. Eleven-year report. *Circulation* 1970, 42(5):935-942.

15. Miettinen M, Turpeinen O, Karvonen MJ, Elosuo R, Paavilainen E: **Cholesterol-lowering diet and mortality from coronary heart-disease**. *Lancet* 1972, **2**(7792):1418-1419.

16. Feinman RD: *The World Turned Upside Down: The Second Low-Carbohydrate Revolution*. Brooklyn, NY: NMS Press; 2014.

17. Masic I. **The Importance of Proper Citation of References in Biomedical Article**s. *Acta Informatica Medica*. 2013;21(3):148-155.

18. Pan A, Sun Q, Bernstein AM, Schulze MB, Manson JE, Willett WC, Hu FB: **Red meat consumption and risk of type 2 diabetes: 3 cohorts of US adults and an updated meta-analysis**. *The Am J Clin Nutrition* 2011, **94**(4):1088-1096.

19. Paul, DR, Rhodes, DG, Kramer, M, Baer, DJ, Rumpler, WV **Validation of a Food Frequency Questionnaire by Direct Measurement of Habitual ad Libitum Food Intake** *American Journal of Epidemiology* 162(8):

20. Ioannidis JP, Tatsioni A, Karassa FB: **Published articles should not be dead and buried: introducing research updates**. *Eur J Clin Invest* 2010, **40**(9):767-769.