Shifting Sands

Unsound Science & Unsafe Regulation

REPORT 5

False Positives: The Irresponsibility Crisis of Science Policy

NATIONAL ASSOCIATION *of* SCHOLARS

S. Stanley Young
Warren Kindzierski
David Randall

Shifting Sands

Unsound Science and Unsafe Regulation

Report V

False Positives:
The Irresponsibility Crisis of Science Policy

A report by the

NATIONAL

ASSOCIATION

of SCHOLARS

Authors

S. Stanley Young
Shifting Sands Project Director
National Association of Scholars

Warren Kindzierski
Former Professor, School of Public Health
University of Alberta-Edmonton

David Randall
Director of Research
National Association of Scholars

About the National Association of Scholars

Mission

The National Association of Scholars is an independent membership association of academics and others working to sustain the tradition of reasoned scholarship and civil debate in America's colleges and universities. We uphold the standards of a liberal arts education that fosters intellectual freedom, searches for the truth, and promotes virtuous citizenship.

What We Do

We publish a quarterly journal, *Academic Questions*, which examines the intellectual controversies and the institutional challenges of contemporary higher education.

We publish studies of current higher education policy and practice with the aim of drawing attention to weaknesses and stimulating improvements.

Our website presents educated opinion and commentary on higher education, and archives our research reports for public access.

NAS engages in public advocacy to pass legislation to advance the cause of higher education reform. We file friend-of-the-court briefs in legal cases defending freedom of speech and conscience and the civil rights of educators and students. We give testimony before congressional and legislative committees and engage public support for worthy reforms.

NAS holds national and regional meetings that focus on important issues and public policy debates in higher education today.

Membership

NAS membership is open to all who share a commitment to its core principles of fostering intellectual freedom and academic excellence in American higher education. A large majority of our members are current and former faculty members. We also welcome graduate and undergraduate students, teachers, college administrators, and independent scholars, as well as non-academic citizens who care about the future of higher education.

NAS members receive a subscription to our journal *Academic Questions* and access to a network of people who share a commitment to academic freedom and excellence. We offer opportunities to influence key aspects of contemporary higher education.

Visit our website, www.nas.org, to learn more about NAS and to become a member.

Our Recent Publications

The Archimedes Standards: Model PreK-12 State Math Standards. 2025.

Ideological Insistence. 2025.

Waste Land: The Education Department's Profligacy, Mediocrity, and Radicalism. 2025.

Instagram the Intifada. 2024.

Shifting Sands: Zombie Psychology. 2024.

Shadows of Influence. 2024.

China and Our Children. 2024.

The Franklin Standards: Model K-12 State Science Standards. 2024.

The Illusion of Institutional Neutrality. 2024.

Curriculum of Liberty. 2024.

Contents

Executive Summary

The National Association of Scholars (NAS) published four reports under the *Shifting Sands: Keeping Count of Government Science* project: $PM_{2.5}$ *Regulation* (2021), *Food Frequency Questionnaires* (2022), *Confounded Errors of Public Health Policy Response* (2023), and *Zombie Psychology, Implicit Bias Theory, and the Implicit Association Test* (2024). These reports examined how irreproducible science affects select areas of government policy and regulation by different federal, state, and local agencies:

- The first report in 2021 focused on the field of environmental epidemiology, which informs the U.S. Environmental Protection Agency's (EPA) policies and regulations.
- The second report in 2022 focused on the field of nutritional epidemiology, which informs the U.S. Food and Drug Administration's (FDA) policies and regulations.
- The third report in 2023 focused on the country's public health bureaucrats' grave mishandling of the federal government's response to the COVID-19 pandemic. This included lapses by the Centers for Disease Control and Prevention (CDC) and the National Institutes of Health (NIH).
- The fourth report in 2024 focused on Implicit Bias theory, and the Implicit Association Test, which have informed a broad range of federal, state, and local statutes and regulations.

Government policies should be built on transparent and accountable research, typically claims made in papers. Regulations developed from these papers are meant to clear a high barrier of proof. The regulations should be based on reproducible scientific research. All the *Shifting Sands* reports proceeded by applying a straightforward statistical examination of scientific studies used to justify government policies.

One statistical approach involved estimating the number of hypotheses tested in the studies to assess the potential for Multiple Testing and Multiple Modeling (MTMM) problems. The other statistical approach—p-value plotting—was used as a severe test to assess the validity (and reproducibility) of research claims made in these studies.

The *Shifting Sands* reports found strong evidence that the irreproducibility crisis had affected all these bodies of research. This in turn had led to an *irresponsibility crisis of science policy* in these agencies. In essence, the EPA, the FDA, the CDC, and a host of states and localities all have used irreproducible science to justify irresponsible science policy that imposes illiberal and economically burdensome regulations or statutes on the American people.

These failures are so widespread that they constitute a crisis of government, and not just a crisis of science. We believe that the irresponsibility crisis of science policy must be addressed by four categories of reforms, which are explained further in the report:

1. Federal government agencies must make systematic changes to their regulatory and funding practices to remove the irreproducibility crisis from American science and the irresponsibility crisis from American government. These reforms will provide immediate solutions for the technical heart of the irreproducibility crisis and the irresponsibility crisis.

2. Federal and state policymakers must make systematic changes to American K-12 and undergraduate science and math education to educate properly a new generation of American scientific professionals and informed citizens and policymakers. These reforms are intended to improve the training of American scientists to reduce a future generation's tendency to engage in slipshod, politicized procedures that fuel the irreproducibility crisis.

3. Federal and state policymakers must end the arbitrary procedures of scientific research and scientific governmental regulation to preserve our liberty from arbitrary government. These reforms focus on aspects of the irresponsibility crisis that imperil Americans' liberty, law, and self-government.

4. Policy institutes must dedicate themselves to science policy as a first-order priority, staff their institutes with personnel dedicated to science policy, and make science policy a priority. These reforms are meant to provide the political infrastructure that will make the other reforms possible.

Policymakers and citizens also should aim at major reform of the structure of the American university, to the federal government's indirect cost funding formulas for university research, to prohibiting discrimination in the guise of illiberal programs and policies such as *diversity, equity, and inclusion*, and much more.

The NAS urges America's citizens, policymakers, and policy institutes to take up this challenge. Science and research procedures should be built on the solid rock of transparent, reproducible, and reproduced scientific inquiry, not on shifting sands. Likewise, science regulatory policy should be built on transparent and accountable procedures. Americans must dedicate themselves to the proper government of science policy, to assure that they retain self-government.

Introduction

America faces an *irresponsibility crisis of science policy*. Technocrats and radical activists embedded in government service have weaponized the powers delegated to federal science regulatory agencies, as well as the authority accorded to putatively nonpartisan scientific experts. They do so by some mixture of malice, self-righteousness, and self-deception. They advance their policy goals without transparency or accountability to elected policymakers or the public.

Radical activists and technocrats do not act solely via the realm of science policy. They seek to impose their policies by bureaucratic fiat in all aspects of government. But science policy is uniquely vulnerable to the ambitions of radical activists. Pervasive politicized groupthink and scientists' reliance on culpably negligent statistical procedures in their research have created the *irreproducibility crisis of modern science*,[1] which produces masses of false positive research results.

Activist bureaucrats actively commission these false positive research results in a host of scientific and social scientific disciplines to justify the mass production of illiberal, radical regulations throughout the range of federal science regulatory agencies, as well as to justify state laws and local ordinances. Radical researchers and radical bureaucrats work in tandem to subordinate constitutional government and democratic accountability to arbitrary, unfounded assertions of scientific authority. Scientific procedures ought to restrain arbitrary, ideological policymaking, but the corruptions of politicized groupthink and misused statistics instead facilitate it. American citizens, policymakers, and policy institutes must meet this challenge by means of an urgent program to institute comprehensive, coherent reform of American science policy.

The National Association of Scholars (NAS) came to this conclusion slowly. We began by publishing *The Irreproducibility Crisis of Modern Science: Causes, Consequences, and the Road to Reform* (2018), which focused upon the nature of the irreproducibility crisis.[2] We continued with our four-report project, *Shifting Sands: Keeping Count of Government Science* (2021-24), which examined how irreproducible science affects four select areas of government policy and regulation by different federal, state, and local agencies.

---

1   Randall (2018).
2   Randall (2018).

- *PM$_{2.5}$ Regulation* (2021) focused on irreproducible research in the field of environmental epidemiology, which informs the U.S. Environmental Protection Agency's (EPA) policies and regulations.[3]
- *Food Frequency Questionnaires* (2022) turned to nutritional epidemiology, which informs the U.S. Food and Drug Administration's (FDA) policies and regulations.[4]
- *Confounded Errors of Public Health Policy Response* (2023) analyzed public health policy response to the COVID-19 pandemic, which informs the policies of the U.S. Centers for Disease Control and Prevention (CDC) and the National Institutes of Health (NIH).[5]
- *Zombie Psychology, Implicit Bias Theory, and the Implicit Association Test* (2024) evaluated implicit bias research and the Implicit Attitude Test, which have informed a broad range of federal, state, and local statute and regulation.[6]

The *Shifting Sands* reports proceeded by applying a straightforward statistical examination of one or more meta-analyses. This process included counting the number of statistical hypotheses tests performed in representative papers from a meta-analysis and doing a p-value plot of the p-values from the papers used in the meta-analysis. The p-value plot is used as a visual check of the heterogeneity (dissimilarity) of test statistics (i.e., p-values) addressing the same research question. This counting and p-value plotting process is applied to research in these four subject areas that had been used to justify government policies. MTMM provides a simple way to assess whether any body of research has been corrupted by the irreproducibility crisis. The four *Shifting Sands* reports found strong evidence that the irreproducibility crisis had affected all these bodies of research. The EPA, the FDA, the CDC, and a host of states and localities all had imposed illiberal and economically burdensome regulations or statutes based upon insufficient scientific support.

These results matter tremendously in themselves. They also indicate a more deep-rooted problem. All these areas of *science policy*—policy informed and justified by scientific and social scientific research—had been distorted by different aspects of the irreproducibility crisis. Above all, *researcher degrees of freedom* had allowed false positive results to be taken for real scientific results and to set in motion vast governmental initiatives. Researcher degrees of freedoms supercharged the irreproducibility crisis in scientific research. *Intervention degrees of freedom* gave

---

3   Young (2021).
4   Young (2022a).
5   Young (2023).
6   Young (2024).

regulators the freedom to use easily manufactured false positive results to intervene in any aspect of the economy, society, or culture.

Americans should not trust that regulators achieved good ends by sloppy means. *The corollary to John Ioannidis' conclusion that more than half of recently published scientific research contains false conclusions[7] is that more than half of government policies based on recently published scientific research are unjustified.*

Activist regulators don't just use intervention degrees of freedom to impose economically burdensome regulations via federal agencies such as the Environmental Protection Agency (EPA) or the Food and Drug Administration (FDA). They also pose severe dangers to American liberty, American law—and, ultimately, to America's system of republican self-government.

Public health policy false positives, combined with the equally arbitrary use of mathematical modeling, contributed heavily to a massive and unnecessary lockdown of the country during the COVID-19 pandemic. The lockdown combined punishing economic damage with pervasive and severe abridgements of individual liberty at the federal, state, and local levels. Public health policy experts also decided that public health policy should include changing public attitudes toward public health policy. In other words, government bureaucrats sought to manufacture consent for their policies rather than to submit them to the judgment of the American people, and they labeled their technocratic propaganda as "public health policy." The bureaucrats combined authoritarianism with dysfunction. They engaged in this end-run around democratic accountability on behalf of scientifically unsubstantiated, mistaken, and practically disastrous policies.

Social psychology false positives in the field of implicit bias research, meanwhile, have justified "diversity training programs" that smack of re-education camps. Radical advocates have begun to propose initiatives justified by implicit bias theory to assault the legal bulwarks of liberty—individual responsibility and the presumption of innocence. They are working to impose implicit bias theory on the judiciary, juries, court personnel, lawyers, and police, so that every component of our judicial and law enforcement systems imposes a prejudgment of "bias" substantiated by nothing more than statistical false positives. They use the same theory to try to impose the arbitrary discrimination of "equity" in every aspect of the law.

Intervention degrees of freedom, in other words, means an unaccountable elite's freedom to impose its preferred policies without accountability to the Constitution, the law, America's elected representatives, or the voters. This elite of progressive activists and technocratic bureaucrats exploits the irreproducibility crisis by manufacturing false positive results to justify calling their policy goals "scientific."

---

7   Ioannidis (2005).

Scientists' refusal to submit their research to scientific tests of transparency and accountability accompanies bureaucrats' refusal to submit scientific policy to political tests of transparency and accountability.

Irreproducible science creates unaccountable government.

Careless statistical reasoning also aligns deeply with careless political reasoning. Scientists who seek statistical associations, shorn of evidence for direct causation, work easily with activists who champion policies based on statistical associations, shorn of respect for individual liberty and individual justice. The irreproducibility crisis erodes scientific standards and the policy crisis erodes standards of liberty. The professional-administrative elite, deeply marinated in progressive ideology, has fused these crises by its arrogation of power.

And this elite is progressive. *Groupthink* ("a psychological drive for consensus at any cost that suppresses dissent and appraisal of alternatives in cohesive decision making groups"[8]) is a central component of the irreproducibility crisis and it is progressive political groupthink that has seized control of the scientific world. No ideology is immune to political groupthink, but scientists *are* overwhelmingly progressive. The groupthink that deforms their scientific conduct therefore includes a predilection for government intervention, congenital mistrust of business, environmentalist ideology, radically egalitarian identity politics, and "equity."

Groupthink includes self-deception, so we cannot tell whether the radical professional-administrative elite consciously exploited the irreproducibility crisis to achieve its political aims. It does not matter in the end. Consciously or unconsciously, this ideological, irresponsible, and incompetent elite in the universities and the government now regularly initiates research programs whose sloppy procedures guarantee the false positive results that will justify more government policies that harm our prosperity and our liberty. The elite's motivations do not matter. Their actions endanger not only the integrity of scientific research but also our freedom. They must be stopped, even if they just act from an incompetence more dangerous than malice.

False Positives provides a policy-oriented conclusion to our *Shifting Sands* reports by outlining Policy Recommendations for how to address the irresponsibility crisis of science policy. First, however, we will summarize the substance of NAS's *Shifting Sands* reports—the nature of the irreproducibility crisis, the procedures we used to evaluate scientific research that informs government policy, the histories of the disciplines and agencies we have investigated, and the results of our investigations. The *Shifting Sands* reports present these results at far greater level of detail. Here we provide a digest of their substance to explain and to justify this report's conclusions and recommendations.

---

8   Janis (1982).

A radical elite's use of the spurious methods of irreproducible science to damage our prosperity and dismantle our liberty poses a grave challenge to Americans. We must pursue these reforms immediately and urgently if we wish to preserve our freedom and our welfare. We must free ourselves from arbitrary science to free ourselves from arbitrary government.

The Irreproducibility Crisis

T he validation of scientific truth requires *replication* or *reproduction*. *Replicability*, most applicable to the laboratory sciences, most commonly refers to obtaining an experiment's results in an independent study, by a different investigator with different data. *Reproducibility*, most applicable to the observational sciences, refers to different investigators using the same data, methods, and/or computer code to reach the same conclusion.[9] *Reproducibility* includes methods reproducibility, results reproducibility, and inferential reproducibility.[10] Scientific knowledge only accrues as multiple independent investigators replicate and reproduce one another's work.

Yet today the scientific process of replication and reproduction no longer works properly. A vast proportion of the scientific claims in published literature have not been replicated or reproduced. Researchers credibly estimate that a majority of these claims cannot be replicated or reproduced—that they are, in fact, false.[11] An extraordinary number of scientific and social-scientific disciplines no longer reliably produce true results—a state of affairs commonly referred to as the *irreproducibility crisis* (*reproducibility crisis, replication crisis*). A substantial majority of 1,500 active scientists recently surveyed by *Nature* called the current situation a crisis. 52% judged the situation a major crisis and another 38% judged it "only" a minor crisis.[12] The increasingly degraded ordinary procedures of modern science display the symptoms of catastrophic failure.[13]

The scientific world's dysfunctional professional incentives contributed substantially to this catastrophic failure. University researchers who publish exciting new results secure tenure, promotion, lateral moves to more prestigious universities, salary increases, grants, professional reputation, and public esteem. The same incentives affect journal editors, who receive acclaim for their journal, and personal reputational awards, by publishing exciting new research—even if they have not vetted that research thoroughly.[14] Grantors want to fund the same sort of exciting

---

9   NASEM (2016); NASEM (2019); Nosek (2020); Pellizzari (2017).
10   Goodman (2016).
11   Halsey (2015); Ioannidis (2005); Randall (2018).
12   Baker (2016).
13   Archer (2020); Chawla (2020); Coleman (2019); Engber (2017); Gobry (2016); Hennen (2019); Herold (2018); Ioannidis (2005); Manuel (2019); NASEM (2019); Randall (2018); Yong (2018); Young (2018); Zeeman (1976); Zimring (2019).
14   Ritchie (2020).

research—and government funders have the added incentive that exciting research with positive results also supports the expansion of their organizational mission.[15] American university administrations want to host grant-winning research, from which they profit by receiving "overhead" costs—frequently a majority of overall research grant costs.[16]

All these incentives reward *published research with new, positive claims*—but not *reproducible research*. Researchers, editors, grantors, bureaucrats, university administrations—each has an incentive to seek out the exciting new research that draws money, status, and power, but they have few or no incentives to double-check their work. Above all, they have little incentive to reproduce their research, to check that the exciting claim holds up—because if it does not, they will lose money, status, and prestige. Each member of the scientific research system, seeking to serve his own interest, acts in ways guaranteed to inflate the production of exciting, but *false*, research claims in peer-reviewed publications.

The scientific world's incentives to publish exciting research rather than reproducible research drastically affect which research scientists submit for publication. Scientists who try to build their careers on checking old findings or publishing negative results are unlikely to achieve professional success. The result is that scientists simply do not submit negative results for publication. Some negative results go to the file drawer. Others somehow turn into positive results as researchers, consciously or unconsciously, massage their data and their analyses. Neither do they perform or publish many replication studies, since the scientific world's incentives do not reward those activities either.[17]

We can quantify this skew by measuring *publication bias*—the skew in published research toward positive results compared with results present in the unpublished literature.[18] A body of scientific literature ought to have a large number of negative results, or results with mixed and inconclusive results. When we examine a given body of literature and find an overwhelmingly large number of positive results, especially when we check it against the unpublished literature and find a larger number of negative results, we have evidence that the discipline's professional literature is skewed to magnify positive effects, or even create them out of whole cloth.[19]

As far back as 1987, a study of the medical literature on clinical trials showed a publication bias toward positive results: "Of the 178 completed unpublished randomized controlled trials (RCTs)[20] with a trend specified, 26 (14%) favored the new

---

15   Lilienfeld (2017); Martino (2017).
16   Cordes (1998); Kaiser (2017); Roche (1994).
17   Randall (2018); Ritchie (2020).
18   Olson (2002); Nissen (2016); Randall (2018).
19   Chambers (2017); Harris (2017); Hubbard (2015); Ritchie (2020).
20   We use RCTs to refer both to "randomized controlled trials" and to "randomized clinical trials"; both terms are common in the literature, and they are roughly equivalent.

therapy compared to 423 of 767 (55%) published reports."[21] Later studies provide further evidence that the phenomenon affects an extraordinarily wide range of fields, including the social sciences generally;[22] climate science;[23] psychology;[24] sociology;[25] research on drug education;[26] and research on "mindfulness-based mental health interventions."[27]

Publication bias especially leads to a skew in favor of research that erroneously claims to have discovered a statistically significant relationship in its data. An extraordinary number of disciplines have shifted toward depending on statistical operations, and research that depends upon statistical operations is peculiarly subject to *p-hacking*.

P-hacking involves the relentless search for statistical significance and comes in many forms, including multiple testing and multiple modeling without appropriate statistical correction.[28] It enables researchers to find nominally statistically significant results even when there is no real effect; to convert a fluke, false positive into a "statistically significant" result.[29] Irreproducible research hypotheses produced by p-hacking send whole disciplines chasing down rabbit holes. For unscrupulous advocates and researchers who seek to game the system and secure a published result regardless of its truth value, p-hacking is a feature not a bug.

Political groupthink, as mentioned above, compounds the effects of scientific professional culture and the careless use of statistical procedures. Political groupthink, the professional culture of modern scientists, and the shift across an extraordinary number of disciplines towards statistical procedures jointly have created the irreproducibility crisis.

We can define the irreproducibility crisis as a professional culture deformed by political groupthink whose incentives and slipshod procedures allow the regular practice of p-hacking without check or consequence.

21   Dickersin (1987).
22   Franco (2014).
23   Michaels (2008).
24   Kühberger (2014).
25   Gerber (2008).
26   McCambridge (2007).
27   Coronado-Montoya (2016). Some meta-researchers prefer to regard the current state of affairs as an "irreproducibility challenge." Fanelli (2018).
28   Ellenberg (2014); Hubbard (2015); Chambers (2017); Harris (2017); Streiner (2018).
29   Boffetta (2008); Ioannidis (2011); McLaughlin (2013); Simonsohn (2014).

P-value Plotting: A Severe Test for Publication Bias, P-hacking, and HARKing

A standard form of p-hacking is for a researcher to run statistical analyses until a statistically significant result appears—and publish the one (likely spurious) result. When researchers ask hundreds of questions, when they are free to use any number of statistical models to analyze associations, it is all too easy to engage in this form of p-hacking. In general, research based on multiple analyses of large complex data sets is especially susceptible to p-hacking, since a researcher can easily produce a p-value < 0.05 by chance alone.[30] Research that relies on combining large numbers of questions and computing multiple models is known as Multiple Testing and Multiple Modeling.[31]

The authors of the *Shifting Sands* reports used *p-value plotting*, a visual form of Multiple Testing and Multiple Modeling (MTMM) analysis, as a way to demonstrate weaknesses in different fields and government agencies' use of *meta-analyses*.

