Types of evidence for diet and chronic disease, session 2

Webinar, July 10, 2025

John P.A. Ioannidis, MD, DSc

Professor of Medicine, of Epidemiology and Population Health, and (by courtesy) of Biomedical Data Science
Co-Director, Meta-Research Innovation Center at Stanford (METRICS)
Stanford University

Disclosures

- No conflicts of interest

Nutri* OR diet OR food

- 2,711,275 papers indexed in PubMed as of Jul 8, 2025
- This includes 34,829 systematic reviews
- While most studies are observational non-randomized, 81,357 papers are classified as randomized controlled trials

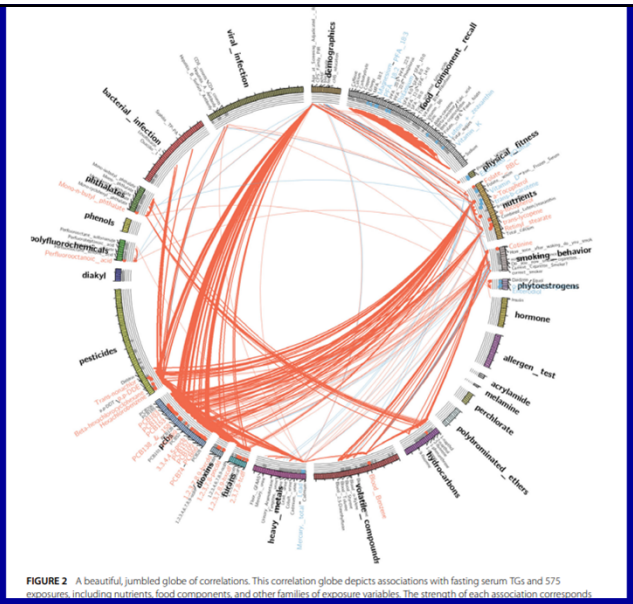


FIGURE 2 A beautiful, jumbled globe of correlations. This correlation globe depicts associations with fasting serum TGs and 575 exposures, including nutrients, food components, and other families of exposure variables. The strength of each association corresponds

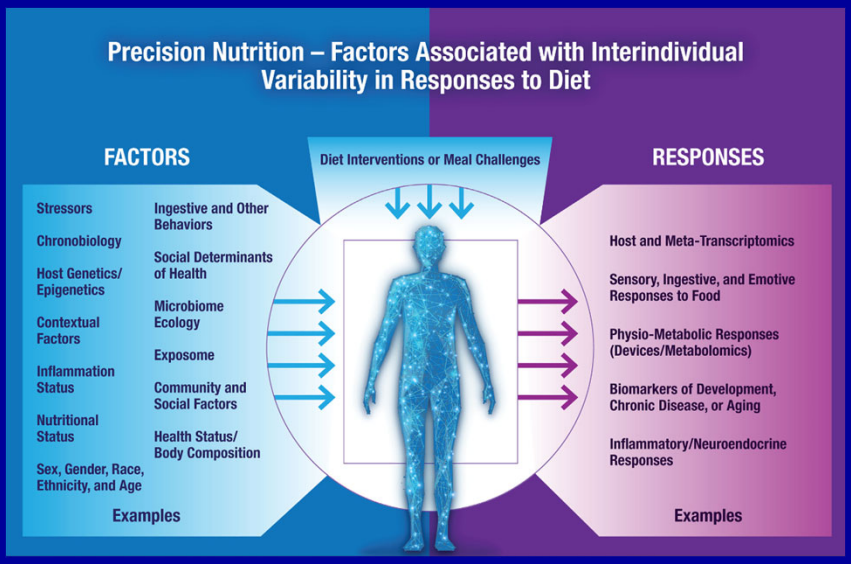
Nutrition effects are very complex and multifaceted

Perspective: Limiting Dependence on Nonrandomized Studies and Improving Randomized Trials in Human Nutrition Research: Why and How

John F Trepanowski[1] and John PA Ioannidis[1-2,3,4,5,6]

[1]Stanford Prevention Research Center; [2]Meta-Research Innovation Center at Stanford (METRICS); and Departments of [3]Medicine, [4]Health Research and Policy, [5]Biomedical Data Science, and [6]Statistics, Stanford University, Stanford, CA

…and personalized



Lee et al. AJCN 2022

5

Still most studies look at epidemiological averages

VIEWPOINT

The Challenge of Reforming Nutritional Epidemiologic Research

John P. A. Ioannidis, MD, DSc
Stanford Prevention Research Center and Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, California.

Some nutrition scientists and much of the public often consider epidemiologic associations of nutritional factors to represent causal effects that can inform public health policy and guidelines. However, the emerging picture of nutritional epidemiology is difficult to reconcile with good scientific principles. The field needs radical reform.

In recent updated meta-analyses of prospective cohort studies, almost all foods revealed statistically significant associations with mortality risk.[1] Substantial deficiencies of key nutrients (eg, vitamins), extreme overconsumption of food, and obesity from excessive calories may indeed increase mortality risk. However, can small intake differences of specific nutrients, foods, or diet patterns with similar calories causally, markedly, and almost ubiquitously affect survival?

Assuming the meta-analyzed evidence from cohort studies represents life span–long causal associations, for a baseline life expectancy of 80 years, nonexperts prelyze in very different ways.[4] Consequently, meta-analyses become weighted averages of expert opinions. In an inverse sequence, instead of carefully conducted primary studies informing guidelines, expert-driven guidelines shaped by advocates dictate what primary studies should report. Not surprisingly, an independent assessment by the National Academies of Sciences, Engineering, and Medicine of the national dietary guidelines suggested major redesign of the development process for these guidelines: improving transparency, promoting diversity of expertise and experience, supporting a more deliberative process, managing biases and conflicts, and adopting state-of-the-art processes.[5]

Proponents of the status quo may maintain that the true associations are even larger than what are reported because of attenuation from nondifferential misclassification. Indeed, self-reported data have error,[6] but there is no guarantee it is nondifferential. Nevertheless, if error is nondifferential and estimated effects are