Government agencies frequently rely on meta-analyses, statistical analyses that combine the results of multiple scientific studies. The formal meta-analysis process is strictly analytic. It computes an overall statistic for those test statistics combined, whereupon scientists make a research claim from the overall statistic. The meta-analysis computational method is flawed, given that "if there is some *garbage* in, then there is only *garbage* out."[32]

Revealing the presence of MTMM identifies a need for statistical corrections when large numbers of questions and models are used to analyze data sets. MTMM, also referred to multiplicity and multiple comparisons in literature, is a statistical analysis strategy that regularly produces large numbers of false positive *statistically significant* results. P-value plotting assesses heterogeneity (dissimilarity) of the test statistics combined in meta-analysis to infer whether "garbage" research is present. When applied to meta-analyses, MTMM analysis and p-value plotting allows

---

30   Chambers (2017); Glaeser (2006); Harris (2017); Hubbard (2015); Ritchie (2020); Westfall (1993).
31   Westfall (1993).
32   Nelson (2018).

researchers to detect the irreproducible meta-analyses produced from irreproducible base studies—*and, therefore, to provide strong evidence that government regulation or statute has been based on scientifically unfounded research*.

P-value plotting can determine by a visual presentation whether a body of literature is really revealing a null result for a statistical relation between a cause and an outcome, which has been cherry picked to justify government regulation or statute. P-value plotting also can detect heterogenous p-values—a body of research that has been distorted by questionable research practices such as publication bias (evidence of "missing papers" in a body of research which would have presented negative evidence), p-hacking, and/or HARKing (to *hypothesize after the results are known*—to look at the data first and then come up with a hypothesis that has a statistically significant result). P-value plotting provides evidence of corrupted base studies used in a meta-analysis, which therefore have rendered the meta-analysis unreliable.

P-value plotting is not the only means available by which to detect questionable research procedures. Scientists have come up with a broad variety of statistical tests to account for frailties in base studies as they compute meta-analyses. Unfortunately, questionable research procedures in base studies severely degrade the utility of the existing means of detection.[33] The *Shifting Sands* authors used p-value plotting not as the first means to detect HARKing and p-hacking in meta-analyses, but as a better means than alternatives which have proven ineffective.[34]

Scientists generally are at least theoretically aware of the dangers posed by publication bias, p-hacking, and HARKing, but they have done far too little to correct their professional practices. Stroup et al., for example, provided a proposal for reporting meta-analysis of observational studies in epidemiology.[35] This proposal is frequently referred to in published literature—16,676 Google Scholar citations as of November 5, 2021.[36] Yet Stroup et al. made no mention of observational studies' MTMM problem and offered no recommendation to control for MTMM. Epidemiologists, for example, are usually silent about MTMM, but when they do address the subject, they often are adamant that no correction for MTMM is necessary.[37] To our knowledge, no epidemiological article, institutional statement, or government regulation has prescribed a MTMM correction for observational studies or directed meta-analysis researchers to account for MTMM bias.

---

33   Carter (2019).
34   For a longer explanation of Multiple Testing Multiple Modeling and of statistical significance, see Young (2022a), Appendix 1 and Appendix 2.
35   Stroup (2000).
36   GS (2021).
37   Rothman (1990).

Methods to adjust for MTMM have existed for decades. The Bonferroni method simply adjusted the p-value by multiplying the p-value by the number of tests. Westfall and Young provided a simulation-based method for correcting an analysis for MTMM.[38] The *Shifting Sands* authors applied this method to disciplines that have persistently failed to incorporate these known methods of correction into their standard operating procedures.

---

38   Benjamini (1995); Westfall (1993).

Four Distorted Disciplines

T he four *Shifting Sands* reports investigated *environmental epidemiology*, *nutritional epidemiology*, *epidemiological public health policy*, and *psychology's implicit bias research*, and the government entities that have based their policies on these bodies of research. These four disciplines vary broadly in their subject matter and their procedures, but all partake in the statistical revolution. All four disciplines, in other words, now establish *statistical associations* to "prove" a sufficient justification for government action.

Deeper philosophical and methodological critiques pose fundamental objections to the substitution of statistical association for causal mechanisms. The *Shifting Sands* reports, however, focused on the insufficient controls of this practice. Scientists theoretically know that they should apply rigorous controls on statistical tests, but in practice researchers in all these disciplines have set up default procedures that facilitate the regular manufacture of false positive results that justify assertive government policy.

Here we outline for each of these four disciplines a sketch of how it became dominated by statistical methodology, how these methodologies were adopted by government agencies, some of the consequences of governmental policy, and the results of our technical studies into select meta-analyses used to substantiate specific government policies. These outlines summarize the four *Shifting Sands* reports, and the published technical studies that formed the basis for the *Shifting Sands* reports.

We provide these outlines to orient readers about the history and the nature of these disciplines, and to suggest the scope and the stakes of government policy in these areas. We also provide these outlines to provide a sufficient justification for the reforms we recommend at the conclusion of this report. We have not investigated every scientific and social scientific discipline that informs government policy, nor every one of extraordinarily large number of regulations and laws based on scientific and social scientific research. We believe we have investigated a representative and significant sample. We provide these four narratives of environmental epidemiology, nutritional epidemiology, epidemiological public health policy, and psychology's implicit bias research, and our four specific technical investigations, to substantiate our general policy recommendations.

Environmental Epidemiology: PM$_{2.5}$ Regulation

A series of local and state regulatory initiatives, particularly in Los Angeles and California, led to the establishment in 1963 of a federal regulatory structure for environmental protection—The Clean Air Act. This was followed by further federal measures, notably the Motor Vehicle Control Act (1965), the Clean Air Act Amendments (1966), the Air Quality Act (1967), the Clean Air Act Amendments (1970), and the establishment of the Environmental Protection Agency (1970). These collectively, but particularly the last two, set the ground rules for the regulatory structure that has persisted to the present day.[39]

The Environmental Protection Agency (EPA) must develop air quality criteria for specific airborne components, informed by expert opinion, and describe their effects singly and in combination on the health and welfare of American citizens. The EPA must set National Ambient Air Quality Standards (NAAQS), as a yardstick by which states and localities can measure their own air quality, and as a legal requirement to enforce reduction of airborne components. The EPA coordinates with different federal government departments, such as the Office of Management and Budget (OMB) and the Council on Environmental Quality (CEQ), but it plays the leading role.[40]

The EPA imposed increasingly restrictive regulations and regularly updated NAAQS. These required data accumulation on both air quality and on health effects, forwarded by the EPA's sponsorship of research that would underpin the emerging regulations. The EPA only shifted from regulation of Total Suspended Particles (TSP) to PM$_{10}$ in 1987. It did not regulate PM$_{2.5}$ (particulate matter less than 2.5 microns in width) explicitly until 1997. The EPA is far older than its current particulate matter regulatory regime.[41]

The current regulations depend on statistical analysis. The EPA and environmental epidemiologists, as a discipline, have not established direct causal biological mechanisms that link air components and health outcomes[42]—save for freak conditions such as prevailed in the Meuse Valley (1930), Donora (1948), and London (1952).[43] Rather, they have relied on statistical analyses to discern *significant associations* between air components and health outcomes. These associations provide the "proof" that an air component, alone or in association with other elements, causes damage to health and to the economy. The debate about whether or not the EPA should make a particular regulatory decision raises questions central to the irreproducibility

---

39   Bachmann (2007).
40   Bachmann (2007).
41   Bachmann (2007); Cao (2013).
42   CASAC (2019); Cox (2017).
43   Milloy (2016); Nemery (2001).

crisis—data accuracy, research protocols, statistical analyses, publication bias, sponsorship bias, etc.

Environmental epidemiological researchers regularly engage in massive hypothesis tests without making Multiple Testing and Multiple Modeling (MTMM) statistical corrections. These tests have associated air quality components with a remarkable number of possible adverse health effects.

These possible, but not proven, effects include but are not limited to: all-cause mortality; cause-specific mortality; all-cause morbidity; low birth weight; miscarriage; COPD exacerbation; inflammation; pulmonary complication; autism; obesity; depression; atopic dermatitis; impaired vestibular function (sense of balance); metabolic disorders; suicide, mental health and well-being; ADHD (Attention Deficit/Hyperactivity Disorder); respiratory complication; pneumonia and acute respiratory infection; reproductive outcomes; high blood pressure; lung and other cancers; and accelerated brain aging.[44]

The costs of insufficiently substantiated regulation can become exorbitant. As a recent example, consider estimated costs requiring ships to use "cleaner fuel" with less sulfur, so as to reduce SO_2 emissions.[45] The EPA argues that the move to low-sulfur ship fuel could save up to 14,000 American and Canadian lives every year. The inferred health-related benefits are estimated to be as much as US$110 billion/year in 2020. The EU claims these regulations will prevent 50,000 premature deaths. On the other hand, the cost of these regulations is estimated at US$3.2 billion/year in 2020 and may rise to a total of one trillion dollars through the year 2050.[46] Yet a growing body of research fails to support the EPA and the EU's mortality claims.[47] This research provides evidence that SO_2 in ambient air has no significant association with mortality,[48] heart attacks,[49] asthma,[50] or lung cancer.[51] Americans may be paying up to $1 trillion dollars to satisfy a regulation with no real scientific foundation.

The EPA issues an extraordinary number of regulations, which affect every area of the economy and constrict everyday freedoms. If the long-term cost of one regulation on one industry amounts to one trillion dollars, the cost of many regulations on every industry is uncountable trillions. The EPA should only impose such costly regulations using fully reproducible science that has survived a battery of severe tests.

---

44   Samet (2019).
45   Cuff (2016); IMO (2020); Tapaninen (2020).
46   Paris (2020).
47   Young (2017).
48   Milojevic (2014); Young (2017).
49   Young (2019a).
50   Kindzierski (2021); Young (2019b).
51   Acharjee (2017); Young (2020).

We note here a conundrum. By an extraordinary number of indicators, Americans' general health has risen remarkably over the last several generations.[52] The gravest recent harm to Americans' life expectancy has been the opioid epidemic, concentrated among poor Americans—an effect entirely unrelated to the remit of the EPA.[53] Yet the EPA produces an ever-lengthening catalogue of studies of things that harm Americans' health.[54] Feinstein charged as far back as 1988 that much of the research suggesting specific health-harms must be the result of misuse of statistics and computers, data dredging to produce dire literature of statistically significant effects that square badly with evidence of general improvements in health and life expectancy.[55]

More narrowly, the EPA constructed its $PM_{2.5}$ regulation from 1997 to the present day upon a series of studies in the generation from the 1970s to the 1990s that sought to establish: 1) significant associations between $PM_{2.5}$ and various health effects; and 2) that the health effects were themselves substantial enough to justify EPA regulation.[56] The regulation from 1997 onward relied on research drawn from the famous Harvard Six Cities and American Cancer Society (ACS) studies—whose original data, on claimed grounds of privacy and confidentiality, have never been made transparently available to other researchers for reproduction or critique.[57]

We may note here that data for environmental epidemiology is difficult to collect—it is an observational science rather than a laboratory one, and one that requires data sets of hundreds of thousands of individuals, sometimes collected over decades, to make any sort of definitive statement. The EPA delayed more rigorous regulation of $PM_{2.5}$ for a generation precisely so as to assemble a data set that they thought would justify such regulation. The EPA and its advocates argue that the difficulty of collecting such data justifies allowing the EPA to base regulation on inaccessible data.

However, it is precisely because the data are so difficult to collect that it is vital to have access to the one available data set, so that it may be subjected to a battery of rigorous tests to see if the analysis is sound. The burden of proof for transparency and reproducibility lies with the research the EPA uses as the basis for its regulations.

When EPA regulation is based on inaccessible data, there are numerous potential weaknesses. We cannot fully account for interaction effects—the effects of "confounding" variables on health effects, such as temperature,[58] atmospheric

---

52   Woolf (2019).
53   Gold (2020).
54   E.g., EPA (2011).
55   Feinstein (1988).
56   Cao (2013).
57   Dockery (1993); Pope (1995); and note especially the critique in Enstrom (2017).
58   Cox (2012).

inversion, or varying demographic predispositions to sickness and mortality.[59] We cannot examine the base information itself for reliability. Death certificates, for example, are not entirely reliable sources of information.[60] Neither do we possess the data that can begin to allow us to determine what are the precise causal mechanisms—the biological mechanisms—by which an airborne component actually induces a health risk.[61]

The Harvard Six Cities/ACS data that underpin the Dockery/Pope research on air quality causing premature deaths (mortality) has never been subjected to Multiple Testing and Multiple Modeling (MTMM), even though an adjustment for MTMM with the widely-used SAS statistical software could easily be applied to the data.[62] Since Dockery and Pope never made their air quality–death data publicly available, no independent, critical researcher can subject the Harvard Six Cities/ACS data to the severe test of MTMM. Since analysis of newer and much larger data sets has found no effects of air quality on mortality, skepticism about the Dockery/Pope results is warranted.[63]

Researchers would prefer to analyze the EPA's $PM_{2.5}$ policy by scrutinizing the data underlying the Harvard Six Cities/ACS studies. Unfortunately, the data's owners have barred public access on the claimed grounds of privacy and confidentiality. In our first *Shifting Sands* report, therefore, the authors applied the MTMM method to portions of the meta-analyses of research underlying the Environmental Protection Agency's (EPA) $PM_{2.5}$ regulations—the regulations based upon research affirming that particulate matter smaller than 2.5 microns in diameter has a deleterious effect on human health. The authors found that there was indeed strong evidence that these meta-analyses had been affected by publication bias, p-hacking, and/or HARKing.

In our first *Shifting Sands* report, the authors conducted four technical investigations about associations between fine particulate matter ($PM_{2.5}$) (and in some cases other air quality components) in ambient air with various health effects. These effects include all-cause mortality, heart attacks and two asthma effects— development of asthma and asthma attacks. They approached these investigations by focusing on meta-analyses that ask the specific question whether *inferred exposure to $PM_{2.5}$ (and other air quality components) is associated with increases in all-cause mortality, heart attacks and asthma*. P-value plots for all four technical studies revealed heterogenous sets of p-values—in visual terms, bilinear patterns.

---

59   CASAC (2019).
60   Goldacre (1993).
61   CASAC (2019). For more discussion about current status of unanswered $PM_{2.5}$-mortality causal mechanisms and several negative studies that invalidate $PM_{2.5}$-mortality causation, see Young (2021), Appendix 4: PM2.5–Mortality Causality—Incomplete Evidence. This evidence, drawn from published literature, does not support a $PM_{2.5}$-mortality causal mechanism.
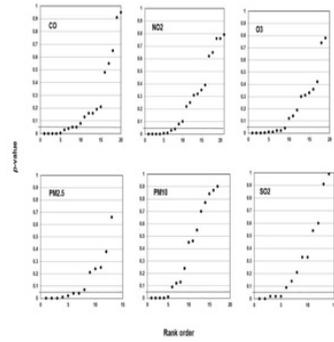62   Westfall (1993).
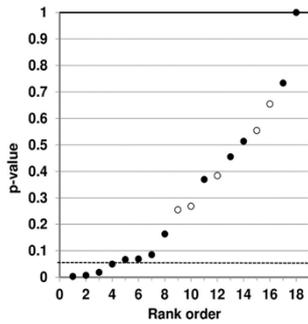63   Greven (2011); Milojevic (2014); Young (2017).
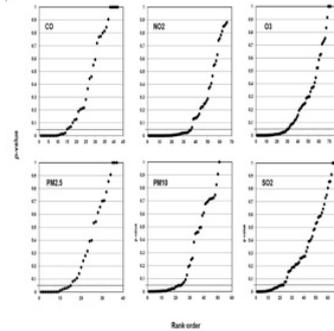
P-value plot, All-Cause Mortality and PM2.5[64]
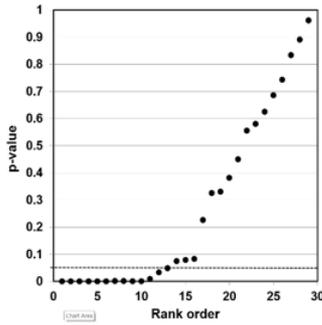
P-value plots, Six air quality components, Air quality—heart attack meta-analysis

18 Cohort Studies, P-Value Plot[65]

P-value plots, Six air quality components, Air quality—asthma attack meta-analysis

Note: solid circles (●) are NO_2 p-values; open circles (o) are $PM_{2.5}$ p-values).

These results provide strong statistical evidence that the EPA has developed policy and regulated $PM_{2.5}$ based upon a field of epidemiology research substantially affected by some combination of sampling bias, publication bias, p-hacking and/or HARKing.

The EPA, in other words, may have imposed trillions of dollars of costs on the American economy based upon irreproducible science.

Nutritional Epidemiology: Red and Processed Meats and Soy Protein

After World War II, and particularly in the aftermath of the Thalidomide scandal of the later 1950s and early 1960s, the United States Food and Drug Administration's (FDA) mandate to enforce drug safety prompted it to adopt rigorous requirements

---

64   Orellano (2020), Appendix A, Figure A.5.
65   Anderson (2013); Young (2022b)

for study design, centered upon the gold standard of the *randomized clinical trial*, and equally rigorous requirements for *statistical analyses* of the data. It adopted these techniques to fulfil the somewhat vague legislative mandate to assess "substantial evidence" by means of "adequate and well-controlled studies." The FDA chose these techniques partly for their technical efficacy and partly because they would pass judicial muster when private manufacturers submitted legal challenges to the scientific validity of FDA regulations.[66]

Since the 1960s, the FDA has reviewed its study design and statistical analysis standards regularly. In collaboration with private industry and academic researchers, it has updated them to match the evolving best practices of scientific research.[67]

The FDA now requires that foods (except for meat from livestock, poultry and some egg products, which are regulated by the U.S. Department of Agriculture) be safe, wholesome, sanitary, and properly labeled.[68] The FDA's labeling requirements include mention of *health claims* that characterize the relationship between a substance (e.g., a food or food component) and a health benefit, a disease (e.g., cancer or cardiovascular disease), or a health condition (e.g., high blood pressure).[69] These *health claims*, whether for good or for ill, must survive an FDA assessment based on rigorous study design and valid statistical analysis. The FDA articulates its regulatory requirements by means of a lengthy catalog of highly detailed Guidance Documents.[70]

FDA regulatory requirements necessitate that evidence to support a health claim should be based on studies in humans.[71] The randomized controlled trial (RCT), especially the randomized, placebo-controlled, double-blind intervention study, provides the strongest evidence among studies in humans.[72] The best RCT certainly trumps the best observational study—and one might argue that a very indifferent RCT is still superior to the best observational study.[73] Yet not all intervention studies on food and food components are RCTs, and frequently an RCT is unavailable and/or impractical. In these cases, the FDA must rely on lower-quality observational studies. It relies especially on cohort studies, dependent on dietary assessments based on FFQ analyses, and now pervasive in nutritional epidemiology.[74]

66   Junod (2008).
67   Junod (2008).
68   U.S. FDA (2021).
69   Kavanaugh (2007).
70   E.g., U.S. Food and Drug Administration Guidance Documents (2006); U.S. Food and Drug Administration Guidance Documents (2009).
71   U.S. Food and Drug Administration Guidance Documents (2009).
72   Schneeman (2007).
73   Barton (2000). RCTs do not as yet standardly account for the latest research, which is broadening our knowledge of the substantial individual and group variation in response to nutritional substances. Cecil and Barton (2020).
74   Byers (1999); Freudenheim (1999); Prentice (2010); Sempos (1999).

The discipline of *nutritional epidemiology* plays a vital role in the FDA's labeling requirements. The FDA uses nutritional epidemiology to provide compelling scientific information to support its nutrition recommendations and more coercive regulations.[75] Nutritional epidemiology applies epidemiological methods to the study at the population level of the effect of diet on health and disease in humans. Nutritional epidemiologists base most of their inferences about the role of diet (i.e., foods and nutrients) in causing or preventing chronic diseases on observational studies.

From the 1980s onward, the increase of computing capabilities facilitated the application of essentially retrospective self-administered dietary assessment instruments—the semi-quantitative food frequency questionnaire (FFQ).[76] FFQs, which are easy to use, place low burdens on participants, and allegedly capture long-term dietary intake, have become the most common method by which scientists measure dietary intake in large observational study populations.[77]

Nutritional epidemiology suffers many weaknesses. Critics have long noted that nutritional epidemiology relies predominantly on observational data, which researchers generally judge to be less reliable than experimental data, and that this generally weakens its ability to establish causality.[78] The discipline's research findings are also afflicted by frequent alterations of study design, data acquisition methods, statistical analysis techniques, and reporting of results.[79] Selective reporting proliferates in published observational studies; researchers routinely test many questions and models during a study, and then only report results that are interesting (i.e., statistically significant).[80]

FDA Guidance Documents acknowledge the shortcomings of food consumption surveys, including FFQs, and generally note that observational studies are less reliable than intervention studies—but still allow FFQs to inform FDA regulations.[81] And even published assessments of shortcomings in nutritional epidemiology procedures[82] usually overlook the problems posed by multiple analysis.

The FDA does acknowledge some dangers from multiplicity analysis, notably in its *Multiple Endpoints in Clinical Trials Guidance for Industry*.[83] Yet nutritional epidemiology suffers from the type of flawed statistical analysis that predictably and chronically inflates claims of statistical significance by failing to adjust for MTMM

---

75   Kavanaugh (2007).
76   Boeing (2013).
77   Boeing (2013); Satija (2015).
78   Satija (2015). Causal criteria in nutritional epidemiology include consistency, strength of association, dose response, plausibility, and temporality. Potischman (1999).
79   Boffetta (2008); NASEM (2016); NASEM (2019); Randall (2018); Sarewitz (2012).
80   Gotzsche (2006); also see Byers (1999).
81   E.g., U.S. Food and Drug Administration Guidance Documents (2006); U.S. Food and Drug Administration Guidance Documents (2009).
82   E.g., Liu (1994); Kristal (2005); Shim (2014).
83   U.S. Food and Drug Administration Guidance Documents (2017).

and by allowing researchers to search for results that are "statistically significant". Scientists have made these points repeatedly in professional and popular venues.[84]

The FDA's health claim reviews examine factors including whether studies are controlled for bias and confounding variables, appropriateness of a study population, soundness of the experimental design and analysis, use of appropriate statistical analysis, and estimates of intake.[85] *These "reviews" do not address the MTMM problem. Nor do they compare the given analysis to a protocol analysis.*

Nutritional epidemiologists have done far too little to correct their professional practices.[86] Scientists also have warned their peers about the particular dangers of multiple testing of cohort studies.[87]

The stakes of these shortcomings are substantial. Inaccurate labels can mislead consumers, not least by encouraging them to adopt fad diets that present health risks.[88] Furthermore, every company in the food sector, which involved $6.22 trillion dollars in annual sales in 2020,[89] depends for its livelihood on accurate labeling of food products. Mislabeling health benefits can give a company a larger market share than it deserves.