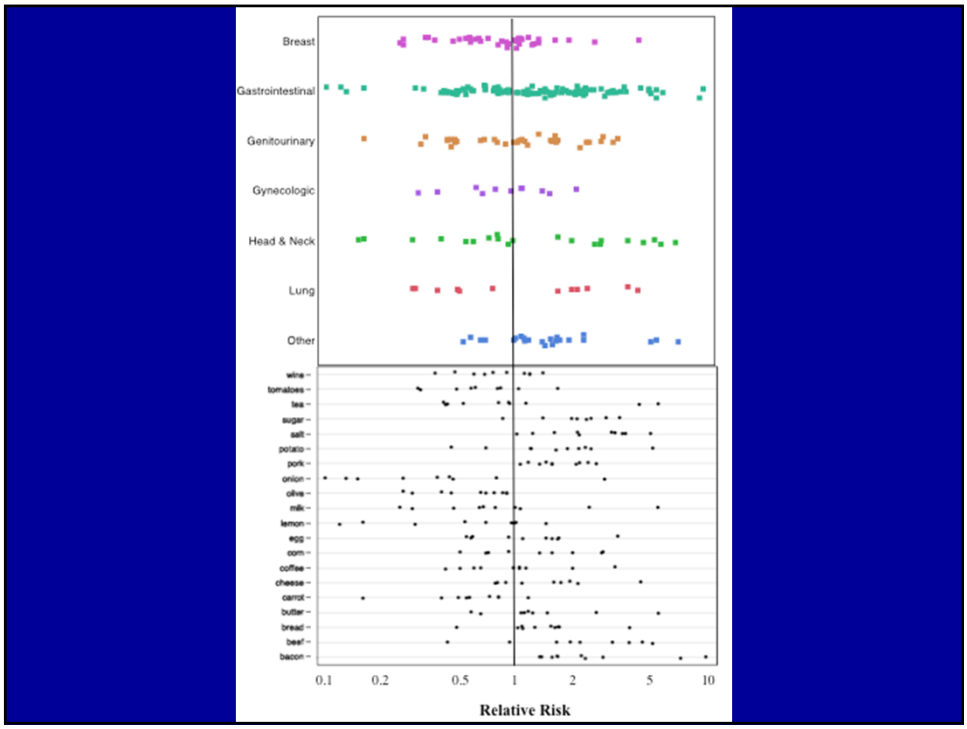
6

Diet causes cancer

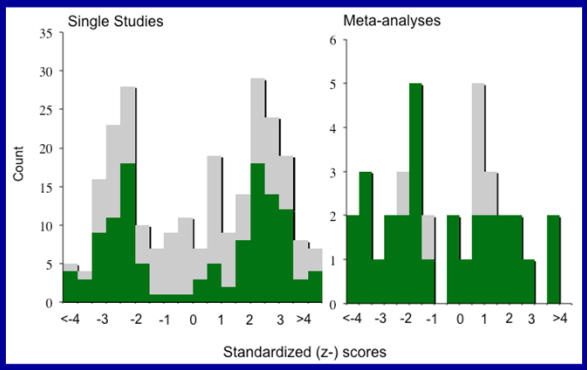
- Open a popular cookbook
- Randomly check 50 ingredients
- How many of those are associated with significantly increased or significantly decreased cancer risk in the scientific literature?

Associated with cancer risk

- veal, salt, pepper spice, flour, egg, bread, pork, butter, tomato, lemon, duck, onion, celery, carrot, parsley, mace, sherry, olive, mushroom, tripe, milk, cheese, coffee, bacon, sugar, lobster, potato, beef, lamb, mustard, nuts, wine, peas, corn, cinnamon, cayenne, orange, tea, rum, raisin



9



Schoenfeld and Ioannidis, AJCN 2013

10

Death by food: it can all kill you

TABLE 1
Relative risks from nonlinear dose-response analysis of 12 predefined food groups and all-cause mortality according to servings per day[1]

	Servings per day						
Associations by food group	0	1	2	3	4	5	6
Inverse association							
Whole grains (30 g/d)	1.00	0.91 (0.89, 0.92)	0.84 (0.82, 0.86)	0.79 (0.76, 0.83)	NA	NA	NA
Vegetables (80 g/d)	1.00	0.94 (0.93, 0.96)	0.91 (0.89, 0.93)	0.89 (0.87, 0.92)	0.89 (0.87, 0.91)	0.89 (0.87, 0.91)	0.89 (0.86, 0.92)
Fruit (80 g/d)	1.00	0.94 (0.93, 0.96)	0.91 (0.89, 0.93)	0.90 (0.88, 0.93)	0.91 (0.88, 0.93)	0.92 (0.89, 0.94)	0.92 (0.89, 0.95)
Nuts (28 g/d)	1.00	0.85 (0.82, 0.89)	NA	NA	NA	NA	NA
Legumes (100 g/d)	1.00	0.90 (0.85, 0.96)	NA	NA	NA	NA	NA
Fish (100 g/d)	1.00	0.93 (0.90, 0.96)	0.90 (0.84, 0.96)	NA	NA	NA	NA
Positive association							
Eggs (55 g/d)	1.00	1.07 (1.01, 1.15)	NA	NA	NA	NA	NA
Red meat (85 g/d)	1.00	1.16 (1.14, 1.18)	1.35 (1.32, 1.38)	NA	NA	NA	NA
Processed meat (30 g/d)	1.00	1.12 (1.10, 1.14)	1.20 (1.17, 1.23)	1.28 (1.23, 1.32)	1.35 (1.28, 1.41)	NA	NA
Sugar-sweetened beverages (250 mL/d)	1.00	1.07 (1.01, 1.14)	NA	NA	NA	NA	NA
Inverse and positive association							
Dairy (200 g/d)	1.00	0.97 (0.95, 0.99)	0.99 (0.97, 1.01)	1.04 (1.01, 1.07)	1.11 (1.05, 1.17)	1.16 (1.08, 1.23)	NA
No association							
Refined grains (30 g/d)	1.00	0.96 (0.92, 1.01)	0.96 (0.90, 1.02)	0.97 (0.91, 1.05)	1.00 (0.92, 1.08)	1.03 (0.92, 1.16)	NA

[1] Values are risk ratios (95% CIs). NA, not applicable.

11

If you eat that many hazelnuts each day, your death risk decreases by 15% (?!)

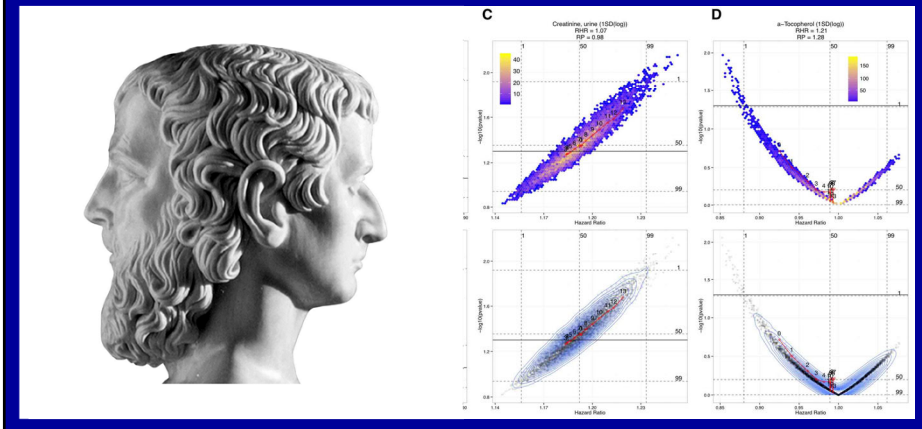
12

If you drink three cups of coffee every day your death risk decreases by 15% (?/!)



13

Almost any result can be obtained: Vibration of effects and the Janus phenomenon



Patel, Burford, Ioannidis. JCE 2015; Patel and Ioannidis, JAMA 2015

14

International Journal of Epidemiology, 2021, 266–278
doi: 10.1093/ije/dyaa164
Advance Access Publication Date: 5 November 2020
Original article

Methods

Examining the robustness of observational associations to model, measurement and sampling uncertainty with the vibration of effects framework

Simon Klau [1,2,*,†] Sabine Hoffmann,[1,3†] Chirag J Patel,[4] John PA Ioannidis,[5,6,7,8,9] and Anne-Laure Boulesteix[1,3]

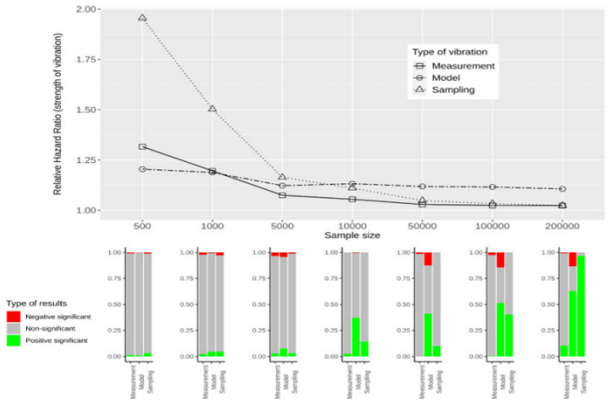


Figure 8 Measurement, model and sampling vibration for different sample sizes (top panel), and bar plots showing the type of results in terms of significance of estimated effects (bottom panel) for the association of HDL-cholesterol with mortality

15

RCTs of lifestyle, non-regulated interventions: many, but fragmented and often non-registered