To take a more concrete example, the Code of Federal Regulations declares that "The scientific evidence establishes that diets high in saturated fat and cholesterol are associated with increased levels of blood total- and LDL-cholesterol and, thus, with increased risk of coronary heart disease," and allows companies to make corollary health claims about reducing the risk of heart disease.[90] The FDA duly notes on its Interactive Nutrition Facts Label that "Diets higher in saturated fat are associated with an increased risk of developing cardiovascular disease."[91]

Yet recent research concludes that "Numerous meta-analyses and systematic reviews of both the historical and current literature reveals that the saturated-fat diet-heart hypothesis was not, and still is not, supported by the evidence. There appears to be no consistent benefit to all-cause or CVD mortality from the reduction of dietary saturated fat."[92] The law rather than the FDA's approach to statistics was at issue here, but the financial consequences have been enormous: consumers have redirected billions of dollars toward producers of foods with less saturated fats, for a diet that may have no discernible health benefit.[93]

---

84   See Byrnes (2001); Støvring (2007); Gotzsche (2006); Gullberg (2009); Kmietowicz (2014). For the Brian Wansink scandal, see Hamblin (2018); Randall and Welser (2018).
85   Schneeman (2007).
86   Aschwanden (2016); Chambers (2017); Harris (2017); Head (2015); Hubbard (2015); Ruxton (2016); Schoenfeld (2013).
87   Bolland (2014).
88   D'Souza (2020); Marks (2011); Schutz (2021).
89   Blázquez (2021).
90   CFR (2020).
91   INFL (n.d.).
92   Gershuni (2018).
93   And see Peretti (2013).

In our second *Shifting Sands* report, the authors conducted two technical investigations about associations between red and processed meats and negative health outcomes, and soy protein and cardiovascular disease risk reduction, and found persuasive circumstantial evidence that the scientific literature (in general) and statistical practices (specifically) had been distorted. The flawed statistical practices centered around the use of the semi-quantitative Food Frequency Questionnaire (FFQ).

The authors evaluated one meta-analysis of red meat and processed meat that used observational base studies[94] and one meta-analysis of soy protein that used RCT base studies[95] as representative of nutritional epidemiologic work in this area. Most nutritional epidemiologists now believe that red and processed meat are associated with severe health effects.[96] The International Agency for Research on Cancer (IARC), the cancer research agency of the World Health Organization, has classified red meat as probably carcinogenic to humans and processed meat as certainly carcinogenic to humans.[97] Some researchers, however, have challenged the nutritional epidemiologists' consensus on other grounds. For example, Vernooij et al. argue the base observational studies are unreliable.[98] The popular press, rather than deferring to a professional consensus, also has pushed back against this paradigm—not least by citing popular low-carbohydrate and high-meat diets (Atkins, etc.) that do not seem to have imposed ill effects on their practitioners.[99] The nutritional epidemiologists' consensus on the carcinogenic effects of red and processed meats does not possess full authority with either professionals or the public.

The Johnston research group (Vernooij et al.) has provided some of the strongest arguments to date against the nutritional epidemiologists' consensus. Their large-scale systematic review and meta-analysis of the 105 base study papers studying the health effects of red and processed meats has provided strong evidence that the base study papers, generally observational studies, provided low- or very-low-certainty evidence[100] according to GRADE criteria.[101]

---

94   Vernooij (2019).
95   Blanco Mejia (2018).
96   For example, see Battaglia (2015); Ekmekcioglu (2018).
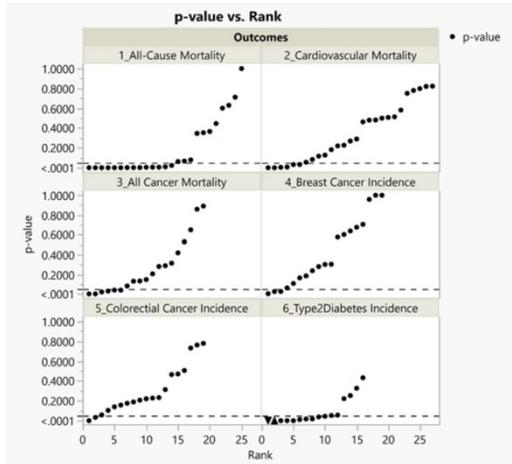97   WHO (2015).
98   Delgado (2021).
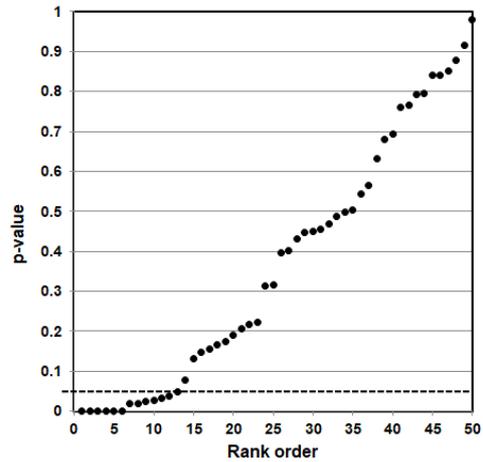99   Bueno (2013); Castellana (2021); Taubes (2021).
100 Vernooij (2019).
101 Guyatt (2008).

P-value plots for meta-analysis of six health outcomes from Vernooij et al.

P-value plot for meta-analysis of the association between soy protein intake and LDL cholesterol reduction from Blanco Mejia et al.

P-value plots for meta-analyses of health effects of red and processed revealed heterogenous sets of p-values—in visual terms, bilinear patterns.

P-value plots for meta-analyses of health effects of soy protein also revealed heterogenous sets of p-values.

These results provide strong statistical evidence that the field of nutritional epidemiology research has been substantially affected by some combination of sampling bias, publication bias, p-hacking and/or HARKing.

We wish to emphasize here that these results partly argued against FDA policy and partly supported it—these results *supported* the FDA's preliminary decision to revoke the 1999 health claim that links soy protein to heart health. The *Shifting Sands* procedures will not disprove the grounds for all existing scientifically informed regulation; they will provide firm support for some portion of our existing regulatory structure. Our procedures, and the policy reforms we suggest, will reform and improve our existing regulatory structure, not just overturn it.

We also wish to emphasize that these results also provided suggestive evidence of research integrity violations. The p-value plot of the randomly selected base studies of the soy protein study also produced a bilinear pattern, even though they were RCTs. The authors discovered 13 p-values below 0.05 supporting an effect and 37 above, supporting no effect. One of these small p-values—0.02037—is for an 'increase' (instead of decrease) in LDL cholesterol. The usual attention to control of MTMM of RCTs as compared with observational studies renders it less likely than that such a bilinear pattern could have emerged from randomness or negligence. The authors' methods and conclusions cannot by themselves prove individual or

systematic research integrity violations, but they do provide circumstantial evidence of widespread research integrity violations in the scientific community.

We wish to have confidence in the good faith of all scientific researchers. This research result, however, leads us to make general policy recommendations to improve research integrity procedures. While most scientific researchers may be acting in good faith, the misconduct of a few is sufficiently important to require changes in the practices of all scientists.

Public Health Policy: COVID-19: Masks and Lockdowns

The Centers for Disease Control and Prevention's (CDC) shift of focus from communicable diseases to environmental health depended upon a parallel shift towards statistical methodology. While the CDC used elementary statistics from its inception, it began to incorporate statistics far more intensively from the late 1950s. The CDC's shift toward statistics relied upon the development of computer hardware and software. Statistics done by hand is intensely time-consuming and the CDC, as every other private and public entity, could not use statistics intensively until computers automated statistical calculation. The successive adoption of programmable calculators, mainframe computers, punch-card technology, and microcomputers each vastly facilitated the CDC's use of statistics. In 2024, the CDC's emphasis on intensive statistical computation using sophisticated computer hardware and software was not yet forty years old.[102]

The CDC's adoption of ever more sophisticated statistical methods has accompanied, and driven, an increase in its technical capacities and a transformation of the way it conceives of disease. Its basic shift in focus from communicable diseases to environmental health ultimately has depended upon the adoption of a statistical framework, which seeks statistical associations between predictors and medical outcomes rather than biological mechanisms. So, for example, "the use of NHANES data combined with data on lead in gasoline from the U.S. Environmental Protection Agency [made it possible] to develop a model to predict human blood lead levels." Further statistical methodological developments that allowed the CDC to operate with increased sophistication include logistic regression models, back-calculation methods, time series analysis, a general integration of surveillance data and epidemic detection and control, work on detection of statistical aberrations, changes in patterns of data over time, cluster investigations, statistics for rare events and small areas, statistics for public health decisions, complicated designs and data

---

102 Bauer (2017); Stroup (2011).

structures, methods for decisions in uncertainty, use of multi-source data, mapping prevalence data, special analysis—and, above all and generally, mathematical modeling.[103]

Mathematical modeling for infectious disease epidemiology possesses its own long history, tracing back to John Graunt's compilation of mortality data in *Natural and Political Observations made upon the Bills of Mortality* (1662), Daniel Bernoulli's 1766 analysis of the effects of smallpox variation, and Ronald Ross' use of mathematical models both to analyze malaria transmission and to evaluate the effectiveness of malaria prevention methods. The latest generations of epidemiological modeling include innovations such as sophisticated use of compartmental models, which distinguish between population subgroups, the use of partial differential equations, and network models.[104] In the late twentieth century, the CDC and the field of epidemiological statistics as a whole began to merge their efforts with mathematical modeling—to supplement, to hybridize with, and even to replace, mechanical models of disease transmission with statistical models. Indeed, statistical methods became necessary simply to determine how to assess samples of data drawn from endless well of Big Data. Epidemiological mathematical modeling became heavily dependent upon statistical models.[105]

This shift in method brought with it a shift in mission. Mathematical modeling, even more than infectious disease epidemiology, was concerned with *public health interventions*—not simply to learn how disease was transmitted, but to reduce its incidence. Ross's work on malaria, we should underline, used mathematical models both to understand how malaria was transmitted and to make sure it could be eradicated. Mathematical epidemiology was always a tool of the state as it sought to improve public health—and we may note that the health of the public thus always implied the health of the state.

But the combination of mathematical modeling, statistics, and environmental health necessarily brought with it an enormous increase of scope for such public health measures. The target was no longer a mosquito or a bacterium, *but every aspect of the environment, including individual and collective human behavior, that possessed a statistical association with a medical outcome*. The discipline in consequence became interdisciplinary, as it drew in data from fields such as microbiology, the social sciences, and the clinical sciences.[106]

Infectious disease epidemiology, statistical methodology, and mathematical modeling transformed the mission of public health. It was no longer content to

---

103 Stroup (2011).
104 Brauer (2017); Kretzschmar (2009); Stroup (2011).
105 Kretzschmar (2009); Stroup (2011).
106 Kretzschmar (2009).

assess every individual and collective human behavior that possessed a statistical association with a medical outcome. It now sought to alter them, and to engage in real-time analysis of these strategies so as to improve their efficacy. Sieber cautions that, "Use of multisource data and further development of record linkage techniques to extract maximal information from existing data sources also will require addressing privacy and confidentiality concerns, as well as appropriate methods of communication of important public health findings to the nation." This caution seems remarkably understated.[107]

Public health has been pioneering the methodologies of the surveillance state—and, since every individual and collective behavior conceivably can influence health, the scope of its methodologies has covered a remarkably broad range of behavior. Its ambitions should provoke concern among Americans who wish to preserve individual liberties from the government.

The COVID-19 epidemic has made the limitations of such public health models crystal clear. The CDC and other epidemiologists used mathematical modeling throughout to estimate transmission, risks, and the effects of different public health interventions.[108] Neil Ferguson's first COVID-19 model proved spectacularly misguided—and spectacularly influential, not least from the nightmare scenario it painted of COVID-19 response absent social distancing: "At one point, the [Ferguson] model projected over 2 million U.S. deaths by October 2020." But even though models are supposed to be evaluated by their usefulness, scientists' enthusiasm for Ferguson's model was not dampened by its failure: "This model proved valuable not by showing us what *is* going to happen, but what *might have been*."[109] Even this encomium would appear to be misguided, since Ferguson's model also predicted a nightmarishly high level of deaths, even with full lockdown policies enacted.

More precisely, Ferguson's model failure, and the failures of other COVID-19 models, did not dampen enthusiasm among a large *part* of the professional community of epidemiological statisticians and modelers.[110] This part of the professional community, which dominates the CDC and peer institutions, takes *model failure* to be a *temporary shortcoming*, data to be used to improve the next generation of models. Such professionals make carefully delimited suggestions for methodological reform: "It has been observed previously for other infectious diseases that an ensemble of forecasts from multiple models perform better than any individual contributing model."[111] They note the rationales for models whose simplicity led to profound policy errors, e.g., that modelers frequently prefer simple, *parsimonious* models,

---

107 Sieber (2006).
108 Adiga (2020); Biggerstaff (2022); Brauer (2017); Ferguson (2006); GAO (2020); Kretzschmar (2009).
109 Adiga (2020); GAO (2020).
110 E.g., Colbourn (2020); Pachetti (2020); Prem (2020); Verity (2020); Walker (2020).
111 Adiga (2020).

particularly to allow policy interventions to proceed quickly.[112] Their retrospective on the history of COVID-19 modeling is one of bland, technocratic success.[113]

The policies that these researchers so blandly endorsed, meanwhile, were astonishingly and troublingly open-ended. In April 2020, for example, the WHO recommended that governments continue lockdowns until such time as they could achieve a set of six conditions alternately arbitrary or implausibly rigorous.[114] These conditions seemed to imply that governments should continue lockdowns until such time as the citizenry's "fully educated" views and behavior coincided in all respects with the recommendations of public health experts. A technical model submitted to the public for judgment should not have the alteration of the public's judgment as a component—much less hold the public hostage to continued lockdowns until they assent to supporting the lockdown policies.

Another part of the professional community has highlighted COVID-19 models' methodological flaws, and their basic failure to predict events—presumably a *sine qua non* in a model.[115] Collins and Wilkinson conducted a systematic review of 145 COVID-19 prediction models published or preprinted between January 3 and May 5, 2020, and discovered pervasive statistical flaws. Different models suffered from small sample size, many predictors, arbitrarily discarded data and predictors, overfitted models, and a general lack of transparency about how they were created. These flaws frequently overlapped. In sum, "all models to date, with no exception, are at high risk of bias with concerns related to data quality, flaws in the statistical analysis, and poor reporting, and none are recommended for use."[116]

The public ought to be able to do more than simply take the word of one scientist or another. Unfortunately, the very complexity of models makes it extraordinarily difficult to provide a standard by which to hold them accountable—aside from the common-sense standard, *did they predict well?* Then, too, while models are considered sufficiently solid to inform policy immediately, they are tentative enough in their claims that a disproven model can always be disclaimed with a shrug and a reply that *we updated the data*. The failure of one parameter informs a new parameterization, not a skepticism of parameters in general. The failure of one prediction can be ignored with resort to the general and the counterfactual: *if you hadn't followed our advice generally, millions would be dead*. To say that a model failed is to invite the inevitable riposte, *we're doing it better now*.

---

112 Bertozzi (2020).
113 Biggerstaff (2022).
114 Chappell (2020).
115 For critiques of lockdown recommendations see Bendavid (2021); Chin (2021); Ioannidis (2021); Melnick (2020).
116 Collins (2021); and see Chin (2020); Howick (2022); Ioannidis (2022a); Levitt (2022a); Levitt (2022b); Nixon (2022); Zavalis (2022).

In our third *Shifting Sands* report, the researchers' technical studies focused on two aspects of nonpharmaceutical intervention response to the COVID-19 pandemic: lockdowns and masking, which were both meant to reduce COVID-19 infections and fatalities. They used p-value plotting to assess specific claims made about the benefit to public health outcomes of these responses.

The researchers found persuasive circumstantial evidence that lockdowns and masking had no proven benefit to public health outcomes.

P-value plot (p-value versus rank) for Herby et al. (2022) meta-analysis of the effect of COVID-19 quarantine (stay-at-home) orders implemented in 2020 on mortality. Symbols (circles) are p-values ordered from smallest to largest (n=20).
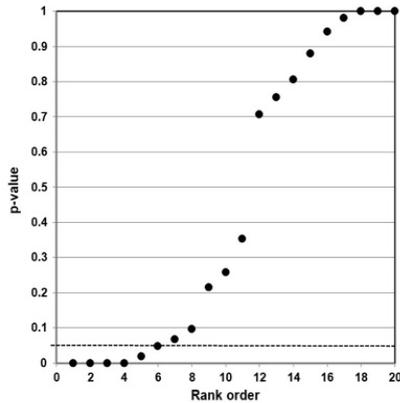
Meta-analysis p-value plots, masks: (a) 15 RCT base studies (Jefferson et al. 2020), (b) 7 RCT base studies (Xiao et al. 2020)
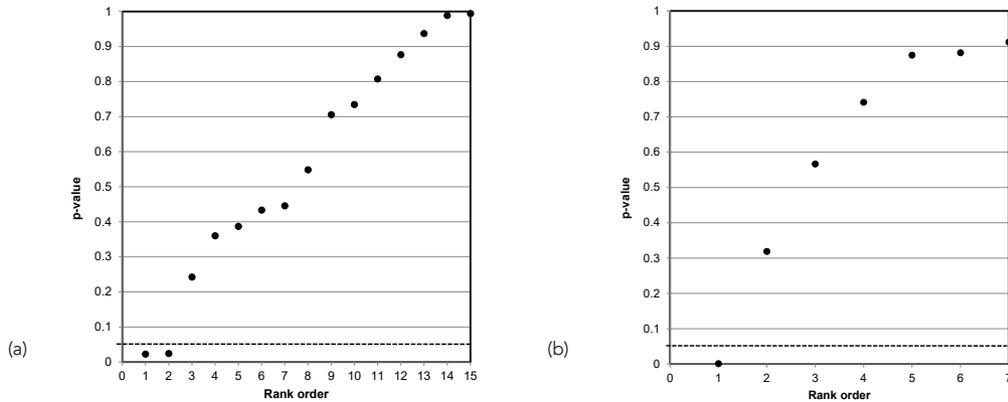
The technical studies suggest a far greater frailty (failure) in the system of epidemiological modeling and policy recommendations. That system, generally, grossly overestimated the potential effects of COVID-19 and, particularly, overestimated the potential benefit of lockdowns and masking. The technical studies support recommendations for policy change to restructure the entire system of government policy based on epidemiological modeling, and not simply to apply cosmetic reforms to the existing system.

Psychology: Implicit Bias Research and the Implicit Association Test

Psychology as a discipline embraced statistics early. Psychologists back to the nineteenth century, who aspired to make psychology a science, sought to use quantitative methods to make universal statements in the study of the mind. Psychologists therefore early seized on statistics, as a means of quantification that offered a way to make bold scientific arguments that acknowledged the inescapable fact that human minds varied.

Yet even beyond the fundamental critique that the ambition to make such universal statements may have no real-world foundation,[117] psychology's statistical revolution has been troubled. Psychologists played a prominent role in the unwieldy marriage of R. A. Fisher's approach to statistics and the "frequentist" approach derived from the work of Jerzy Neyman and Egon Pearson. Theoretical inconsistency about how to treat p-values fueled a "practical" ability to design psychological-statistical experiments. Psychology suffers as a discipline from running experiments with small sample sizes, and hence low *statistical power*—a low probability of a significance test detecting a true effect. The irreducible difficulties in defining mental characteristics, much less in establishing their comparability from individual to individual, limit the discipline's ability to conduct rigorous statistical experiments. It also embraces the loose definition of statistical significance at $p \leq 0.05$, rather than the tighter definitions embraced by other disciplines—some branches of physics, for example, use the "five sigma" standard of $p \leq 0.00005$. The social psychology subdiscipline appears to be unusually subject to politicized groupthink. Psychology, and especially social psychology, for all these reasons has been unusually afflicted by the irreproducibility crisis. Any psychological research conclusion based upon statistical techniques warrants especially close scrutiny of its methodological foundations.[118]

Implicit bias theory draws upon a series of pivotal psychology articles in the 1990s, above all the 1995 article of Anthony Greenwald and Mahzarin Banaji that first defined the concept of *implicit* or *unconscious bias*.[119] These articles argued, drawing upon the broader theory of implicit learning,[120] that individuals' behavior was determined regardless of their individual intent, by "implicit bias" or "unconscious bias." These biases significantly and pervasively affected individuals' actions, and were irremovable, or very difficult to remove, by conscious intent.

---

117 Staddon (2019).
118 Danziger (1990); Lamiell (2019); Nuttgens (2023); Randall (2018); Smedslund (2021); Stigler (1992); Vankov (2014); Yong (2018).
119 Greenwald (1995); WIPR (2020); Greenwald (2006).
120 Reber (1989).

Notably, researchers measured such biases in terms of race and sex—the categories of identity politics that fit with radical ideology, and which were at issue in antidiscrimination law.