BMJ 2015;350:h1323 doi: 10.1136/bmj.h1323 (Published 27 March 2015)

ANALYSIS

Call to improve transparency of trials of non-regulated interventions

The public and clinicians require transparent, quality evidence for all interventions. Trials of non-regulated interventions are common, and efforts to improve their registration and publication compared with drug trials are overdue, say **Rafael Dal-Ré, Michael Bracken,** and **John Ioannidis**

16

Long-term randomized trials in nutrition are doable, and indispensable

TABLE 1 ATBC study (enrolled in 1985–1988): initial and postintervention-period results[1]

	β-Carotene		α-Tocopherol	
	All deaths	Lung cancer	All deaths	Prostate cancer
Intervention to April 1993 (72)	1.08 (1.01, 1.16)	1.18 (1.03, 1.36)	1.02 (0.95, 1.09)	0.68 (0.53, 0.88)
Postintervention				
To April 1999 (73)	—	1.06 (0.94, 1.20)	—	0.88 (0.76, 1.03)
To April 2001 (73)	1.07 (1.02, 1.12)	—	1.01 (0.96, 1.05)	—
To December 2009 (74)	1.02 (0.99, 1.05)	1.04 (0.96, 1.01)	1.02 (0.98, 1.05)	0.97 (0.89, 1.05)

[1]Values are relative risk (95% CIs). ATBC, Alpha-Tocopherol, Beta-Carotene Cancer Prevention.

17

Pilot trials need to be assessed for generalizability biases

Identification and evaluation of risk of generalizability biases in pilot versus efficacy/effectiveness trials: a systematic review and meta-analysis

Michael W. Beets[1*], R. Glenn Weaver[1], John P. A. Ioannidis[2], Marco Geraci[1], Keith Brazendale[1], Lindsay Decker[1], Anthony D. Okely[3], David Lubans[4], Esther van Sluijs[5], Russell Jago[6], Gabrielle Turner-McGrievy[1], James Thrasher[1], Xiaming Li[1] and Andrew J. Milat[7,8]

Results: A total of 39 pilot and larger trial pairs were identified. The frequency of the biases varied: delivery agent bias (19/39 pairs), duration bias (15/39), implementation support bias (13/39), outcome bias (6/39), measurement bias (4/39), directional conclusion bias (3/39), target audience bias (3/39), intervention intensity bias (1/39), and setting bias (0/39). In meta-analyses, delivery agent, implementation support, duration, and measurement bias were associated with an attenuation of the effect size of − 0.325 (95CI − 0.556 to − 0.094), − 0.346 (− 0.640 to − 0.052), − 0.342 (− 0.498 to − 0.187), and − 0.360 (− 0.631 to − 0.089), respectively.

Conclusions: Pre-emptive avoidance of RGBs during the initial testing of an intervention may diminish the voltage drop between pilot and larger efficacy/effectiveness trials and enhance the odds of successful translation.

18

Impact of risk of generalizability biases in adult obesity interventions: A meta-epidemiological review and meta-analysis

Obesity Rev, 2022

Michael W. Beets[1] | Lauren von Klinggraeff[1] | Sarah Burkart[1] | Alexis Jones[1] | John P. A. Ioannidis[2] | R. Glenn Weaver[1] | Anthony D. Okely[3] | David Lubans[4] | Esther van Sluijs[5] | Russell Jago[6] | Gabrielle Turner-McGrievy[7] | James Thrasher[7] | Xiaoming Li[7]