Greenwald and Banaji also promoted the Implicit Association Test as a way to measure implicit bias. The IAT is one a series of attempts by psychologists since *ca.* 1970 to find a way to avoid false self-report and assess "true" individual bias. The IAT was meant to supersede the known frailties of earlier techniques, although it seems rather to have recapitulated them.[121]

Mitchell judges that "A review of the public record leaves little doubt that the seminal event in the public history of the implicit prejudice construct was the introduction of the IAT in 1998, followed closely by the launching of the Project Implicit website in that same year."[122] This publicity continued over the decades, notably including Banaji and Greenwald's popularizing 2013 book, *Blindspot: Hidden Biases of Good People*.[123] Greenwald et al. explicitly have sought to use their research to affect the operations of the law: "The central idea is to use the energy generated by research on unconscious forms of prejudice to understand and challenge the notion of intentionality in the law." These researchers, and others of their colleagues, have acted in legal education, as expert witnesses, promoting diversity trainings, promoting paid consulting services by Project Implicit, Inc., collecting federal grant money, and more. In so doing, they made bold but unsubstantiated claims about the solidity of implicit bias research and the importance of the effect. Acolytes have disseminated their arguments to audiences including the police, public defenders, human resource advisors, and doctors.[124]

All this has happened even as the intellectual underpinnings of implicit bias theory came under sustained and devastating assault. Implicit bias never acquired consensus support from psychologists—some published articles arguing for its validity, while others then critically examined the evidence and the theory. As it so happens, an increasing number of psychologists have provided evidence for devastating problems with implicit bias theory.[125] A further scholarly literature reviews the evidence critiquing implicit bias theory.[126] Other scholars have provided extensive critiques of the Implicit Association Test (IAT).[127] Yet more scholars have

---

121 Blanton (2023).
122 Mitchell (2017).
123 Banaji (2013).
124 Mitchell (2017).
125 For critiques of implicit bias research, see Andreychik (2012); Arkes (2004); Cone (2017); Corneille (2020a); Corneille (2020b); Corneille (2022); Cyrus-Lai (2022); Hahn (2020); Jussim (2018); Jussim (2020a); Mitchell (2017); Oswald (2015); Rubinstein (2018); Skov (2020); and see Chin (2023).
126 Blanton (2008); Cesario (2022); Gawronski (2019); Gawronski (2022); Jussim (2020b); Jussim (2023); Lai (2021); Mitchell (2017).
127 Anselmi (2011); Blanton (2009); Blanton (2015a); Blanton (2015b); Blanton (2017); Bluemke (2009); Carlsson (2016); Chequer (2021); van Dessel (2020); Fiedler (2006); Forscher (2019); Hahn (2014); Henry (2021); Hughes (2022); LeBel (2011); Oswald (2013); Oswald (2015); van Ravenzwaaij (2011); Rezaei (2011); Schimmack (2021); Unkelbach (2020).

critiqued interventions based on implicit bias theory.[128] A further scholarly literature reviews the evidence critiquing the IAT.[129]

Jussim (2020b) notes the real-world context of implicit bias research and the IAT:

- The coiners of implicit bias made public claims far beyond the scientific evidence.
- Activists find implicit bias and implicit bias trainings politically useful, as associated consultants find it financially lucrative, and bureaucrats find it useful as a way to address rhetorical and legal accusations of discriminatory behavior.
- Scientific and activist bias overstates the power and pervasiveness of implicit bias.[130]

Machery (2022) scathingly concludes that,

> We do not know what indirect measures measure; indirect measures are unreliable at the individual level, and people's scores vary from occasion to occasion; indirect measures predict behavior poorly, and we do not know in which contexts they could be more predictive; in any case, the hope of measuring broad traits is not fulfilled by the development of indirect measures; and there is still no reason to believe that they measure anything that makes a causal difference. These issues would not be too concerning for a budding science; they are anomalies for a 30-year-old research tradition that has been extremely successful at selling itself to policy makers and the public at large.[131]

On the whole, the defenders of implicit bias theory and the IAT simply have not responded to the full implications of these critiques. Such defenses as they have made are not very persuasive: one defense of the IAT and implicit bias theory, for example, is that "it doesn't follow from a particular measure being flawed that the phenomenon we are attempting to measure is not real."[132] Greenwald et al. in 2022 summarized their sense of implicit bias theory, and included responses to some, although not all, of the critiques presented above. While they indeed have provided counter-arguments for some of these critiques, their response is far from adequate or persuasive.[133]

---

128 Lai (2023); Paluck (2021).
129 Blanton (2008); Blanton (2023); Jussim (2020a); Jussim (2020b); Jussim (2023); Lai (2021); Meissner (2019).
130 Jussim (2020b).
131 Machery (2022a); and see Machery (2022b).
132 Payne (2018); Sulzer (2022).
133 Greenwald (2022).

Yet implicit bias theory has fueled a series of policy changes, completed and proposed, in the law, medicine, the Equal Employment Opportunity Commission, science, court personnel, the police, the jury system, and education. All of these policies, many of which center on required "implicit bias trainings" or "diversity trainings," subject an ever-expanding number of Americans to training in a concept which does not exist, to support imaginary charges of "systemic racism" or "systemic sexism," using a measure (IAT) that does not work. In the legal realm, adopting the implicit bias standard would allow lawyers to seize on the law stating that a "hostile environment" is an actionable offense under antidiscrimination law. Implicit bias raises any inequity, not least those detected by a statistical study, to be evidence of implicit bias, and hence a hostile environment. To adopt an implicit bias standard would replace individual intent with statistical associations—disparate impact—in anti-discrimination law.

It is dubious that such a thing as implicit bias even exists, and if there is such a thing, it is unlikely to be so hard-edged and pervasive as its proponents claim. Nor does the IAT, the tool that is supposed best to measure implicit bias, appear to measure it accurately or reliably. Policies devised to reduce implicit bias, moreover, seem to be either ineffective or counterproductive.

In our fourth *Shifting Sands* report, the researchers' technical studies used p-value plotting to assess the validity of the Implicit Association Test (IAT) by reviewing claims for IAT–real-world behavior correlations relating to race and sex. The researchers found persuasive circumstantial evidence that there was no association between IAT measures and real-world behavior or real-world perception, either for race (white behavior toward and perception of blacks) or for sex (effects of implicit bias upon womens' achievement in high-ability careers).

The second study also examined the role of *confounders*—unexamined variables that affect the analyzed variables, and which, when accounted for, alter their putative relationship—that implicit bias theory should have considered, and which further weaken that theory's evidentiary basis. P-value plotting of differing male-female vocational interests (personal interests and behaviors), for example, revealed a statistically significant correlation that the proponents of implicit bias theory ought to have considered.

The technical studies provide further evidence that there is no scientific support for implicit bias theory. The technical studies support recommendations to rescind all laws, regulations, and private initiatives based upon implicit bias theory. Policymakers should give priority to rescinding laws and regulations that affect the personnel involved in executing law and order, such as judges, lawyers, and policemen, as well as medical personnel. Private institutions and enterprises should be encouraged by public opinion to rescind all activities, such as diversity trainings, based on implicit bias theory.

Race: P-value plot of 87 correlations between IAT results and real-world microbehaviors.
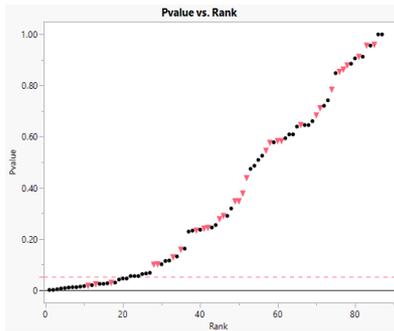
Race: P-value plot of 75 correlations between IAT results and real-world person perception measures.
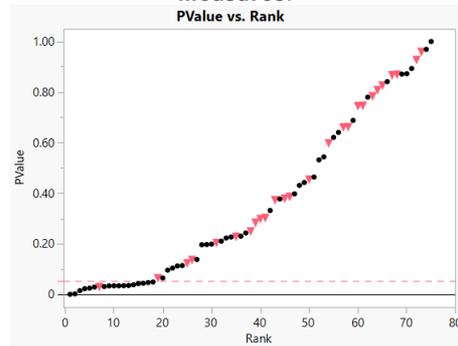
Note: black circle (●) ≡ +ve correlation, i.e., IAT result is positively correlated with micro-behavior; triangle (▼) ≡ −ve correlation, i.e., IAT result is negatively correlated with micro-behavior.

Rank-ordered p-values computed for 27 ICC (implicit−criteria measure) correlations from the Kurdi et al. and Kurdi and Banaji meta-analysis dealing with sex.
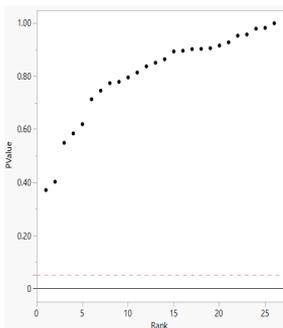
Note: p-values were computed from mean correlation coefficient (r) values for each study.

Rank-ordered p-values computed for 11 different vocational interest dimensions reported by Su et al.
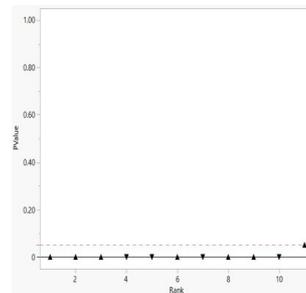
Note: (▼) vocational interest dimension favoring females; (▲) vocational interest dimension favoring males.

Policy Conclusions

T he *Shifting Sands* authors provide strong evidence that federal, state, and local governments have enacted policy based on false positive research results in four distinct areas: environmental epidemiology, nutritional epidemiology, public health epidemiology, and social psychology (implicit bias theory). Irreproducible research in all four areas has prompted extensive regulation at the federal level (EPA, FDA, CDC) and statutes and ordinances among the states and localities. These regulations and statutes are both economically burdensome and detrimental to Americans' individual liberty.

These authors' research reveals massive deformation of American scientific practice. It also reveals that this deformation in turn has deformed American regulatory practice. While there are many commenters on the irreproducibility crisis who advocate for voluntary changes of scientific practice, we do not believe that this is a realistic or adequate solution. America needs far greater changes to address the irreproducibility crisis—and the larger danger revealed by the irreproducibility crisis, the irresponsibility crisis of science policy. These failures are so widespread that they constitute a crisis of government, and not just a crisis of science.

We believe that the irresponsibility crisis of science policy must be addressed by four categories of reforms.

1. **Federal government agencies must make systematic changes to their regulatory and funding practices to remove the irreproducibility crisis from American science and the irresponsibility crisis from American government.**

    Scientists will not change their practices unless the federal government credibly warns them that it will withhold government grant dollars until they adopt stringent reproducibility reforms. Nor will federal regulators adopt stringent new tests of science underlying regulation unless policymakers explicitly require them to do so. Federal government agencies must make systematic changes to their regulatory and funding practices to remove the irreproducibility crisis from American science and the irresponsibility crisis from American government.

2. **Federal and state policymakers must make systematic changes to American K-12 and undergraduate science and math education to educate properly a new generation of American scientific professionals and informed citizens and policymakers.**

    American science's problems begin long before researchers acquire PhDs and start to apply for government grants. American K-12 and undergraduate science and mathematics education require systematic overhaul. American students, both those seeking STEM careers and those we prepare to be informed citizens and policymakers, need a proper education in statistics, experimental design, and the irreproducibility crisis. Indeed, they need a more rigorous preparation in general in mathematics and science. Above all, they need a depoliticized STEM education, which will not introduce politicized groupthink into basic science education, nor educate students from the beginning to believe that the point of science is (radical) policy activism. Federal and state policymakers must make systematic changes to American K-12 and undergraduate science and math education to educate properly a new generation of American scientific professionals and informed citizens and policymakers.

3. **Federal and state policymakers must end the arbitrary procedures of scientific research and scientific governmental regulation so as to preserve our liberty from arbitrary government.**

    The scientific weaknesses of American government, academia, and K-12 education articulate aspects of a far graver challenge. Radical activists permeating the body of American scientists and government regulators have weaponized the irreproducibility crisis. Wittingly or unwittingly, they now manufacture false positive research results as part of a sustained campaign to achieve their political goals via specious claim to scientific authority, the powers policymakers have delegated to regulatory authorities, and illiberal laws based on hollow science. Researcher degrees of freedom lead to intervention degrees of freedom. Intervention degree of freedom not only burden American prosperity but also, and more consequentially, erode American liberty, law, and republican self-government. Americans must recognize the coherent and extensive challenge that weaponized, politicized pseudo-science poses to American liberty. We must reform science policy as a coherent whole, to preserve our liberty from this radical, technocratic elite. We must put an end to both the arbitrary procedures of scientific research and the arbitrary powers of scientific governmental regulation.

4. **Policy institutes must dedicate themselves to science policy as a first-order priority, staff their institutes with personnel dedicated to science policy, and make science policy a priority.**

    Policy institutes must alert policymakers and the public to this danger—and they have not. They do not properly staff their institutes with personnel dedicated to science policy, they do not make science policy a priority, and they do not appear to be even aware that science policy exists as a coherent entity, which requires comprehensive policy solutions. Policy institutes must dedicate themselves to science policy as a first-order priority. Policy institutes must address not only the irresponsibility crisis of science policy but every aspect of science policy, with policy solutions for each discipline and each regulatory agency.

Americans must engage in linked, systematic reforms of Federal government agencies' regulatory and funding practices, state and local laws and regulations, and K-12 and undergraduate STEM education. They must do so by means of a coherent reform of science policy that aims to preserve our liberty by ending the arbitrary procedures of scientific research and arbitrary powers of scientific governmental regulation. To make this possible, policy institutes must dedicate themselves to science policy as a first-order priority, staff their institutes with personnel dedicated to science policy, and make science policy a priority, with policy solutions for each discipline and each regulatory agency.

Policymakers and citizens also should aim at major reform of the structure of the American university, to the federal government's indirect cost funding formulas, to prohibiting discrimination in the guise of programs and policies such as *diversity, equity, and inclusion*, and much more. But they should start by reforming the twin crises of the irreproducibility crisis of modern science and the irresponsibility crisis of modern science policy.

Policy Recommendations

We make four different categories of policy recommendations: *Government, STEM Education, Liberty*, and *Policy Institutes*. We provide the most specific recommendations in *Government*; our recommendations for each category are at increasing levels of generality. While we do not suggest a priority order, our suggested Government reforms will provide immediate solutions for the technical heart of the irreproducibility crisis and the irresponsibility crisis. Our suggested STEM Education reforms are intended to improve the training of American scientists in the long run, and (among other goals) reduce a future generation's propensity to engage in the slipshod, politicized procedures that fuel the irreproducibility crisis. Since our suggested Government recommendations should do much to defend Americans' prosperity from unfounded regulation, our suggested *Liberty* reforms focus on those aspects of the irresponsibility crisis that imperil Americans' liberty, law, and self-government. Our suggested Policy Institutes reforms are meant to provide the political infrastructure that will make these other reforms possible. We need policy institutes to dedicate themselves to science policy to make sure that the public and policymakers make these policy recommendations a priority, and to translate these policy recommendations into programmatic detail and statutory language.

Government

In our *Shifting Sands* reports, the authors made precise recommendations to the EPA, the FDA, and the CDC. Here we collate and synthesize those recommendations and apply them to all government agencies. All these recommendations are intended to bring federal agency methodologies up to the level of *best available science*, as per the mandate of The Information Quality Act.[134]

(Policymakers also should look at the overlapping suggestions in the National Association of Scholars' **Model Science Policy Code**: https://www.nas.org/policy/model-science-policy-code.)

We direct these recommendations partly to the personnel within those agencies, partly to the federal legislators who oversee these federal agencies, and partly to the

134 IQA (2000); OMB (2019).

executive branch personnel who may draft executive orders that apply to all federal agencies. All these individuals, as well as the broader world of policy institutes and American citizens, will need to work together to conclude how best to apply these recommendations in detail to each agency and each professional discipline.

These recommendations assume that policymakers largely preserve the existing structure of federal science regulation and funding. We make these recommendations to give policymakers an outline for immediate science policy reform. We emphasize, however, that these reforms should be considered a beginning, not an end.

We specify that federal science regulatory agencies should adopt these recommendations—but so too should state and local regulatory agencies. If the federal EPA should be reformed, so too should California's. State policy institutes should consider how to adapt these policy recommendations to the state and local level.

1. **Federal agencies should reform the statistical procedures by which they assess research.**

   a. **Federal agencies should adopt resampling methods (Multiple Testing and Multiple Modeling) as part of their standard battery of tests applied to research.**

   This resampling-based multiple testing procedure already has been incorporated into a variety of disciplines, including genomics[135] and economics,[136] and has been shown to be optimal for a broad class of testing problems.[137] Any discipline using statistics can incorporate these procedures into their regular tests. Every government agency that relies on scientific research should require the use of such procedures to test scientific research, before it is used to justify regulation, or qualify as *best available science*.

   All federal science regulatory agencies, in other words, should only rely on base studies and meta-analyses that use a resampling methodology (MTMM) to correct their results. The agencies also should subject all such research to independent MTMM analyses.

   b. **Federal agencies should take greater cognizance of the difficulties associated with subgroup analysis.**

   Groups and individuals vary sufficiently in their responses to the same substances that it is conceivable that federal science regulatory agencies

---

135 Ge (2003).
136 Jones (2019a); Jones (2019b); and see Romano (2016).
137 Cox (2008); Meinshausen (2011).

should not be attempting to give general advice to the public. Scientists and regulators therefore rightly aim to consider whether particular substances have different effects on different subgroups, defined by categories such as race and sex. Yet subgroup analysis multiplies the number of statistical operations and therefore multiplies the possibility of producing false positives. Federal agency policies for MTMM correction should include explicit and detailed consideration of how to apply them to subgroup analysis.[138]

c. **Federal agencies should require all studies that do not correct for MTMM be labeled "exploratory."**

Research that does not correct for MTMM is exploratory rather than confirmatory and should be labeled clearly as such.[139] Federal science regulatory agencies should follow up on this reform either by ruling that their regulatory decisions cannot rely on exploratory research or, as a second best, by requiring regulators to explain in detail why they include exploratory research in their weight-of-evidence assessments.

2. **Federal agencies should reform their procedures generally to address the irreproducibility crisis and the irresponsibility crisis.**

a. **Federal agencies should report the proportion of positive results to negative results in the research they fund.**

Federal science regulatory agencies' bureaucratic self-interest—and its mandate—will always incline their employees, consciously or unconsciously, to fund research that supports regulation. Federal agencies must make a conscious effort to ensure that the research they fund does not put a thumb on the scales of a field's research as a whole—that it does not fund an overabundance of false positive results and then say that the "weight of evidence" justifies regulation. Federal agencies should report the proportion of positive to negative results in the research they fund, with data reported for every program and discipline. Any program or discipline that reports more than 65% positive results in the research it funds should initiate a reform of its granting program, to counter the effects of bureaucratic self-interest and groupthink.

b. **Federal agencies should place greater weight on reproduced research.**

---

138 Cf. Van der Laan (2011).
139 Even an observational study corrected for multiple testing ultimately is only exploratory. Further policy reform might expand our recommendation here to apply to all observational studies, including those that do correct for multiple testing.

Improved statistical techniques will reduce the effects of the irreproducibility crisis in different scientific field. But such statistical tests cannot catch every sort of questionable research procedure. Indeed, research that passes every statistical test might still be a false positive. Federal science regulatory agencies therefore should increase the weight they assign to research that is not only reproducible, *but also reproduced*— and decrease the weight they assign to research that has not yet been reproduced.

c. **Federal agencies should constrain the use of "weight of evidence" to take account of the irresponsibility crisis.**

The "weight of evidence" principle generally facilitates arbitrary judgments as to what science should inform regulation. Self-interest will inevitably incline scientists and regulators, consciously or unconsciously, to weigh more heavily research that facilitates regulation. Groupthink redoubles the effects of consensus-thinking, which too easily discards research that fails to endorse the consensus. Wherever possible, federal science regulatory agencies should substitute transparent rules for "weight of evidence" judgments, in particular the rules for accepting or rejecting papers to be used in a meta-analysis. Federal agencies also should require regulators to elaborate in detail whenever they apply a "weight of evidence" judgment, by means of a coherent argument which can be falsified by independent critique.

3. **Federal agencies should reduce regulatory reliance on research that uses questionable research procedures, HARKing, and p-hacking.**

a. **Federal agencies should require preregistration and registered reports of all research that informs regulation.**

Preregistration and registered reports, using the procedures and resources of organizations such as the Center for Open Science (https://www.cos.io), will constrain the ability of scientists to HARK, and generally inhibit p-hacking and questionable research procedures. Preregistration and registered reports are not cures. Determined scientists in time undoubtedly will devise methods to undermine the effectiveness of these precautions. But preregistration and registered reports *will* substantially improve the reliability of research used by federal science regulatory agencies. Federal agencies should stipulate that all

preregistration and registered reports must detail the MTMM methods that will be used to assess results.

b. **Federal agencies should require public access to all research data used to justify regulations.**

Federal science regulatory agencies should require that all research used to justify regulation must provide public access to the underlying research data. Scientists often claim that analysis data sets cannot be made public because they must prevent the disclosure of the identity of human subjects. This claim is not persuasive, since we now possess standard methods such as micro-aggregation that can prevent such disclosures.[140] Scientists should be expected to use such procedures as standard practices. Federal agencies should direct all necessary funding to ensure de-identification of human data.[141] and provide an adequate means to address all privacy and confidentiality concerns. But these are challenges that the federal government can and must meet, not convenient obstacles that prevent public access.[142]

c. **Federal agencies should fund data set building and data set analysis separately.**

Researchers who combine data collection and data analysis possess an incentive to selectively adjust data to improve results of their analyses. Federal science regulatory agencies should separate these two functions, to remove this incentive. They also should require researchers to provide a hold-out data set to a trusted third party before analysis, so that analysis claims can be tested independently using the hold-out data set.

d. **Federal agencies should rely for regulation exclusively on meta-analyses that use severe tests for endemic questionable research procedures, HARKing, and p-hacking.**

When federal science regulatory agencies use meta-analyses or a systematic review to justify regulation, they should rely only on meta-analyses that conduct rigorous tests to detect whether a field's base studies have been affected by questionable research procedures, HARKing, and p-hacking. Since so many base studies are unreliable, the meta-analyses

---

140 El Emam (2009).
141 For a beginning, see Gal (2014); Kushida (2012).
142 Cecil and Griffin have noted how an agency can insulate its actions from public scrutiny by funding a grant for controversial research and then basing its action on those findings. As long as the agency does not take possession or control of the records, FOIA requests—or other procedures to facilitate public oversight—will not assist those who wish to challenge the findings the agency relies on to justify its actions. Cecil (1985). The requirement for public access to research data also will ensure that federal agencies do not undertake maneuvers of this nature.

which collate these base studies likewise have become unreliable—
Garbage In, Garbage Out. While we will not prescribe further particular
methods here, we state that existing tests are not sufficient.[143] Federal
agencies should adopt tests substantially more stringent than those they
currently accept.