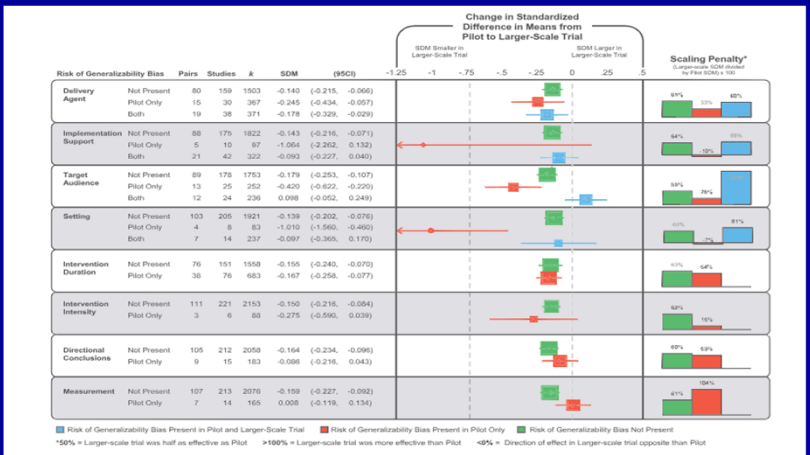


FIGURE 4 Forest plot of the change in the standardized difference in means (SDM) of the presence, absence, or carry forward of risk of generalizability biases from a pilot/feasibility study to a larger scale trial. No pairs contained directional conclusion bias in both the pilot and larger scale trial. Intervention duration, intervention intensity, and measurement describe differences between smaller and larger scale studies, so they cannot be present in both studies

19

Precision on top? E.g. N-of-1 trials were placed at the top in the mid-90s

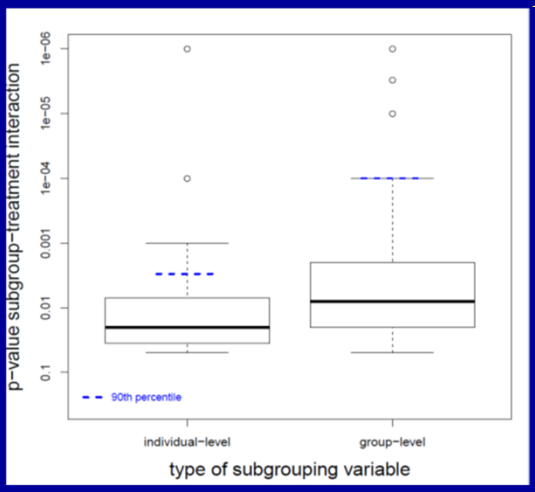
20

Why were N-of-1 trials largely abandoned 30 years ago?

- Not good if the disease/condition does not have a steady natural history
- Not good if there is carry over effect
- Not good if there are priming effects and if effects depend on previous choices
- Not good if the disease has a fatal outcome and a relatively short course
- Not good if there is poor/unpredictable compliance/adherence/tolerability

21

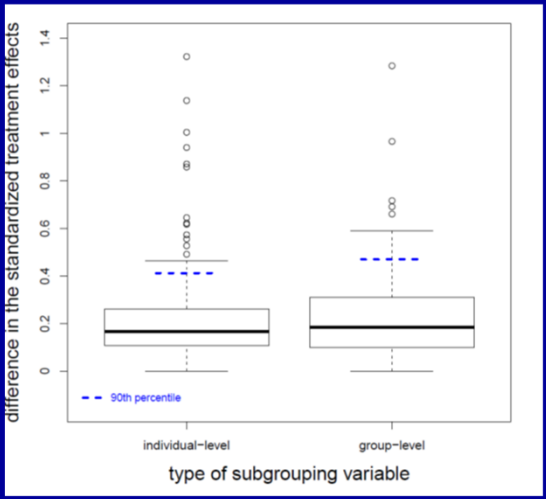
Subgroup differences in large-scale MIPDs: few and with low support



Schuit et al, Int J Epidemiol 2018

22

Treatment effect modifications for individual and group level subgrouping variables: typically small