4. **Federal agencies should reform their use of mathematical modeling.**

    a. **Federal agencies should reconceive of modeling as measuring**
        uncertainty.

        Gelman has severely criticized the use of the term *confidence interval*,
        which gives unwary researchers the mistaken impression that a statis-
        tical operation can and should be used to establish sufficient knowledge.
        He prefers the term *uncertainty interval*, although Greenland prefers
        compatibility interval. These changes in nomenclature are intended to
        reinforce the truth that statistics can and should aim at measuring un-
        certainty rather than establishing certainty.[144] This concept also should
        be applied to modeling, especially where it depends upon statistical op-
        erations. As Briggs puts it, "The goal of probability models is to quanti-
        fy uncertainty in an observable Y given assumptions or observations X.
        That and nothing more."[145] Federal science regulatory agencies should
        formulate guidelines that make explicit that modeling is meant to quan-
        tify uncertainty, and that models should communicate to policymakers
        a quantification of the uncertainties of action rather than a prescription
        of certainty to justify action.

    b. **Federal agencies should require preregistration of mathematical**
        modeling studies.

        Federal science regulatory agencies should formulate rules requiring
        the pre-registration of mathematical modeling studies, including pro-
        spective validation practices; pre-specified, agreed-upon rules for judg-
        ing success and/or the need for recalibration; registries of existing past
        models; data, code, and software sharing and reporting transparency;
        and unbiased reporting and complete documentation of past model
        performance.[146]

---

143 Carter (2019).
144 Gelman (2019).
145 Briggs (2018); and see Briggs (2016).
146 Ioannidis (2022b).

c. **Federal agencies should require mathematical modeling transparency and reproducibility.**

Federal science regulatory agencies should formulate rules requiring greater reliance on unbiased data and less reliance on theoretical speculation; transparent release of underlying data and models, to allow anyone to analyze model input data, model predictions, and model outcome data; division of data set construction from data set analysis; modeling the entire predictive distribution, with a particular focus on accurately quantifying uncertainty; continuously monitoring the performance of any model against real data and either re-adjusting or discarding models based on accruing evidence; avoiding unrealistic assumptions about the benefits of interventions; using up-to-date and well-vetted tools and processes that minimize the potential for error through auditing loops in the software and code; maintaining an open-minded approach and acknowledging that most forecasting is exploratory, subjective, and non-pre-registered research; and articulating efforts to avoid selective reporting bias.[147] Federal agencies also should limit and require articulate defenses of all arbitrary "weight of evidence" judgments that inform mathematical models.[148]

5. **Federal agencies should reform their funding to researchers and organizations to address the irreproducibility crisis and the irresponsibility crisis.**

a. **Federal agencies should not fund or rely on research claims of other organizations until these organizations adopt sound scientific practices.**

Federal science regulatory agencies often fund external organizations, such as the World Health Organization (WHO), the International Agency for Research on Cancer (IARC), and the Health Effects Institute (HEI). These organizations are effectively beyond the reach of effective oversight. Federal agencies should not fund or rely on research claims of other organizations until these organizations adopt the best available science methodologies we recommend here.

b. **Federal agencies should only fund research that adopts sound scientific practices.**

---

147 Ioannidis (2022a); Young (2021).
148 Cox (2020).

Federal science regulatory agencies provide the largest single source of science funding in the world, and hence possess vast power over the conduct of scientific research. Federal science granting agencies should establish procedures to require grantees to adopt the best available science methodologies we recommend here.

c. **Federal agencies should increase funding to investigate direct causal biological links between substances and health outcomes.**

Epidemiology, both environmental and nutritional, depends on establishing statistical associations in default of establishing direct causal biological links between substances and health outcomes. The EPA's and FDA's reliance on association rather than causation weakens the justifications of its regulations. Federal science regulatory agencies should redirect grant funding toward investigating direct causal biological links between substances and health outcomes, to minimize their reliance on statistical associations. Federal agencies also should place substantially greater weight on negative results in research to establish direct causal biological links. They should also establish a set procedure by which a sufficient number of such negative results preclude regulation absent research that proves statistical association to a substantially higher standard of rigor than at present required.

6. **Federal agencies should seek regulatory stability as they put more rigorous science standards into effect.**

Federal science regulatory agencies should not overturn previous regulations arbitrarily as they put more rigorous science standards into effect. *Regulatory stability* is an important goal for the Federal government, and indeed for any system of laws and regulations. American enterprises have invested substantial resources in regulatory compliance, and their investments should not casually be set at naught.[149] These reforms should be introduced via the federal agencies' regular, planned regulatory reviews, which will allow the reform of procedures to enact new regulations to proceed in an orderly manner.[150] But these regulatory reviews should not exempt existing regulations. We should not grandfather bad science forever—or even very long.

---

149 Randall (2020).
150 Randall (2020).

For a highly relevant example, consider the Harvard Six Cities/ACS studies that are cited in support of current $PM_{2.5}$ regulation.[151] The government should announce that it will cease using the Harvard Six Cities/ACS studies, and similarly irreproducible data sources, by some reasonably near date, unless the underlying data have been made publicly available. As the same time, the government should immediately begin to fund a high-priority program to create a new, substitute data set, with born-open, publicly accessible data and built-in de-identification to address any privacy concerns. These data will then be available for the EPA to use once it ceases using the Harvard Six Cities/ACS studies and similarly irreproducible data sources. If the new data do not justify the regulations, then the regulations can be withdrawn in an orderly manner. If the new data do justify the regulations, then the regulations can be continued. This multi-part reform should maximize reproducibility reforms and regulatory stability.

Similarly crafted multi-part reforms, enacted throughout the federal agencies' remit, ought to maximize the twin goods of good science and stable regulation.

7. **Federal agencies should establish systematic procedures to inhibit research integrity violations.**

Federal science regulatory agencies should establish systematic procedures to inhibit research integrity violations. We may phrase this positively as a call for the federal agencies to mandate a system of Good Institutional Practices (GIP) for all recipients of federal research money, and for all research that informs federal regulatory decisions.[152] GIP should include practices such as:

a. annual training for principal investigators and students in applying research ethics to data analysis (e.g., lessons to avoid bad practices such as p-hacking and HARKING);

b. random audits of laboratory note books;

c. whistleblower systems for research integrity violations;

d. real consequences for delinquent researchers, including bars on grant applications, lost lab space, and bars on accepting new members into their research groups;

e. annual reporting requirements by institutions receiving federal research funds;

---

151 Dockery (1993); Pope (1995).
152 Begley (2015).

    f.   real consequences for institutions that fail to enforce GIP in their institutions, including institutional loss of eligibility for government funding; and

    g.   established procedures within the federal government to ensure compliance with GIP guidelines.

8.   **Federal policymakers should establish an Irreproducibility Crisis Commission.**

Federal policymakers should charter a commission to advise federal science regulatory agencies how to achieve the recommendations we have outlined in this section. This commission should include experts such as William M. Briggs, Andrew Gelman, and John Ioannidis, as well as regulatory draftsmen who can articulate these recommendations in a form that they can be used at once to reform federal regulatory procedure. These recommendations should be articulated in an appropriate manner for each individual science regulatory agency.

9.   **Federal policymakers should appoint mission-oriented personnel to carry out these reforms.**

The existing cadre of federal government personnel created or has grown familiar and attached to the procedures that have created the irresponsibility crisis. Too many are activists who actively foment that crisis to pursue their political ends. Policymakers and the public must expect that the bulk of existing federal personnel will work to cripple these reforms. Federal policymakers must appoint to every relevant federal agency a new body of mission-oriented administrators and experts, dedicated to carrying out all the Government policy recommendations enumerated above. This final reform is essential to put all the previously listed statutory and regulatory reforms into effect.

STEM Education

The irreproducibility crisis and the irresponsibility crisis derive in good measure from weaknesses in America's existing STEM education. STEM education throughout should focus on instruction in statistics and experimental design, to educate future scientists to avoid the procedures that lead to these twin crises, and to educate future citizens and policymakers to be aware of and detect irreproducible research and irresponsible policy. STEM education also must be depoliticized, to prevent groupthink—and to prevent the emergence of a generation of scientists who

believe that the point of scientific research is to forward (radical) political activism rather than to discover the truth.

1. **K-12 Education**

   K-12 science education has been softened and politicized by misguided education school pedagogy, by the effect of Common Core mathematics standards and Next Generation Science Standards, and by licensure standards that allow unqualified education school majors to teach science and prevent qualified science majors from teaching. K-12 science education should be reformed by:

   a. replacing the Next Generation Science Standards (NGSS) with rigorous and depoliticized science standards (such as the National Association of Scholars' and Freedom in Education's *The Franklin Standards: Model K-12 State Science Standards*, https://www.nas.org/reports/the-franklin-standards);

   b. requiring basic instruction in the history of science, statistical literacy, risk analysis, the scientific method, experimental design, and the irreproducibility crisis;

   c. strengthening subject matter requirements and removing education school requirements from K-12 science teacher licensure and professional development (**Education Licensure Certificate Act**, National Association of Scholars, https://www.nas.org/policy/model-education-licensure-code/education-licensure-certificate-act);

   d. drafting model professional development courses for science teachers that include advanced training in the history and philosophy of science, statistical and mathematical numeracy, the major branches of the natural sciences and how they differ in practice and philosophy, and how science can constructively serve public interests;

   e. ensuring that high school graduates know enough mathematics and statistics to equip them for science majors and science careers.

2. **Undergraduate Education**

   Undergraduate science and social science education should be reformed by:

   a. strengthening General Education Requirements in the history of science, statistical literacy, the scientific method, experimental design, and the irreproducibility crisis;

b. strengthening departmental major and minor requirements, in the sciences and the social sciences, in the history of science, statistical literacy, the scientific method, experimental design, and the irreproducibility crisis;

c. removing politicized science and social science courses from General Education Requirements and departmental major and minor requirements.

3. **Graduate Education**

Federal grants have transformed graduate science education into a quasi-feudal structure, where university professors who receive grants in turn fund cohorts of graduate students who learn to defer to their paymaster rather than to cultivate an ethic of free and independent inquiry. Federal research support has subsidized the overproduction of science graduates, independent of the nation's employment needs; has enabled the imposition of discriminatory practices in recruitment of young scientific talent; and has undermined merit-based criteria in identifying promising scientific talent. Policymakers should establish a Portable Graduate Fellowship (PGF) to replace the existing dysfunctional model of graduate science education. The PGF should:

a. Make awards to individual graduate students, not to institutions or tied to specific research grants. PGFs will go where the student goes, as does the award money for National Science Foundation's existing Graduate Research Fellow Program.[153]

b. Base awards upon evaluation of demonstrable promise of scientific talent. Awards will include both salary and tuition support, as well as funds for conducting research.

c. Ensure that awards will allow the recipient to enroll in a university graduate program, in a commercial research and development laboratory, or in a government agency with a legitimate research interest, such as military or national security research.

d. Require that the award recipient, and those the recipients work with, adhere to government protocols for open science, reproducibility, and research integrity.

4. **Legal and judicial education**

---

153 NSF (n.d.).

Federal and state policymakers should mandate that law schools, continuing legal education, and continuing judicial education provide courses for lawyers and judges on the irreproducibility crisis, science and social science research, and best legal and judicial practices for assessing science and social science research and the testimony of expert witnesses.

Liberty

Researchers and government officials attempt to subsume a remarkably large number of subject matters under "public health," including secondary ("perimetric") boycotts of institutions funded by tobacco companies,[154] fossil fuel divestment,[155] Independence Day fireworks,[156] so-called "anti-racism,"[157] the anti-Israel Boycott, Divest, and Sanction (BDS) movement,[158] and "social policy" generally.[159]

With such a wide remit, Americans well may fear that researchers and government officials will abridge free speech in the name of public health, while using public health techniques. Twitter, for example, already has blacklisted dissenters from the government's COVID-19 policy to reduce the influence of their skepticism.[160] Even broader interventions are more than plausible. Outside the realm of epidemiology, for example, machine-learning experts have been exploring how to remove what they call "hate speech."[161]

Epidemiology already concerns itself with "surveillance" in the health context. It is reasonable to worry about the conflation of public health modeling and the parallel work by computer scientists to establish a broader surveillance state, to fear the marriage of the epidemiological model with the computer science algorithm. Meme transmission can be modeled; so can "public health" efforts to inhibit the reproduction of memes.

Policymakers need to act broadly to protect liberty from the challenges of arbitrary policymaking, using the techniques of epidemiology and modeling.

1. **Federal agencies should reduce intervention degrees of freedom.**

   Federal science regulatory agencies should formulate rules to reduce intervention degrees of freedom. These rules will overlap with those for pre-registration, transparency, and reproducibility, but they should be framed explicitly to reduce regulatory bureaucrats' degrees of freedom to enact

---

154 Offen (2005).
155 Cooper (2019).
156 Mousavi (2021).
157 Blanding (2021).
158 BNC (2021).
159 Terris (2011).
160 Hart (2022); Nelson (2022).
161 Vishwamitra (2021).

policy without full transparency and accountability to policymakers and the public.

2. **Federal policymakers should establish a Liberty Commission.**

    Congress and the president should jointly convene an expert commission, drawing upon noted defenders of civil liberties such as Greg Lukianoff and Glenn Greenwald, as well as epidemiological experts in different agencies and professions, to delimit the areas of private life which may be subject to public health interventions. This commission also should draft rules articulating the principles it has drafted as detailed guidelines limiting what public health interventions, or research regarding health interventions, any federal government may fund, conduct, or allow.

    a. **Define Scope of Public Health Interventions.**

        The Liberty Commission's rules should limit explicitly the scope of public health interventions to physical health, narrowly and carefully defined, and explicitly define public health not to include any aspect of concepts such as *mental health*, *environmental health*, or *social health*. Public health authorities should be prohibited from intervening in matters that properly should be decided freely by individuals or by their elected policymakers.

    b. **Define Scope Narrowly.**

        The Liberty Commission's rules should limit explicitly and narrowly how public health interventions may change individual and collective behavior, and that all such public health interventions be required to receive explicit sanction from both houses of Congress. Above all, public health interventions should not be allowed to aim to alter public judgment of a public policy. Public judgment should determine public health policy, not vice versa.

3. **Federal policymakers should establish a COVID-19 Commission.**

    Federal policymakers should commission a full-scale report on the origins and nature of COVID-19, as well as of public health policy errors committed during the response to COVID-19. Errors to be investigated should include every instance of politicization of COVID-19 public health policy, and censorship of discussion of COVID-19 policy, as well as the role of public and private entities (e.g., social media companies) in forwarding politicization and

censorship.[162] This commission should be empowered to subpoena data from all relevant government agencies and private entities and to publicize it. It should also present concrete suggestions for reforms to prevent the recurrence of policy errors, politicization, and censorship.

While such a commission should include articulate defenders of what the government did correctly, it also should include large numbers of professional critics of government policy, such as John Ioannidis, Jay Bhattacharya, and Martin Kulldorf. This commission, moreover, should be directed not to require a consensus report, but to welcome divisions of opinion, with majority and minority reports. The public should welcome, and be accustomed to, the idea that experts disagree.

4. **Federal policymakers should establish a Computer Science Commission.**

    Public health modeling naturally aligns with the use of computer science algorithms. Social media censorship of COVID-19 policy discourse depended on both. Public health modeling is well suited to provide a plausible justification for using computer science algorithms to limit public debate—and, despite all its methodological flaws, may provide useful techniques for censorship that abrogates Americans' First Amendment rights. When public health defines the transmission of ideas as a communicable disease that threatens public health, it has a broad arsenal of tools to inhibit such transmission. Federal policymakers also should establish a commission to provide guidelines for federal funding, conduct, and regulation of the use of computer science algorithms, particularly as they are used by the federal government and by social media companies. This commission, moreover, should provide guidelines to ensure that artificial-intelligence programming is not similarly subverted to inhibit liberty.

5. **Federal, state, and local policymakers should rescind all laws, regulations, and programs based on implicit bias theory.**

    Implicit bias theory and the IAT have no scientific validity. No Americans should be subject to policy based on nonsense—much less policy intended to promote radical identity politics ideology. Policymakers should give priority to rescinding regulations that affect the personnel involved in executing law and order, such as judges, lawyers, and policemen, and medical personnel. Private enterprises should be encouraged by public opinion to rescind all activities, such as diversity trainings, based on implicit bias theory.

---

162 Editorial Board (2021); Hart (2022); Kulldorff (2020); Mosher (2022); Nelson (2022); Wittkowski (2022).

6. **Federal policymakers should establish a Social Sciences Commission.**

    Federal agencies such as the NIH or the EPA have procedures for requiring that regulations be founded on substantial scientific research. The procedures may not yet take due account of the irreproducibility crisis, but they exist. Generally, no equivalent exists to guide, for example, the invocation of implicit bias by the U.S. Department of Education's *Guiding Principles: A Resource Guide for Improving School Climate and Discipline* (2014).[163] Social sciences such as psychology are not conceived to be as rigorous as physics or chemistry, and a resource guide does not have the immediate effect of an EPA regulation. Nevertheless, some procedures need to be applied to all publications by the federal government, including both regulations and resource guides, to determine whether research that invokes concepts such as implicit bias has sufficient scientific justification. Federal policymakers should establish a Social Sciences Commission to determine general guidelines, with due weight given to transparent data, preregistration, proper statistical controls, publication bias, politicized groupthink, and all the aspects of the irreproducibility crisis. Each individual department and agency should then be required to apply the commission's guidelines to their own procedures and publications.

7. **Federal and state legislators should establish committees to oversee social scientific support for proposed laws and regulations.**

    Federal and state legislatures should establish dedicated committees to investigate and provide judgment on all bills and new laws that use social science research to justify their policies. Permanent committees, with permanent staff, will be able to provide informed judgment on all such bills and new laws. These committees should have the power to inform their fellow policymakers and the public about the social scientific support, or lack thereof, for new bills and new laws. Other committees should have the option to send a new bill to these social science committees for their judgment, although they should not be required to do so.

8. **Federal and state legislators should pass resolutions that state guiding principles for the judicial system, legal education, and the operation of the law.**

    Federal and state legislators should pass resolutions that state guiding principles for the judicial system, legal education, and the operation of the law.

---

163 USDE (2014).

These resolutions should state that individual behavior and events, and the first principles of due process, the presumption of innocence, and individual responsibility, should govern the operations of the law and determine the course of justice; and that no argument or policy based on statistical disparities should have any role in the operations of the law. This principle should apply at least to:

a. The education of judges, jurors, policemen, court personnel, and any other state employee involved in executing law and order;
b. The training, work requirements, and promotion requirements for judges, jurors, policemen, court personnel, and any other state employee involved in executing law and order;
c. Police enforcement of the law;
d. Jury selection;
e. Jury verdicts; and
f. Judicial decisions.

Policy Institutes

America faces an extraordinary challenge to its ideals and institutions of liberty and republican self-government—and only a few, uncoordinated champions work to oppose this onslaught. The vast majority of scientists, and government experts employed to judge scientific research, either are on the political left, hence naïve or Machiavellian practitioners of political groupthink, or do not wish to jeopardize their chances of receiving government grants by opposing the politicized majority. A scattering of scientists in every discipline speak up against the false consensuses imposed by radical activists, but they lack the institutional support to defend rigorous science or constitutional liberty.

Policy institutes ought to provide an institutional network to substitute for the weakness of scientific dissent in the academy—but they do not. Since 2016, for example, the Heritage Foundation, arguably the most influential traditional-minded policy institute, has published only four reports on science policy.[164] The American Enterprise Institute likewise devotes limited resources to science policy and largely produces op-ed pieces, not policy papers.[165] The Cato Institute's Center for the Study of Science was largely focused on climate change, and it was shut down in 2019.[166] The Competitive Enterprise Institute focuses on select aspects of science policy, but

---

164 Heritage (n.d.).
165 AEI (n.d.).
166 Waldman (2019).

not on science policy as a whole.[167] *The New Atlantis* (https://www.thenewatlantis.com) provides a home for intelligent journalism on science policy, but it is a journal seeking to influence the culture rather than to achieve programmatic reform of science policy.

Policy institutes do not provide dedicated personnel or an institutional focus on science policy; hence neither policymakers nor the public make science policy a priority. Indeed, policymakers, policy institute personnel, and members of the public are scarcely aware that science policy as a coherent whole exists. Much less do they realize that progressive activist science policy as a whole poses a clear and present danger to Americans' liberty and republican self-government, and that they need urgently and immediately to create a new range of coherent science policy initiatives to ward off the danger presented by radical activists camouflaged as scientists.

Policy institutes should re-articulate their mission to include a coherent focus on science policy. They also should declare that this focus is a first-order priority. They should implement these statements by funding permanent personnel to articulate and publicize science policy, and by establishing websites and journals dedicated to science policy reform. Federal policy institutes should focus on federal regulations and the federal judicial system, while state policy institutions should focus on public K-12 and undergraduate education, as well as state judicial systems.

The NAS strongly urges that policy institutes dedicate themselves at once to fulfil this new mission.

---

167 CEI (n.d.).

Conclusion

Policy that seeks to restrict freedom must justify itself against the null hypothesis of a free republic—that it is better for government to do nothing and for the republic's citizens to exercise their freedoms untrammeled. This has long been the spirit of American science policy. Our policymakers, representing the American people, long ago decided that science regulations must justify themselves with the *best available science*—that is, science that has passed the severest tests. They used this phrase to defend liberty, not to facilitate its abrogation; to restrict regulation to the least necessary and not to facilitate the expansion of government regulation. *Best available science* was meant to restrict government bureaucrats, not to authorize them to build regulatory empires.[168]

That principle has been reversed in practice. Activists and technocrats jointly use science regulation to advance their policy goals. The irreproducibility crisis of modern science, fueled above all by scientific researchers' shift to statistical procedures, provides the occasion for the mass production of false positive research results, to forward the mass production of illiberal, radical regulations. The activists and technocrats in government service are not accountable to policymakers or the public. The irreproducibility crisis of modern science is the irresponsibility crisis of modern science policy.

The NAS builds upon its research in *The Irreproducibility Crisis* and its *Shifting Sands* reports to suggest a wide variety of reforms to address this crisis—reforms to government regulation and funding, reforms to science education, reforms to defend liberty coherently from government science policy, and reforms to focus the attention of policy institutes on science policy. These reforms must be enacted if we are to defend American liberty from the arbitrary exercise of power by radical activists acting in the name of science.

These reforms are only part of what must be done. The universities must be reformed. The illiberal groupthink of *diversity, equity, and inclusion* must be banished from American science. The fanatics, the money-grubbers, the power-mad bureaucrats, and the pliable time-servers must be removed from America's scientific

---

168 Buchanan (2004).

establishment. The spirit of American science must be rekindled, devoted to the discovery and free communication of scientific truth.

The NAS urges America's citizens, policymakers, and policy institutes to take up this challenge. Science and research procedures should be built on the solid rock of transparent, reproducible, and reproduced scientific inquiry, not on shifting sands; and science regulatory policy likewise should be built on transparent and accountable procedures. Americans must dedicate themselves to the proper government of science policy, to assure that they retain self-government. We have been inattentive for too long to what is done in the name of science. We must, forthwith, take responsibility for the world of science policy.

Bibliography

Acharjee M. K., Das, K., Young, S. S. 2017. Air quality and lung cancer: Analysis via Local Control. Conference: Joint Statistical Meetings. https://www.research-gate.net/publication/321695891_Air_quality_and_lung_cancer_Analysis_via_Local_Control.

Adiga, A., Dubhashi, D., Lewis, B. et al. 2020. Mathematical models for COVID-19 pandemic: A comparative analysis. *Journal of the Indian Institute of Science* 100, 4: 793–807. https://doi.org/10.1007/s41745-020-00200-6.

AEI (American Enterprise Institute). N.d. Science Policy. https://www.aei.org/tag/science-policy/.

Anderson, H. R., Favarato, G., Atkinson, R. W. 2013. Long-term exposure to air pollution and the incidence of asthma: Meta-analysis of cohort studies. *Air Quality, Atmosphere & Health* 6: 47–56. https://link.springer.com/article/10.1007/s11869-011-0144-5.

Andreychik, M. R., and Gill, M. J. 2012. Do negative implicit associations indicate negative attitudes? Social explanations moderate whether ostensible "negative" associations are prejudice-based or empathy. *Journal of Experimental Social Psychology* 48: 1082–93. https://doi.org/10.1016/j.jesp.2012.05.006.

Anselmi, P., Vianello, M., and Robusto, E. 2011. Positive associations primacy in the IAT: A many-facet rasch measurement analysis. *Experimental psychology* 58, 5: 376–84. https://doi.org/10.1027/1618-3169/a000106.

Archer, E. 2020. The intellectual and moral decline in academic research. *The James G. Martin Center for Academic Renewal*, January 29, 2020. https://www.jamesgmartin.center/2020/01/the-intellectual-and-moral-decline-in-academic-research/.

Arkes, H. R., and Tetlock, P. E. 2004. Attributions of Implicit Prejudice, or "Would Jesse Jackson 'Fail' the Implicit Association Test?" *Psychological Inquiry* 15, 4: 257–78. https://doi.org/10.1207/s15327965pli1504_01.

Aschwanden, C. 2016. You Can't Trust What You Read About Nutrition. *FiveThirtyEight*. https://fivethirtyeight.com/features/you-cant-trust-what-you-read-about-nutrition/.

Bachmann, J. D. 2007. Will the circle be unbroken: A history of the US National Ambient Air Quality Standards. *Journal of the Air & Waste Management Association* 57, 6: 652–97. https://doi.org/10.3155/1047-3289.57.6.652.

Baker, M. 2016. 1,500 scientists lift the lid on reproducibility. *Nature* 533, 7604: 452–4. http://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970.

Banaji, M., and Greenwald, G. 2013. *Blindspot: Hidden Biases of Good People*. New York, NY: Delacorte Press.

Barton, S. 2000. Which clinical studies provide the best evidence? The best RCT still trumps the best observational study. *BMJ (Clinical research ed.)* 321, 7256: 255–56. https://doi.org/10.1136/bmj.321.7256.255.

Battaglia Richi, E., Baumer, B., Conrad, B., Darioli, R., Schmid, A., and Keller, U. 2015. Health risks associated with meat consumption: A review of epidemiological studies. *International Journal For Vitamin and Nutrition Research* 85, 1—2: 70–8. https://doi.org/10.1024/0300-9831/a000224.

Begley, C. G., Buchan, A. M., and Dirnagl, U. 2015. Robust research: Institutions must do their part for reproducibility. *Nature* 525, 7567: 25–7. https://doi.org/10.1038/525025a.

Bendavid, E., Oh, C., Bhattacharya, J., and Ioannidis, J. P. A. 2021. Assessing mandatory stay-at-home and business closure effects on the spread of COVID-19. *European Journal of Clinical Investigation* 51, 4: e13484. https://doi.org/10.1111/eci.13484.

Benjamini, Y., and Hochberg, Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B* 57, 1: 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Bertozzi, A. L., Franco, E., Mohler, G., Short, M. B., and Sledge, D. 2020. The challenges of modeling and forecasting the spread of COVID-19. *Proceedings of the National Academy of Sciences of the United States of America* 117, 29: 16732. https://doi.org/10.1073/pnas.2006520117.

Biggerstaff, M., Slayton, R. B., Johansson, M. A., and Butler, J. C. 2022. Improving pandemic response: Employing mathematical modeling to confront coronavirus disease 2019. *Clinical Infectious Diseases* 74, 5: 913–7. https://academic.oup.com/cid/article/74/5/913/6338173.

Blanco Mejia, S., Messina, M., Li, S. S., Viguiliouk, E., Chiavaroli, L., Khan, T. A., Srichaikul, K., Mirrahimi, A., Sievenpiper, J. L., Kris-Etherton, P., and Jenkins, D. J. A. 2019. A meta-analysis of 46 studies identified by the FDA demonstrates that soy protein decreases circulating LDL and total cholesterol concentrations

in adults. *The Journal of Nutrition* 149, 6: 968—81. https://doi.org/10.1093/jn/nxz020.

Blanding, M. 2021. Symposium encourages 'anti-racism' focus for public health. *Harvard T. H. Chan School of Public Health*, September 29, 2021. https://www.hsph.harvard.edu/news/features/symposium-encourages-anti-racism-focus-for-public-health/.

Blanton, H., and Jaccard, J. 2008. Unconscious racism: A concept in pursuit of a measure. *Annual Review of Sociology* 34: 277–97. https://doi.org/10.1146/annurev.soc.33.040406.131632.

Blanton, H., Jaccard, J., Klick, J., Mellers, B., Mitchell, G., and Tetlock, P. E. 2009. Strong claims and weak evidence: Reassessing the predictive validity of the IAT. *The Journal of Applied Psychology* 94, 3: 567–603. https://doi.org/10.1037/a0014665.

Blanton, H., Jaccard, J., and Burrows, C. N. 2015a. Implications of the Implicit Association Test D-transformation for psychological assessment. *Assessment* 22, 4: 429–40. https://doi.org/10.1177/1073191114551382.

Blanton, H., Jaccard, J., Strauts, E., Mitchell, G., and Tetlock, P. E. 2015b. Toward a meaningful metric of implicit prejudice. *The Journal of Applied Psychology* 100, 5: 1468–81. https://doi.org/10.1037/a0038379.

Blanton, H., and Jaccard, J. 2017. You can't assess the forest if you can't assess the trees: Psychometric challenges to measuring implicit bias in crowds. *Psychological Inquiry* 28, 4: 249–57. https://doi.org/10.1080/1047840X.2017.1373550.

Blanton, H., and Jaccard, J. 2023. Listening to measurement error: Lessons from the IAT. In J. A. Krosnick, T. H. Stark, and A. L. Scott, eds., *The Cambridge Handbook of Implicit Bias and Racism*. Cambridge: Cambridge University Press. https://osf.io/ar4u6.

Blázquez, Andrea. 2021. U.S. food retail industry statistics & facts. *Statista*, Sept. 10, 2021. https://www.statista.com/topics/1660/food-retail/.

Bluemke, M., and Fiedler, K. 2009. Base rate effects on the IAT. *Consciousness and Cognition* 18, 4: 1029–38. https://doi.org/10.1016/j.concog.2009.07.010.

BNC (Palestinian BDS National Committee). 2021. EqualHealth Campaign Against Racism Issue Statement of Solidarity and Endorse BDS. *BDS*, May 17, 2021. https://bdsmovement.net/EqualHealth-Campaign-Against-Racism-Issue-Statement-of-Solidarity-Endorse-BDS.

Boeing, H. 2013. Nutritional epidemiology: New perspectives for understanding the diet-disease relationship? *European Journal of Clinical Nutrition* 67, 5: 424–9. https://doi.org/10.1038/ejcn.2013.47.

Boffetta, P., McLaughlin, J. K., Vecchia, C. L., Tarone, R. E., Lipworth, L., and Blot, W. J. 2008. False-positive results in cancer epidemiology: A plea for epistemological modesty. *Journal of The National Cancer Institute* 100: 988–95. https://doi.org/10.1093/jnci/djn191.

Bolland, M., and Grey A. 2014. Rapid Response to: Oral contraceptive use and mortality after 36 years of follow-up in the Nurses' Health Study: prospective cohort study. *BMJ* 349: g6356. https://doi.org/10.1136/bmj.g6356.

Brauer F. 2017. Mathematical epidemiology: Past, present, and future. *Infectious Disease Modelling* 2, 2: 113–27. https://doi.org/10.1016/j.idm.2017.02.001.

Briggs, W. M. 2016. *Uncertainty: The Soul of Modeling, Probability, & Statistics*. New York, NY: Springer.

Briggs, W. M. 2018. Uncertainty: The Soul of Models, Probability & Statistics. Chapter Abstracts. William M. Briggs, March 14, 2018. https://www.wmbriggs.com/post/18724/.

Buchanan, J. M., and Tullock, G. 2004. *The Calculus of Consent: Logical Foundations of Constitutional Democracy*. Indianapolis: Liberty Fund, Inc. http://files.libertyfund.org/files/1063/Buchanan_0102-03_EBk_v6.0.pdf.

Bueno, N. B., de Melo, I. S., de Oliveira, S. L., and da Rocha Ataide, T. 2013. Very-low-carbohydrate ketogenic diet v. low-fat diet for long-term weight loss: A meta-analysis of randomised controlled trials. *British Journal of Nutrition* 110, 7: 1178–87. https://doi.org/10.1017/S0007114513000548.

Byers, T. 1999. Preface. *American Journal of Clinical Nutrition* 69, 6: 1303S. https://doi.org/10.1093/ajcn/69.6.1303S.

Byrnes, G. 2001. Maternal age and risk of type 1 diabetes in children. Flawed analysis invalidates conclusions. *BMJ* 322, 7300: 1489; author reply 1490–1. https://pubmed.ncbi.nlm.nih.gov/11430374/.

Cao, J., Chow, J. C., Lee, F. S. C., and Watson, J. G. 2013. Evolution of $PM_{2.5}$ measurements and standards in the U.S. and future perspectives for China. *Aerosol and Air Quality Research* 13, 4: 1197–211. http://dx.doi.org/10.4209/aaqr.2012.11.0302.

Carlsson, R., and Agerström, J. 2016. A closer look at the discrimination outcomes in the IAT literature. *Scandinavian Journal of Psychology* 57, 4: 278–87. https://doi.org/10.1111/sjop.12288.

Carter, E. C., Schönbrodt, F. D., Gervais, and W. M., Hilgard, J. 2019. Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science* 2, 2: 115–44. https://doi.org/10.1177/2515245919847196.

CASAC (Clean Air Scientific Advisory Committee). 2019. CASAC Review of the EPA's Integrated Science Assessment for Particulate Matter (External Review

Draft—October 2018). https://yosemite.epa.gov/sab/sabproduct.nsf-/LookupW ebReportsLastMonthCASAC/6CBCBBC3025E13B4852583D90047B352/-%24File/ EPA-CASAC-19-002+.pdf.

Castellana, M., Conte, E., Cignarelli, A., Perrini, S., Giustina, A., Giovanella, L., Giorgino, F., and Trimboli, P. 2020. Efficacy and safety of very low calorie ketogenic diet (VLCKD) in patients with overweight and obesity: A systematic review and meta-analysis. *Reviews in Endocrine and Metabolic Disorders* 21, 1: 5–16. https://doi.org/10.1007/s11154-019-09514-y.

Cecil, J. S., and Griffin, E. 1985. The Role of Legal Policies in Data Sharing. In *Sharing Research Data*, eds. Fienberg, S.E., Martin, M. E., Straf, M. L. Washington, DC: National Academy Press. 148–198. https://www.nap.edu/read/2033/chapter/15.

Cecil, J. E., and Barton, K. L. 2020. Inter-individual differences in the nutrition response: From research to recommendations. *The Proceedings of the Nutrition Society* 79, 2: 171–73. https://doi.org/10.1017/S0029665119001198.

CEI (Competitive Enterprise Institute). N.d. Energy and Environment. https://cei. org/issues/energy-and-environment/.

Cesario, J. 2022. So close, yet so far: Stopping short of killing implicit bias. *Psychological Inquiry* 33, 3: 162–6. https://doi.org/10.1080/1047840X.2022.2106753.

CFR (Code of Federal Regulations). 2020. CFR - Code of Federal Regulations Title 21. Revised as of April 21, 2020. https://www.accessdata.fda.gov/scripts/cdrh/cf-docs/cfcfr/cfrsearch.cfm?fr=101.75.

Chambers, C. 2017. *The Seven Deadly Sins of Psychology, A Manifesto for Reforming the Culture of Scientific Practice*. Princeton, NJ: Princeton University Press.

Chappell, B. 2020. WHO Sets 6 Conditions For Ending A Coronavirus Lockdown. NPR, April 15, 2020. https://www.npr.org/sections/goatsandso-da/2020/04/15/834021103/who-sets-6-conditions-for-ending-a-coronavirus-lockdown.

Chawla, D. S. 2020. Russian journals retract more than 800 papers after 'bombshell' investigation. *Science*, January 8, 2020. https://www.sciencemag.org/news/2020/01/russian-journals-retract-more-800-papers-after-bomb-shell-investigation.

Chequer, S., and Quinn, M. G. 2021. More Error than Attitude in Implicit Association Tests (IATs), a CFA-MTMM analysis of measurement error. PsyArXiv Preprints. https://psyarxiv.com/afyz2.

Chin, V., Samia, N. I., Marchant, R., Rosen, O., Ioannidis, J. P. A., Tanner, M. A., and Cripps, S. 2020. A case study in model failure? COVID-19 daily deaths and ICU bed utilisation predictions in New York state. *European Journal of Epidemiology* 35, 8: 733–42. https://doi.org/10.1007/s10654-020-00669-6.

Chin, V., Ioannidis, J. P. A., Tanner, M. A., Cripps, S. 2021. Effect estimates of COVID-19 non-pharmaceutical interventions are non-robust and highly model-dependent. *Journal of Clinical Epidemiology* 136: 96–132. https://doi.org/10.1016/j.jclinepi.2021.03.014.

Chin, J., Holcombe, A., Zeiler, K., Forscher, P., and Guo, A. 2023. Metaresearch, Psychology, and Law: A Case Study on Implicit Bias. Boston University, Law and Psychology Commons. https://scholarship.law.bu.edu/faculty_scholarship/3422/.

Colbourn T. 2020. COVID-19: Extending or relaxing distancing control measures. *The Lancet Public Health* 5, 5: e236–e237. https://doi.org/10.1016/S2468-2667(20)30072-4.

Coleman. L. 2019. How to tackle the unfolding research crisis. *Quillette*, December 14, 2019. https://quillette.com/2019/12/14/how-to-tackle-the-unfolding-research-crisis/.

Collins, G. S., and Wilkinson, J. 2021. Statistical issues in the development of COVID-19 prediction models. *Journal of Medical Virology* 93, 2: 624. https://doi.org/10.1002/jmv.26390.

Cone, J., Mann, T. C., and Ferguson, M. J. 2017. Chapter Three - Changing our implicit minds: How, when, and why implicit evaluations can be rapidly revised. *Advances in Experimental Social Psychology* 56: 131–99. https://doi.org/10.1016/bs.aesp.2017.03.001.

Cooper, R. 2019. Divestment in Fossil Fuels: A Preventive Public Health Strategy. *Psychiatric Times* 36, 4, April 12, 2019. https://www.psychiatrictimes.com/view/divestment-fossil-fuels-preventive-public-health-strategy.

Cordes, C. 1998. Overhead Rates for Federal Research are as High as Ever, Survey Finds. *The Chronicle of Higher Education*, January 23, 1998. https://www.chronicle.com/article/Overhead-Rates-for-Federal/99293.

Corneille, O., and Mertens, G. 2020a. Behavioral and physiological evidence challenges the automatic acquisition of evaluations. *Current Directions in Psychological Science* 29, 6. https://doi.org/10.1177/09637214209641.

Corneille, O., and Hütter, M. 2020b. Implicit? What do you mean? A comprehensive review of the delusive implicitness construct in attitude research. *Personality and Social Psychology Review* 24, 3: 212–32. https://doi.org/10.1177/1088868320911325.

Corneille, O., and Béna, J. 2022. The "Implicit Bias" wording is a Relic. Let's move on and study unconscious social categorization effects. *Psychological Inquiry* 33, 3: 167–72. https://doi.org/10.1080/1047840X.2022.2106754.

Coronado-Montoya, S., Levis, A. W., Kwakkenbos, L., Steele, R. J., Turner, E. H., Thombs, B. D. 2016. Reporting of positive results in randomized controlled trials

of mindfulness-based mental health interventions. *PLoS One* 11, 4. https://doi.org/10.1371/journal.pone.0153220.

Cox, D. D., and Lee, J. S. 2008. Pointwise testing with functional data using the Westfall–Young randomization method. *Biometrika* 95, 3: 621–34. https://doi.org/10.1093/biomet/asn021.

Cox Jr., L. A. [Tony], Popken, D., and Ricci, P. F. 2012. Temperature, not fine particulate Matter (PM2.5), is causally associated with short-term acute daily mortality rates: Results from one hundred United States cities. *Dose-Response* 11, 3: 319–43. https://doi.org/10.2203/dose-response.12-034.Cox.

Cox Jr., L.A. 2017. Do causal concentration–response functions exist? A critical review of associational and causal relations between fine particulate matter and mortality. *Critical Reviews in Toxicology* 47, 7: 609–37. https://doi.org/10.1080/10408444.2017.1311838.

Cox, L. A., Jr, and Popken, D. A. 2020. Should air pollution health effects assumptions be tested? Fine particulate matter and COVID-19 mortality as an example. *Global Epidemiology* 2: 100033. https://doi.org/10.1016/j.gloepi.2020.100033.

Cuff, M. 2016. Shipping industry agrees to cap sulphur emissions by 2020. The Guardian. https://www.theguardian.com/environment/2016/oct/28/shipping-industry-agrees-to-cap-sulphur-emissions-by-2020.

Cyrus-Lai, W., et al. 2022. Avoiding bias in the search for implicit bias. *Psychological Inquiry* 33, 3: 203–12. https://doi.org/10.1080/1047840X.2022.2106762.

Danziger, K. 1990. *Constructing the Subject: Historical Origins of Psychological Research*. Cambridge, MA: Cambridge University Press.

Delgado, J., Ansorena, D., Van Hecke, T., Astiasarán, I., De Smet, S., and Estévez, M. 2021. Meat lipids, NaCl and carnitine: Do they unveil the conundrum of the association between red and processed meat intake and cardiovascular diseases?_Invited Review. *Meat Science* 171: 108278. https://doi.org/10.1016/j.meatsci.2020.108278.

van Dessel, P., Cummins, J., Hughes, S., Kasran, S., Cathelyn, F., and Moran, T. 2020. Reflecting on 25 years of research using implicit measures: Recommendations for their future use. *Social Cognition* 38, Suppl.: S223-S242. https://doi.org/10.1521/soco.2020.38.supp.s223.

Dickersin, K., Chan, S., Chalmers, T. C., Sacks, H. S., and Smith, H., Jr. 1987. Publication bias and clinical trials. *Controlled Clinical Trials* 8, 4: 343–53. https://doi.org/10.1016/0197-2456(87)90155-3.

Dockery, D. W., Pope III, C. A., Xu, X., Spengler, J. D., Ware, J. H., Fay, M. E., Ferris, B. G., and Speizer, F. E. 1993. An association between air pollution and mortality in six

U.S. cities. *New England Journal of Medicine* 329: 1753–9. https://doi.org/10.1056/nejm199312093292401.

D'Souza, M. S., Dong, T. A., Ragazzo, G., Dhindsa, D. S., Mehta, A., Sandesara, P. B., Freeman, A. M., Taub, P., and Sperling, L. S. 2020. From fad to fact: Evaluating the impact of emerging diets on the prevention of cardiovascular disease. *American Journal of Medicine*. 133, 10: 1126–34. https://doi.org/10.1016/j.amjmed.2020.05.017.

Editorial Board (*Wall Street Journal*). 2021. How Fauci and Collins Shut Down Covid Debate. They worked with the media to trash the Great Barrington Declaration. *Wall Street Journal*, December 21, 2021. https://www.wsj.com/articles/fauci-collins-emails-great-barrington-declaration-covid-pandemic-lockdown-11640129116?page=1.

Ekmekcioglu, C., Wallner, P., Kundi, M., Weisz, U., Haas, W., and Hutter, H. P. 2018. Red meat, diseases, and healthy alternatives: A critical review. *Critical Reviews in Food Science and Nutrition* 8, 2: 247–61. https://doi.org/10.1080/10408398.2016.1158148.