23

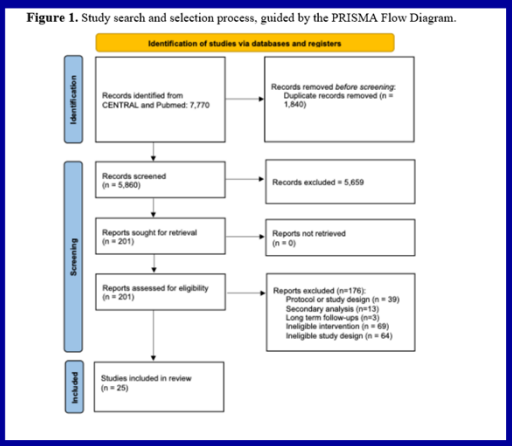
Biomarker-driven precision trial designs

	Enrichment	Randomize-all	Adaptive design	Umbrella	Basket
Histology	dependent	dependent	dependent	dependent	independent
Number of targeted therapies	1	1	≥1	>1	≥1
Number of biomarkers	1	1	≥1	>1	≥1
Type of biomarkers	Bm+	Bm+ and Bm−	Bm+ and Bm−	Bm+ if exploratory Bm+ and Bm− if confirmatory	Usually Bm+
Biomarker credentials (a priori knowledge)	very strong	+/−	+/−	strong	very strong
Biomarker assay	single, locally	single, locally	single, locally	multiplex, centralized	single, locally
Provides information on the Biomarker-treatment benefit association (is the biomarker predictive?)	−	+/−	+/−	+	−
Number of patients required to screen	Prevalence-dependent	Prevalence-dependent	Prevalence-dependent	Prevalence-dependent	Prevalence-dependent
Sufficiently large sample size (depends on the rarity of the mutation)*	+	++	+/−	+++	++
Overlap of patients	−	+/−	+/−	+	−
Statistical complexity	+	+	+++	++	++
Tradeoff between power versus sample size	−	+	++	+++	+++
Subgroup analyses – multiplicity	−	−	+++	++	+++
Type 1 error problems	+	+	++	++	+++
Flexibility†	−	−	+++	+	+
Time efficiency and cost savings	−−	−−	+	++	++

Janiaud, Serghiou, Ioannidis, Cancer Treatment Reviews 2019

24

Precision genetics? A systematic review of 25 RCTs on polygenic risk scores



25

Conclusions of the 25 RCTs

- 13 had favorable claims
- 5 claimed further research is warranted
- 7 stated no significant differences were found

26

All meta-analyses show null benefits

Outcome	Measure	N° of studies (intervention; control)	value	95%CI	I
Total cholesterol	mean difference	2 (391; 391)	-2.01	[-8.27; 4.26]	0%
HDL cholesterol	mean difference	3 (411; 437)	-0.21	[-2.65; 2.23]	1.7%
LDL cholesterol	mean difference	4 (514; 537)	-3.64	[-7.88; 0.60]	0%
Sistolic Blood Pressure	mean difference	3 (247; 276)	-1.26	[-4.44; 1.92]	10.1%
Diastolic Blood Pressure	mean difference	3 (247; 276)	-1.88	[-4.17; 0.42]	0%
Weight	mean difference	4 (482; 443)	-0.33	[-0.87; 0.20]	0%
BMI	mean difference	3 (319; 271)	-0.12	[-0.64; 0.39]	0%
Physical activity	SDM	4 (508; 511)	-0.01	[-0.13; 0.11]	0%
Anxiety	SDM	5 (702; 671)	-0.02	[-0.13; 0.08]	0%
Perceived risk	SDM	2 (379; 379)	-0.10	[-0.40; 0.19]	77.2%
Worry	SDM	3 (718; 733)	-0.06	[-0.23; 0.10]	42.4%
Peak sun exposure	SDM	2 (533; 548)	0.02	[-0.10; 0.14]	0%
Intentional tanning	SDM	4 (1109; 1120)	-0.03	[-0.11; 0.06]	0%
Total Sun exposure	SDM	3 (1091; 1101)	0.00	[-0.09; 0.08]	0%
Sun Protection	SDM	2 (533; 548)	0.04	[-0.08; 0.16]	0%
Incidence of the disease	relative risk	3 (467; 418)	0.945	[0.320; 2.79]	68.3%
Screening attendance	relative risk	4 (990; 748)	1.115	[0.770; 1.61]	25.5%