El Emam, K., Dankar, F. K., Issa, R., Jonker, E., Amyot, D., Cogo, E., Corriveau, J. P., Walker, M., Chowdhury, S., Vaillancourt, R., Roffey, T., and Bottomley, J. 2009. A globally optimal k-anonymity method for the de-identification of health data. *Journal of the American Medical Informatics Association: JAMIA* 16, 5: 670–82. https://doi.org/10.1197/jamia.M3144.

Ellenberg, J. 2014. *How Not to Be Wrong: The Power of Mathematical Thinking*. New York, NY: Penguin Press.

Engber, D. 2017. Daryl Bem proved ESP Is real. Which means science is broken. *Slate*, June 7, 2017. https://slate.com/health-and-science/2017/06/daryl-bem-proved-esp-is-real-showed-science-is-broken.html.

Enstrom, J. E. 2017. Fine particulate matter and total mortality in Cancer Prevention Study cohort reanalysis. *Dose Response* 15, 1: 1–12. https://doi.org/10.1177/1559325817693345.

EPA (Environmental Protection Agency). 2011. EPA Report Underscores Clean Air Act's Successful Public Health Protections/Landmark law saved 160,000 lives in 2010 alone. United States Environmental Protection Agency, https://archive.epa.gov/epapages/newsroom_archive/newsreleases/f8ad3485e788be5a852578460054049.html.

Fanelli, D. 2018. Opinion: Is science really facing a reproducibility crisis, and do we need it to? *Proceedings of the National Academy of Sciences of the United States of America* 115, 11: 2628–31. https://doi.org/10.1073/pnas.1708272114.

Feinstein, A. R. 1988. Scientifc standards in epidemiologic studies of the menace of daily life. *Science* 242: 1257–63. https://doi.org/10.1126/science.3057627.

Ferguson, N. M., Cummings, D. A. T., Fraser, C., Cajka, J. C., Cooley, P. C., and Burke, D. S. 2006. Strategies for mitigating an influenza pandemic. *Nature* 442, 7101: 448–52. https://www.nature.com/articles/nature04795.

Fiedler, K., Messner, C., and Bluemke, M. 2006. Unresolved problems with the "I", the "A", and the "T": A logical and psychometric critique of the Implicit Association Test (IAT). *European Review of Social Psychology* 17: 74–147. https://doi.org/10.1080/10463280600681248.

Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., and Nosek, B. A. 2019. A meta-analysis of procedures to change implicit measures. *Journal of Personality and Social Psychology* 117, 3: 522–59. https://doi.org/10.1037/pspa0000160.

Franco, A., Malhotra, N., and Simonovits, G. 2014. Publication bias in the social sciences: Unlocking the file drawer. *Science* 345: 1502–5. https://doi.org/10.1126/science.1255484.

Freudenheim, J. L. 1999. Study design and hypothesis testing: issues in the evaluation of evidence from research in nutritional epidemiology. *The American Journal of Clinical Nutrition* 69, 6: 1315S—21S. https://doi.org/10.1093/ajcn/69.6.1315S.

Gal, T. S., Tucker, T. C., Gangopadhyay, A., and Chen, Z. 2014. A data recipient centered de-identification method to retain statistical attributes. *Journal of Biomedical Informatics* 50: 32–45. https://doi.org/10.1016/j.jbi.2014.01.001.

GAO (U.S. Government Accountability Office). 2020. Disease modeling: How Math Can Help In A Pandemic. U.S. Government Accountability Office, June 9, 2020. https://www.gao.gov/blog/disease-modeling-how-math-can-help-pandemic.

Gawronski, B. 2019. Six lessons for a cogent science of implicit bias and its criticism. *Perspectives on Psychological Science* 14, 4. https://doi.org/10.1177/1745691619826015.

Gawronski, B., Ledgerwood, A., and Eastwick, P. W. 2022. Implicit bias≠Bias on implicit measures. *Pyschological Inquiry* 33, 3: 139–55. https://doi.org/10.1080/1047840X.2022.2106750.

Ge, Y., Dudoit, S., Speed, T. P. 2003. Resampling-based multiple testing for microarray data analysis. Technical Report #633: 1–41. https://statistics.berkeley.edu/sites/default/files/tech-reports/633.pdf.

Gelman, A., and Greenland, S. 2019. Are confidence intervals better termed "uncertainty intervals"? *BMJ* 366: I5381. https://pubmed.ncbi.nlm.nih.gov/31506269/.

Gerber, A. S., and Malhotra, N. 2008. Publication bias in empirical sociological research: Do arbitrary significance levels distort published results? *Sociological Methods and Research* 37, 1: 3–30. https://doi.org/10.1177/0049124108318973.

Gershuni, V. M. 2018. Saturated fat: Part of a healthy diet. *Current Nutrition Reports* 7, 3: 85–96. https://doi.org/10.1007/s13668-018-0238-x.

Glaeser, E. L. 2006. Researcher incentives and empirical methods. NBER Technical Working Papers 0329, National Bureau of Economic Research, Inc. https://www.nber.org/papers/t0329.pdf.

Gobry, P.-E. 2016. Big Science is Broken. *The Week*, April 18, 2016. https://theweek.com/articles/618141/big-science-broken.

Gold, M. S. 2020. The role of alcohol, drugs, and deaths of despair in the U.S.'s falling life expectancy. *Missouri Medicine* 117, 2: 99–101. https://www.ncbi.nlm.nih.gov/pubmed/32308224.

Goldacre, M. J. 1993. Cause-specific mortality: Understanding uncertain tips of the disease iceberg. *Journal of Epidemiology and Community Health* 47, 6: 491–6. https://doi.org/10.1136/jech.47.6.491.

Goodman, S. N., Fanelli, D., and Ioannidis, J. P. A. 2016. What does research reproducibility mean? *Science Translational Medicine* 8, 341: 1–6. https://doi.org/10.1126/scitranslmed.aaf5027.

Gotzsche, P. C. 2006. Believability of relative risks and odds ratios in abstracts: Cross sectional study. *BMJ* 333: 231–4, https://doi.org/10.1136/bmj.38895.410451.79.

Greenwald, A. G., and Banaji, M. R. 1995. Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review* 102, 1: 4–27. https://doi.org/10.1037/0033-295x.102.1.4.

Greenwald, A. G., and Krieger, L. H. 2006. Implicit Bias: Scientific Foundations. *California Law Review* 94, 4: 945–67. https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/cccf922b-2a03-441a-940a-90960f3b442c/content.

Greenwald, A. G., Dasgupta, N., Dovidio, J. F., Kang, J., Moss-Racusin, C. A., and Teachman, B. A. 2022. Implicit-bias remedies: Treating discriminatory bias as a public-health problem. *Psychological Science in the Public Interest* 23, 1: 7–40. https://doi.org/10.1177/15291006211070781.

Greven, S., Dominici, F., and Zeger, S. 2011. An approach to the estimation of chronic air pollution effects using spatio-temporal information. *Journal of the American Statistical Association* 106, 494: 396–406. https://doi.org/10.1198/jasa.2011.ap09392.

GS (Google Scholar). 2021. https://scholar.google.com/scholar?cites=16315716240118231868&as_sdt=5,33&sciodt=0,33&hl=en, November 5, 2021.

Gullberg, B., and Ranstam, J. 2009. Flawed analysis of risk factors for coronary heart disease. *Journal of Internal Medicine* 266, 6: 574–5; author reply 576–7. https://doi.org/10.1111/j.1365-2796.2009.02161.x.

Guyatt, G. H., Oxman, A. D., Vist, G. E., et al; GRADE Working Group. 2008. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 336: 924–6. https://doi.org/10.1136/bmj.39489.470347.AD.

Hahn, A., Judd, C. M., Hirsh, H. K., and Blair, I. V. 2014. Awareness of implicit attitudes. *Journal of Experimental Psychology. General* 143, 3: 1369–92. https://doi.org/10.1037/a0035028.

Hahn, A., and Goedderz, A. 2020. Trait-unconsciousness, state-unconsciousness, preconsciousness, and social miscalibration in the context of implicit evaluation. *Social Cognition* 38, Supp.: S114–S134. https://guilfordjournals.com/doi/pdf/10.1521/soco.2020.38.supp.s115.

Halsey, L. G., Curran-Everett, D., Vowler, S. L., and Drummond, G. B. 2015. The fickle P value generates irreproducible results. *Nature Methods* 12, 3: 179–85. https://doi.org/10.1038/nmeth.3288.

Hamblin, J. 2018. A Credibility Crisis in Food Science. *The Atlantic*, September 24, 2018. https://www.theatlantic.com/health/archive/2018/09/what-is-food-science/571105/.

Harris, R. 2017. *Rigor Mortis: How Sloppy Science Creates Worthless Cures, Crushes Hope, and Wastes Billions*. New York, NY: Basic Books.

Hart, J. 2022. The Twitter Blacklisting of Jay Bhattacharya. *The Wall Street Journal*, December 9, 2022. https://www.wsj.com/articles/the-twitter-blacklisting-of-jay-bhattacharya-medical-expert-covid-lockdown-stanford-doctor-shadow-banned-censorship-11670621083.

Head, M. L., Holman L., Lanfear, R., Kahn, A. T., and Jennions, M. D. 2015. The extent and consequences of p-hacking in science. *PLoS Biology* 13, 3: e1002106. https://doi.org/10.1371/journal.pbio.1002106.

Hennen, A. 2019. The Credibility Issue in Nutrition Science is a Sign for All of Higher Ed. *The James G. Martin Center for Academic Renewal*, November 27, 2019. https://www.jamesgmartin.center/2019/11/the-credibility-issue-in-nutrition-science-is-a-sign-for-all-of-higher-ed/.

Henry, P. J. 2021. A survey researcher's response to the Implicit revolution: Listen to what people say. In J. A. Krosnick, T. H. Stark, and A. L. Scott, eds., *The Cambridge Handbook of Implicit Bias and Racism*. Cambridge, MA: Cambridge University Press. https://osf.io/y62ct.

Heritage Foundation. N.d. Science Policy. https://www.heritage.org/science-policy?f%5B0%5D=content_type%3Areport.

Herold, E. 2018. Researchers Behaving Badly: Known Frauds Are "the Tip of the Iceberg." *Leapsmag*. October 19, 2018. https://leapsmag.com/researchers-behaving-badly-why-scientific-misconduct-may-be-on-the-rise/.

Howick, J., Koletsi, D., Ioannidis, J. P. A., et al. 2022. Most healthcare interventions tested in Cochrane Reviews are not effective according to high quality evidence: A systematic review and meta-analysis. *Journal of Clinical Epidemiology* 148: 160–9. https://doi.org/10.1016/j.jclinepi.2022.04.017.

Hubbard, R. 2015. *Corrupt Research: The Case for Reconceptualizing Empirical Management and Social Science*. London, UK: Sage Publications.

Hughes, S., Cummins, J., and Hussey, I. 2023. Effects on the Affect Misattribution Procedure are strongly moderated by influence awareness. *Behavior Research Methods* 55: 1558–86. https://doi.org/10.3758/s13428-022-01879-4.

IMO (International Maritime Organization). 2020. Sulphur 2020—cutting sulphur oxide emissions. IMO, London, UK. http://www.imo.org/en/MediaCentre/HotTopics/Pages/Sulphur-2020.aspx.

INFL (Interactive Nutrition Facts Label). N.d. Interactive Nutrition Facts Label. U.S. Food & Drug Administration. https://www.accessdata.fda.gov/scripts/interactivenutritionfactslabel/saturated-fat.cfm.

Ioannidis, J. P. A. 2005. Why most published research findings are false. *PLoS Medicine* 2, 8: e124. https://doi.org/10.1371/journal.pmed.0020124.

Ioannidis, J. P., Tarone, R., and McLaughlin, J. K. 2011. The false-positive to false-negative ratio in epidemiologic studies. *Epidemiology* 22, 4: 450–6. https://doi.org/10.1097/EDE.0b013e31821b506e.

Ioannidis, J. P. A. 2021. Precision shielding for COVID-19: Metrics of assessment and feasibility of deployment. *BMJ Global Health* 6, 1: e004614. https://doi.org/10.1136/bmjgh-2020-004614.

Ioannidis, J. P. A., Cripps, S., and Tanner, M. A. 2022a. Forecasting for COVID-19 has failed. *International Journal of Forecasting* 38, 2: 423–38. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7447267/.

Ioannidis, J. P. A. 2022b. Pre-registration of mathematical models. *Mathematical Biosciences* 345: 108782. https://doi.org/10.1016/j.mbs.2022.108782.

IQA (Information Quality Act). 2000. Sec. 515, Treasury and General Government Appropriations Act for Fiscal Year 2001 (Public Law 106-554), https://www.fws.gov/informationquality/section515.html.

Janis, L. 1982. *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*. Boston: Houghton Mifflin.

Jones, D., Molitor, D., and Reif, J. 2019a. What do workplace wellness programs do? Evidence from the Illinois Workplace Wellness Study. *The Quarterly Journal of Economics* 134, 4: 1747–91. https://doi.org/10.1093/qje/qjz023.

Jones, D., Molitor, D., Reif, J. 2019b. Documentation for Illinois Workplace Wellness Study. https://www.nber.org/workplacewellness/s/wyoung.pdf.

Junod, S. W. 2008. "FDA and Clinical Drug Trials: A Short History." In *A Quick Guide to Clinical Trials*, Madhu Davies and Faiz Kerimani, eds. Washington: Bioplan, Inc. 25–55. https://www.fda.gov/media/110437/download.

Jussim, L., Stevens, S. T., and Honeycutt, N. 2018. Unasked questions about stereotype accuracy. *Archives of Scientific Psychology* 6, 1: 214–29. https://doi.org/10.1037/arc0000055.

Jussim, L., Careem, A., Goldberg, Z. Honeycutt, N., and Stevens, S. T. 2020a. IAT Scores, Racial Gaps, and Scientific Gaps. In J. A. Krosnick, Stark, T. H., and Scott, A. L., eds., *The Future of Research on Implicit Bias* (Cambridge: Cambridge University Press, forthcoming). https://osf.io/4nhdm.

Jussim, L. 2020b. Implicit Bias: Racial Gaps and Scientific Gaps. OSFHome. https://osf.io/vmd38.

Jussim, L., Thulin, E., Fish, J., and Wright, J. D. 2023. Articles Critical of the IAT and Implicit Bias. OFSHome. https://osf.io/74whk/.

Kaiser, J. 2017. NIH plan to reduce overhead payments draws fire. *Science*, June 2, 2017. https://www.sciencemag.org/news/2017/06/nih-plan-reduce-overhead-payments-draws-fire.

Kavanaugh, C. J., Trumbo, P. R., and Ellwood, K. C. 2007. The U.S. Food and Drug Administration's evidence-based review for qualified health claims: Tomatoes, lycopene, and cancer. *Journal of the National Cancer Institute* 99, 14: 1074–85. https://doi.org/10.1093/jnci/djm037.

Kindzierski, W., Young, S., Meyer, T., and Dunn, J. 2021. Evaluation of a meta-analysis of ambient air quality as a risk factor for asthma exacerbation. *Journal of Respiration* 1, 3: 173–96. https://doi.org/10.3390/jor1030017.

Kmietowicz, Z. 2014. Study claiming Tamiflu saved lives was based on "flawed" analysis. *BMJ* 348: g2228. https://doi.org/10.1136/bmj.g2228.

Kretzschmar, M., and Wallinga, J. 2009. Mathematical Models in Infectious Disease Epidemiology. In Krämer, A., Kretzschmar, M., and Krickeberg, K., eds., *Modern Infectious Disease Epidemiology. Statistics for Biology and Health*. New York, NY: Springer. https://doi.org/10.1007/978-0-387-93835-6_12.

Kristal, A. R., Peters, U., Potter, J. D. 2005. Is it time to abandon the food frequency questionnaire? *Cancer Epidemiology, Biomarkers & Prevention* 14: 2826–8. https://doi.org/10.1158/1055-9965.EPI-12-ED1.

Kühberger, A., Fritz, A., and Scherndl, T. 2014. Publication bias in psychology: A diagnosis based on the correlation between effect size and sample size. *PLoS One* 9, 9: e105825. https://doi.org/10.1371/journal.pone.0105825.

Kulldorff, M., Gupta, S., and Bhattacharya, J. 2020. Great Barrington Declaration. https://gbdeclaration.org/.

Kushida, C. A., Nichols, D. A., Jadrnicek, R., Miller, R., Walsh, J. K., and Griffin, K. 2012. Strategies for de-identification and anonymization of electronic health record data for use in multicenter research studies. *Medical Care* 50 Suppl: S82–S101. https://doi.org/10.1097/MLR.0b013e3182585355.

Lai, C. K., and Wilson, M. E. 2021. Measuring implicit intergroup biases. *Social and Personality Psychology Compass* 15, 1, Article e12573. https://doi.org/10.1111/spc3.12573.

Lai, C. K., and Lisnek, J. A. 2023. The Impact of Implicit Bias-Oriented Diversity Training on Police Officers' Beliefs, Motivations, and Actions. *Psychological Science*. https://osf.io/74whk/.

Lamiell, J. T. 2019. *Psychology's Misuse of Statistics and Persistent Dismissal of Its Critics*. London, UK: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-12131-0.

LeBel, E. P., and Paunonen, S. V. 2011. Sexy but often unreliable: The impact of unreliability on the replicability of experimental findings with implicit measures. *Personality & Social Psychology Bulletin* 37, 4: 570–83. https://doi.org/10.1177/0146167211400619.

Levitt, M., Zonta, F., and Ioannidis, J. P. A. 2022a. Comparison of pandemic excess mortality in 2020–2021 across different empirical calculations. *Environmental Research* 213: 113754. https://doi.org/10.1016/j.envres.2022.113754.

Levitt, M., Zonta, F., and Ioannidis, J. P. A. 2022b. Excess death estimates from multiverse analysis in 2009–2021. *medRxiv*, 2022.09.21.22280219. https://doi.org/10.1101/2022.09.21.22280219.

Lilienfeld, S. O. 2017. Psychology's replication crisis and the grant culture: Righting the ship. *Perspectives on Psychological Science* 12, 4: 660–64. https://doi.org/10.1177/1745691616687745.

Liu, K. 1994. Statistical issues related to semiquantitative food-frequency questionnaires. *The American Journal of Clinical Nutrition* 59, 1 Suppl: 262S-265S. https://doi.org/10.1093/ajcn/59.1.262S.

Machery, E. 2022a. Anomalies in implicit attitudes research. *Wiley Interdisciplinary Reviews. Cognitive Science* 13, 1, e1569. https://doi.org/10.1002/wcs.1569.

Machery, E. 2022b. Anomalies in implicit attitudes research: Not so easily dismissed. *Wiley Interdisciplinary Reviews. Cognitive Science* 13, 3, e1591. https://doi.org/10.1002/wcs.1591.

Manuel, T. 2019. Why the way we use statistical significance has created a crisis in science. *Science: The Wire*, March 31, 2019. https://science.thewire.in/the-sciences/why-the-way-we-use-statistical-significance-has-created-a-crisis-in-science/.

Marks, J. H. 2011. On regularity and regulation, health claims and hype. *Hastings Center Report* 41, 4: 11–12. https://doi.org/10.1002/j.1552-146x.2011.tb00113.x.

Martino, J. P. 2017. *Science Funding: Politics and Porkbarrel*. New York, NY: Routledge.

McCambridge, J. 2007. A case study of publication bias in an influential series of reviews of drug education. *Drug and Alcohol Review* 26, 5: 463–8. https://doi.org/10.1080/09595230701494366.

McLaughlin, J. K., and Tarone, R. E. 2013. False positives in cancer epidemiology. *Cancer Epidemiology, Biomarkers and Prevention* 22, 1: 11–15. https://doi.org/10.1158/1055-9965.EPI-12-0995.

Meinshausen, N. Maathuis, M. H., and Bühlmann, P. 2011. Asymptotic optimality of the Westfall--Young permutation procedure for multiple testing under dependence. *Annals of Statistics* 39, 6: 3369–91. https://projecteuclid.org/euclid.aos/1330958683.

Meissner, F., Grigutsch, L. A., Koranyi, N., Müller, F., and Rothermund, K. 2019. Predicting behavior with implicit measures: Disillusioning findings, reasonable explanations, and sophisticated solutions. *Frontiers in Psychology* 10, 2483. https://doi.org/10.3389/fpsyg.2019.02483.

Melnick, E. R., and Ioannidis, J. P. A. 2020. Should governments continue lockdown to slow the spread of covid-19? *BMJ (Clinical Research Ed.)* 369: m1924. https://doi.org/10.1136/bmj.m1924.

Michaels, P. J. 2008. Evidence for "publication bias" concerning global warming in *Science* and *Nature*. *Energy & Environment* 19, 2: 287–301. https://doi.org/10.1260/095830508783900735.

Milloy, S. J. 2016. *Scare Pollution: Why and How to Fix the EPA*. USA: Bench Press.

Milojevic, A., Wilkinson, P., Armstrong, B., Bhaskaran, K., Smeeth, L., and Hajat, S. 2014. Short-term effects of air pollution on a range of cardiovascular events in England and Wales: Case-crossover analysis of the MINAP database, hospital admissions and mortality. *Heart (British Cardiac Society)* 100, 14: 1093–8. https://doi.org/10.1136/heartjnl-2013-304963.

Mitchell, G., and Tetlock, P. E. 2017. Popularity as a poor proxy for utility: The case of implicit prejudice. In S. O. Lilienfeld and I. D. Waldman, eds., *Psychological Science under Scrutiny: Recent Challenges and Proposed Solutions* (pp. 164–195). Wiley Blackwell. https://doi.org/10.1002/9781119095910.ch10.

Mosher, S. W. 2022. Government censorship should scare us just as much as COVID once did. *New York Post*, September 17. 2022. https://nypost.com/2022/09/17/government-censorship-should-scare-us-as-much-as-covid-did/.

Mousavi, A., Yuan, Y., Masri, S., Barta, G., and Wu, J. 2021. Impact of 4th of July fireworks on spatiotemporal PM2.5 concentrations in California based on the PurpleAir Sensor Network: Implications for policy and environmental justice.

International Journal of Environmental Research and Public Health 18, 11: 5735. https://doi.org/10.3390/ijerph18115735.

NASEM (National Academies of Sciences, Engineering, and Medicine). 2016. *Statistical Challenges in Assessing and Fostering the Reproducibility of Scientific Results: Summary of a Workshop*. Washington, DC: The National Academies Press. https://www.nap.edu/read/21915/.

NASEM (National Academies of Science, Engineering, and Medicine). 2019. *Reproducibility and Replicability in Science*. Washington, D.C.: The National Academies Press. https://www.nap.edu/read/25303/.

Nelson, L. D., Simmons, J., and Simonsohn, U. 2018. Psychology's Renaissance. *Annual Review of Psychology* 69: 511–34. https://doi.org/10.1146/annurev-psych-122216-011836.

Nelson, A. 2022. Doctors slam COVID government censorship exposed in 'Twitter Files': 'On the road to totalitarianism.' *Fox News*, December 27, 2022. https://www.foxnews.com/media/doctors-slam-covid-government-censorship-exposed-twitter-files-road-totalitarianism.

Nemery, B., Hoet, P. H. M., and Nemmar, A. 2001. The Meuse Valley fog of 1930: An air pollution disaster. *Lancet* 357, 9257: 704–8. https://doi.org/10.1016/s0140-6736(00)04135-0.

Nissen, S. B., Magidson, T., Gross, K., and Bergstrom, C. T. 2016. Publication bias and the canonization of false facts. *eLife* 5: e21451. https://doi.org/10.7554/elife.21451.

Nixon, K., Jindal, S., Parker, F., et al. 2022. An evaluation of prospective COVID-19 modelling studies in the USA: from data to science translation. *The Lancet. Digital Health* 4, 10: e738–e747. https://doi.org/10.1016/S2589-7500(22)00148-0.

Nosek, B., and Errington, T. M. 2020. What is replication? *PloS Biology* 18, 3: e3000691. https://doi.org/10.1371/journal.pbio.3000691.

NSF. N.d. Graduate Research Fellowship Program (GRFP), National Science Foundation, https://beta.nsf.gov/funding/opportunities/nsf-graduate-research-fellowship-program-grfp.

Nuttgens, S. 2023. Making psychology "count": On the mathematization of psychology. *Europe's Journal of Psychology*: 19, 1: 100–12. https://doi.org/10.5964/ejop.4065.

Offen, N., Smith, E. A., and Malone, R. E. 2005. The perimetric boycott: A tool for tobacco control advocacy. *Tobacco Control* 14, 4: 272–7. https://doi.org/10.1136/tc.2005.011247.

Olson, C.M., Rennie, D., Cook, D., Dickersin, K., Flanagin, A., Hogan, J. W., Zhu, Q., Reiling, J., and Pace, B. 2002. Publication bias in editorial decision making. *Journal of the American Medical Association* 287, 21: 2825–8. https://doi.org/10.1001/jama.287.21.2825.

OMB (Office of Management and Budget). 2019. Improving Implementation of the Information Quality Act. https://www.whitehouse.gov/wp-content/uploads/2019/04/M-19-15.pdf.

Orellano, P., Reynoso, J., Quaranta, N., Bardach, A., and Ciapponi, A. 2020. Short-term exposure to particulate matter (PM_{10} and $PM_{2.5}$), nitrogen dioxide (NO_2), and ozone (O_3) and all-cause and cause-specific mortality: Systematic review and *meta*-analysis. *Environment International* 142: 105676. https://doi.org/10.1016/j.envint.2020.105876.

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., and Tetlock, P. E. 2013. Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology* 105, 2: 171–92. https://doi.org/10.1037/a0032734.

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., and Tetlock, P. E. 2015. Using the IAT to predict ethnic and racial discrimination: small effect sizes of unknown societal significance. *Journal of Personality and Social Psychology* 108, 4: 562–71. https://doi.org/10.1037/pspa0000023.

Pachetti, M., Marini, B., Giudici, F., et al. 2020. Impact of lockdown on Covid-19 case fatality rate and viral mutations spread in 7 countries in Europe and North America. *Journal of Translational Medicine* 18, 1: 338. https://doi.org/10.1186/s12967-020-02501-x.

Paluck, E. L., Porat, R., Clark, C. S., and Green, D. P. 2021. Prejudice reduction: Progress and challenges. *Annual Review of Psychology* 72: 533–60. https://doi.org/10.1146/annurev-psych-071620-030619.

Paris, Costas. 2020. Report Puts $1 Trillion Price Tag on Cutting Ship Carbon Emissions. *The Wall Street Journal*, January 21, 2020. https://www.wsj.com/articles/report-puts-1-trillion-price-tag-on-cutting-ship-carbon-emissions-11579627855.

Payne, K., Niemi, L., and Doris, J. M. 2018. How to think about 'Implicit Bias'. *Scientific American*, March 27, 2018. https://www.scientificamerican.com/article/how-to-think-about-implicit-bias/.

Pellizzari, E., Lohr, K. Blatecky, A., and Creel, D. 2017. *Reproducibility: A Primer on Semantics and Implications for Research*. Research Triangle Park, NC: RTI Press. https://www.rti.org/sites/default/files/resources/18127052_Reproducibility_Primer.pdf.

Peretti, J. 2013. Food Giants Making Fat Profits. *Independent*, August 19, 2013. https://www.independent.ie/life/health-wellbeing/fitness/food-giants-making-fat-prof-its-29509349.html.

Pope III, C. A., Thun, M. J., Namboodiri, M. M., Dockery, D. W., Evans, J. S., Speizer, F. E., and Heath Jr., C. W. 1995. Particulate air pollution as a predictor of mortality

in a prospective study of U.S. adults. *American Journal of Respiratory and Critical Care Medicine* 151, 3, Pt. 1: 669–74. https://doi.org/10.1164/ajrccm/151.3_Pt_1.669.

Potischman, N., and Weed, D. L. 1999. Causal criteria in nutritional epidemiology. *The American Journal of Clinical Nutrition* 69, 6: 1309S–1314S. https://doi.org/10.1093/ajcn/69.6.1309S.

Prem, K., Liu, Y., Russell, T. W., et al., Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group. 2020. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: A modelling study. *The Lancet. Public Health* 5, 5: e261–e270. https://doi.org/10.1016/S2468-2667(20)30073-6.

Prentice, R. L. 2010. Dietary assessment and the reliability of nutritional epidemiology research reports. *Journal of the National Cancer Institute* 102, 9: 583–5. https://doi.org/10.1093/jnci/djq100.

Randall, D., and Welser, C. 2018. *The Irreproducibility Crisis of Modern Science: Causes, Consequences, and the Road to Reform*. New York: National Association of Scholars. https://www.nas.org/reports/the-irreproducibility-crisis-of-modern-science.

Randall, D. 2020. Regulatory Science and the Irreproducibility Crisis. Fixing Science Conference, February 7-8, 2020, Independent Institute, Oakland, California. https:// www.youtube.com/watch?v=p6ysi65ekSA.

van Ravenzwaaij, D., van der Maas, H. L., and Wagenmakers, E. J. 2011. Does the name-race implicit association test measure racial prejudice? *Experimental Psychology* 58, 4: 271–7. https://doi.org/10.1027/1618-3169/a000093.

Reber, A. S. 1989. Implicit learning of tacit knowledge. *Journal of Experimental Psychology. General* 118, 3: 219–35. https://doi.org/10.1037/0096-3445.118.3.219.

Rezaei, A. R. 2011. Validity and reliability of the IAT: Measuring gender and ethnic stereotypes. *Computers in Human Behavior* 27, 5: 1937–41. https://doi.org/10.1016/j.chb.2011.04.018.

Ritchie, S. 2020. *Science Fictions: How Fraud, Bias, Negligence, and Hype Undermine the Search for Truth*. New York, NY: Henry Holt and Company.

Roche, G. C. 1994. *The Fall of the Ivory Tower: Government Funding, Corruption, and the Bankrupting of American Higher Education*. Washington, D.C.: Regnery.

Romano, J. P., and Wolf, M. 2016. Efficient Computation of Adjusted p-Values for Resampling-Based Stepdown Testing. University of Zurich, Department of Economics, Working Paper Series, Working Paper No. 219. http://www.econ.uzh.ch/static/wp/econwp219.pdf.

Rothman, K. J. 1990. No adjustments are needed for multiple comparisons. *Epidemiology* 1, 1: 43–46. https://www.jstor.org/stable/pdf/20065622.pdf?seq=1.

Rubinstein, R. S., Jussim, L., and Stevens, S. T. 2018. Reliance on individuating information and stereotypes in implicit and explicit person perception. *Journal of Experimental Social Psychology* 75: 54–70. https://doi.org/10.1016/j.jesp.2017.11.009.

Ruxton, C. H. 2016. Food science and food ingredients: the need for reliable scientific approaches and correct communication, Florence, 24 March 2015. *International Journal of Food Sciences and Nutrition* 67, 1: 1–8. https://doi.org/10.3109/09637486.2015.1126567.

Samet, J. M. 2019. Current Knowledge on Adverse Effects of Low-Level Air Pollution: Have We Filled the Gap? *Health Effects Institute Annual Meeting Session.* Seattle, WA. https://www.healtheffects.org/sites/default/files/Samet-low-levels-HEI-2019_0.pdf.

Sarewitz, D. 2012. Beware the creeping cracks of bias. *Nature* 485: 149. https://doi.org/10.1038/485149a.

Satija, A., Yu, E., Willett, W. C., and Hu, F. B. 2015. Understanding nutritional epidemiology and its role in policy. *Advances in Nutrition* 6, 1: 5–18. https://doi.org/10.3945/an.114.007492.

Schimmack, U. 2021. The Implicit Association Test: A method in search of a construct. *Perspectives on Psychological Science* 16, 2: 396–414. https://doi.org/10.1177/1745691619863798.

Schneeman, B. 2007. FDA's review of scientific evidence for health claims. *The Journal of Nutrition* 137, 2: 493–4. https://doi.org/10.1093/jn/137.2.493.

Schoenfeld, J. D., and Ioannidis, J. P. 2013. Is everything we eat associated with cancer? A systematic cookbook review. *The American Journal of Clinical Nutrition* 97, 1: 127–34. https://doi.org/10.3945/ajcn.112.047142.

Schutz, Y., Montani, J. P., and Dulloo, A. G. 2021. Low-carbohydrate ketogenic diets in body weight control: A recurrent plaguing issue of fad diets? *Obesity Reviews* 22, Suppl 2: e13195. https://doi.org/10.1111/obr.13195.

Sempos, C. T., Liu, K., and Ernst, N. D. 1999. Food and nutrient exposures: What to consider when evaluating epidemiologic evidence. *The American Journal of Clinical Nutrition* 69, 6: 1330S—38S. https://doi.org/10.1093/ajcn/69.6.1330S.

Shim, J. S., Oh, K., and Kim, H. C. 2014. Dietary assessment methods in epidemiologic studies. *Epidemiology and Health* 36, e2014009. https://doi.org/10.4178/epih/e2014009.

Sieber, W. K., Jr, Green, T., Williamson, G. D., and Centers for Disease Control and Prevention. 2006. Statistics and public health at CDC. *MMWR supplements* 55, 2: 22–24. https://www.cdc.gov/mmwr/preview/mmwrhtml/

su5502a9.htm#:~:text=Use%20of%20multisource%20data%20and,health%20findings%20to%20the%20nation.

Simonsohn, U., Nelson, L. D., and Simmons, J. P. 2014. P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General* 143, 2: 534–47. https://doi.org/10.1037/a0033242.

Skov, T. 2020. Unconscious gender bias in academia: Scarcity of empirical evidence. *Societies* 10, 2: 31. https://doi.org/10.3390/soc10020031.

Smedslund, J. 2021. From statistics to trust: Psychology in transition. *New Ideas in Psychology* 61, 100848. https://doi.org/10.1016/j.newideapsych.2020.100848.

Staddon, J. 2019. Object of inquiry: Psychology's other (non-replication) problem. *Academic Questions* 32, 2: 246–56. https://dukespace.lib.duke.edu/server/api/core/bitstreams/6892c41a-655f-472e-96ea-75b9e40a5663/content.

Stigler, S. M. 1992. A historical view of statistical concepts in psychology and educational research. *American Journal of Education* 101, 1: 60–70. http://www.medicine.mcgill.ca/epidemiology/hanley/bios601/Surveys/StiglerPsychStats.pdf.

Støvring, H., Hansen, D. G., Jarlbaek, L., Kildemoes, H. W., Lous, J., and Andersen, M. 2007. Statin use and age at death: Evidence of a flawed analysis. *American Journal of Cardiology* 99, 8: 1181–2; author reply 1182. https://doi.org/10.1016/j.amjcard.2007.01.003.

Streiner, D. L. 2018. Statistics commentary series, commentary no. 27: P-hacking. *Journal of Clinical Psychopharmacology* 38: 286–8. https://doi.org/10.1097/JCP.0000000000000901.

Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., Moher, D., Becker, B. J., Sipe, T. A., and Thacker, S. B. 2000. Meta-analysis of observational studies in epidemiology: A proposal for reporting. *Journal of the American Medical Association* 283, 15: 2008–12. https://doi.org/10.1001/jama.283.15.2008.

Stroup, D. F., Lyerla, R., and Centers for Disease Control and Prevention (CDC). 2011. History of statistics in public health at CDC, 1960–2010: the rise of statistical evidence. *MMWR supplements* 60, 4: 35–41. https://www.cdc.gov/mmwr/preview/mmwrhtml/su6004a7.htm.

Sulzer, S. H. 2022. Implicit Bias and Public Health Law. The Network for Public Health Law, January 11, 2022. https://www.networkforphl.org/news-insights/implicit-bias-and-public-health-law/.

Tapaninen, U. 2020. The Environmental Impact of Maritime Transport (and How to Combat Emissions). Kogan Page, London, UK. https://www.koganpage.com/article/environmental-impact-of-maritime-transport.

Taubes. G. 2021. The Keto Way: What If Meat Is Our Healthiest Diet? *The Wall Street Journal*. https://www.wsj.com/articles/

the-keto-way-what-if-meat-is-our-healthiest-diet-11611935911 (accessed August 4, 2021).

Terris, M. 2011. A social policy for health. *American Journal of Public Health* 101, 2: 250–52. https://doi.org/10.2105/ajph.101.2.250.

Unkelbach, C., and Fiedler, K. 2020. The challenge of diagnostic inferences from implicit measures: The case of non-evaluative influences in the evaluative priming paradigm. *Social Cognition* 38, Suppl.: S208–S222. https://doi.org/10.1521/soco.2020.38.supp.s208.

USDE (U.S. Department of Education). 2014. Guiding Principles: A Resource Guide for Improving School Climate and Discipline. https://files.eric.ed.gov/fulltext/ED544743.pdf.

U.S. Food and Drug Administration (FDA). 2017. Food Labeling: Health Claims; Soy Protein and Coronary Heart Disease. Docket No. FDA-2017-N-0763. https://www.fda.gov/media/108701/download.

U. S. Food and Drug Administration Guidance Documents. 2006. Guidance for Industry: Estimating Dietary Intake of Substances in Food. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-industry-estimating-dietary-intake-substances-food.

U.S. Food and Drug Administration Guidance Documents. 2009. Guidance for Industry: Evidence-Based Review System for the Scientific Evaluation of Health Claims. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-industry-evidence-based-review-system-scientific-evaluation-health-claims.

U.S. Food and Drug Administration Guidance Documents. 2017. Multiple Endpoints in Clinical Trials Guidance for Industry. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/multiple-endpoints-clinical-trials-guidance-industry.

Van Der Laan, M., Malani, A., and Van Der Benbom, O. 2011. Improving the FDA Approval Process. University of Chicago Public Law & Legal Theory Working Paper No. 367. https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1148&context=public_law_and_legal_theory.

Vankov, I., Bowers, J., and Munafò, M. R. 2014. On the persistence of low power in psychological science. *Quarterly Journal of Experimental Psychology (2006)*: 67, 5: 1037–40. https://doi.org/10.1080/17470218.2014.885986.

Verity, R., Okell, L. C., Dorigatti, I., et al. 2020. Estimates of the severity of coronavirus disease 2019: A model-based analysis. *The Lancet: Infectious Diseases* 20, 6: P669–P677. https://doi.org/10.1016/S1473-3099(20)30243-7.

Vernooij, R. W. M., Zeraatkar, D., Han, M. A., El Dib, R., Zworth, M., Milio, K., Sit, D., Lee, Y., Gomaa, H., Valli, C., Swierz, M. J., Chang, Y., Hanna, S. E., Brauer, P. M., Sievenpiper, J., de Souza, R., Alonso-Coello, P., Bala, M. M., Guyatt, G. H., and Johnston, B. C. 2019. Patterns of red and processed meat consumption and risk for cardiometabolic and cancer outcomes: A systematic review and meta-analysis of cohort studies. *Annals of Internal Medicine* 171: 732–4. https://doi.org/10.7326/M19-1583.

Vishwamitra, N., et al. 2021. On Analyzing COVID-19-related Hate Speech Using BERT Attention. 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA). 669–76. https://par.nsf.gov/servlets/purl/10223825.

Waldman, S. 2019. "U.S. think tank shuts down prominent center that challenged climate science." *Science*, May 29, 2019. https://www.science.org/content/article/us-think-tank-shuts-down-prominent-center-challenged-climate-science.

Walker, P. G. T., Whittaker, C., Watson, O. J., et al. 2020. The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. *Science* 369, 6502: 413–22. https://doi.org/10.1126/science.abc0035.

Westfall, P. H., and Young, S. S. 1993. *Resampling-Based Multiple Testing*. New York, NY: John Wiley & Sons.

WIPR (World Intellectual Property Review). 2020. Hidden depths: The problem of unconscious bias. Sterne Kessler. November 4, 2020. https://www.sternekessler.com/news-insights/news/hidden-depths-problem-unconscious-bias.

Wittkowski, K. M. 2020. The first three months of the COVID-19 epidemic: Epidemiological evidence for two separate strains of SARS-CoV-2 viruses spreading and implications for prevention strategies. *medRxiv*. https://doi.org/10.1101/2020.03.28.20036715.

Woolf, S. H., and Schoomaker, H. 2019. Life expectancy and mortality rates in the United States, 1959–2017. *Journal of the American Medical Association* 322, 20: 1996–2016. https://doi.org/10.1001/jama.2019.16932.

World Health Organization (WHO). 2015. IARC Monographs evaluate consumption of red meat and processed meat. Press release No. 240. https://www.iarc.who.int/wp-content/uploads/2018/07/pr240_E.pdf (accessed August 4, 2021).

Yong, E. 2018. Psychology's replication crisis is running out of excuses. *The Atlantic*, November 19, 2018. https://www.theatlantic.com/science/archive/2018/11/psychologys-replication-crisis-real/576223/.

Young, S. S., and Miller, H. 2018. Junk Science Has Become a Profitable Industry. Who Will Stop It? *Real Clear Science*, November 26, 2018. https://www.realclearscience.com/articles/2018/11/26/junk_science_has_become_a_profitable_industry_110810.html.

Young, S. S., Smith, R. L., and Lopiano, K. K. 2017. Air quality and acute deaths in California, 2000-2012. *Regulatory Toxicology and Pharmacology* 88: 173–84. https://doi.org/10.1016/j.yrtph.2017.06.003.

Young, S. S., and Kindzierski, W. B. 2019a. Evaluation of a meta-analysis of air quality and heart attacks, a case study. *Critical Reviews in Toxicology* 49, 1: 85–94. https://doi.org/10.1080/10408444.2019.1576587.

Young, S. S., Acharjee, M. K., and Das, K. 2019b. The reliability of an environmental epidemiology meta-analysis, a case study. *Regulatory Toxicology and Pharmacology* 102: 47–52. https://doi.org/10.1016/j.yrtph.2018.12.013.

Young, S. S., and Kindzierski, W. B. 2020. Particulate Matter Exposure and Lung Cancer: A Review of two Meta-Analysis Studies. *arXiv*. https://arxiv.org/abs/2011.02399.

Young, S. S., Kindzierski, W. B., and Randall, D. 2021. *Shifting Sands, Unsound Science and Unsafe Regulation Report 1. Keeping Count of Government Science: P-Value Plotting, P-Hacking, and PM2.5 Regulation*. New York, NY: National Association of Scholars. https://www.nas.org/reports/shifting-sands-report-i.

Young, S. S., Kindzierski, W. B., and Randall, D. 2022a. *Shifting Sands, Unsound Science and Unsafe Regulation Report 2. Flimsy Food Findings: Food Frequency Questionnaires, False Positives, and Fallacious Procedures in Nutritional Epidemiology*. New York, NY: National Association of Scholars. https://www.nas.org/reports/shifting-sands-report-ii.

Young, S. S., Cheng, K.-C., Chen, J. H., Chen, S.-C., Kindzierski, W. B. 2022b. Reliability of a meta-analysis of air quality–asthma cohort studies. *International Journal of Statistics and Probability*, 11(2): 61–76. https://doi.org/10.5539/ijspv11n2p61.

Young, S. S., Kindzierski, W., and Randall, D. 2023. *Shifting Sands: Report III. The Confounded Errors of Public Health Policy Response to the COVID-19 Pandemic*. New York, NY: National Association of Scholars. https://www.nas.org/reports/shifting-sands-report-iii/full-report.

Young, S. S., Kindzierski, W., and Randall, D. 2024. *Shifting Sands: Report IV: Zombie Psychology, Implicit Bias Theory, and the Implicit Association Test*. https://www.nas.org/reports/shifting-sands-report-iv/full-report.

Zavalis, E. A., and Ioannidis, J. P. A. 2022. A meta-epidemiological assessment of transparency indicators of infectious disease models. *PLOS One* 17, 10: e0275380. https://doi.org/10.1371/journal.pone.0275380.

Zeeman, E. C. 1976. Catastrophe theory. *Scientific American* 234, 4: 65–83. https://doi.org/10.1038/scientificamerican0476-65.

Zimring, J. C. 2019. *What Science Is and How It Really Works*. Cambridge, MA: Cambridge University Press.