27

Nutrition science would benefit from reproducible practices

OPEN

A manifesto for reproducible science

Marcus R. Munafò[1,2*], Brian A. Nosek[3,4], Dorothy V. M. Bishop[5], Katherine S. Button[6], Christopher D. Chambers[7], Nathalie Percie du Sert[8], Uri Simonsohn[9], Eric-Jan Wagenmakers[10], Jennifer J. Ware[11] and John P. A. Ioannidis[12,13,14]

Improving the reliability and efficiency of scientific research will increase the credibility of the published scientific literature and accelerate discovery. Here we argue for the adoption of measures to optimize key elements of the scientific process: methods, reporting and dissemination, reproducibility, evaluation and incentives. There is some evidence from both simulations and empirical studies supporting the likely effectiveness of these measures, but their broad adoption by researchers, institutions, funders and journals will require iterative evaluation and improvement. We discuss the goals of these measures, and how they can be implemented, in the hope that this will facilitate action toward improving the transparency, reproducibility and efficiency of scientific research.

28

Learning to live with small/tiny effects

Published by Oxford University Press on behalf of the International Epidemiological Association
International Journal of Epidemiology 2011;40:1292–1307
doi:10.1093/ije/dyr099

Risk factors and interventions with statistically significant tiny effects

George CM Siontis[1] and John PA Ioannidis[1,2]*

[1]Clinical Trials and Evidence-Based Medicine Unit and the Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece and [2]Stanford Prevention Research Center, Department of Medicine, Stanford University School of Medicine, Stanford, USA

*Corresponding author. Stanford Prevention Research Center, Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA. E-mail: jioannid@stanford.edu

Accepted 19 May 2011

Background Large studies may identify postulated risk factors and interventions with very small effect sizes. We aimed to assess empirically a large number of statistically significant relative risks (RRs) of tiny magnitude and their interpretation by investigators.

Methods RRs in the range between 0.95 and 1.05 were identified in abstracts of articles of cohort studies; articles published in *NEJM*, *JAMA* or *Lancet*; and Cochrane reviews. For each eligible tiny effect and the respective study, we recorded information on study design, participants, risk factor/intervention, outcome, effect estimates, *P*-values and interpretation by study investigators. We also calculated the probability that each effect lies outside specific intervals around the null (RR interval 0.97–1.03, 0.95–1.05, 0.90–1.10).

Results We evaluated 51 eligible tiny effects (median sample size 112 786 for risk factors and 36 021 for interventions). Most (37/51) appeared in articles published in 2006–10. The effects pertained to nutrition ($n = 19$), genetic and other biomarkers ($n = 8$), correlates of health care ($n = 8$) and diverse other topics ($n = 16$) of clinical or public health importance and mostly referred to major clinical outcomes. A total of 15 of the 51 effects were >80% likely to lie outside the RR interval 0.97–1.03, but only 8 were >40% likely to lie outside the RR interval 0.95–1.05 and none was >1.7% likely to lie outside the RR interval 0.90–1.10. The authors discussed at least one concern for 23 effects (small magnitude $n = 19$, residual confounding $n = 11$, selection bias $n = 1$). No concerns were expressed for 28 effects.

Conclusions Statistically significant tiny effects for risk factors and interventions of clinical or public health importance become more common in the literature. Cautious interpretation is warranted, since most of these effects could be eliminated with even minimal biases and their importance is uncertain.

29

Concluding comments

- Nutrition science has made major contributions for understanding and treating clinical syndromes and extreme situations (e.g. deficiencies and toxic doses)
- Working in the range of subtle chronic disease associations has been notoriously frustrating and millions of papers have contributed mostly confusion
- A new paradigm is needed combining relevant randomized trials, reproducible research practices and rigorous exploration of precision options.

30