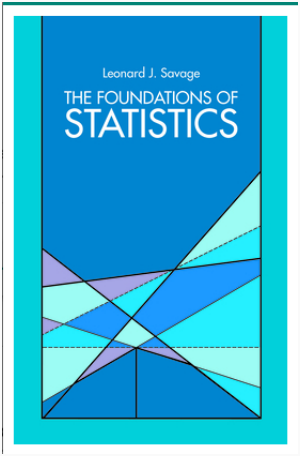
Theoretical statistics in practice

Nancy Reid
University of Toronto
March 8 2024

Outline

1. Theory and Foundations
2. Applications
3. Back to Theory

Theory and Foundations



It is unanimously agreed that statistics depends somehow on probability. But, as to what probability is and how it is connected with statistics, there has seldom been such complete disagreement and breakdown of communication since the Tower of Babel. Doubtless, much of the disagreement is merely terminological and would disappear under sufficiently sharp analysis.

- Statistics needs a healthy interplay between theory and applications
- theory meaning foundations, rather than theoretical analysis of specific techniques
- must be continually tested against new applications
- "the practical application of general theorems is a different art from their establishment by mathematical proof" Fisher 1958 SMRW

What are the foundations?

- probability, analysis, applied mathematics — modelling
- Bayes, Neyman, Fisher — approaches to inference
- nature of uncertainty — epistemic, empirical
- nature of induction — belief functions, inferential models
- interpretation of *p*-values, confidence regions, credibility intervals
- role of sufficiency, ancillarity, conditioning, asymptotic theory
- sparsity, causality, assumption-free/lean inference, stability, prediction, decisions

What use are foundations?

- provide a rigorous basis for the development of techniques
- provide a common language for particular classes of problems
- help to clarify the nature of uncertainty in scientific conclusions
- highlight aspects of data analysis which are likely to raise difficult issues
- suggest strategies for tackling highly complex problems
- avoid ‘re-inventing the wheel’ for each new application

Applications

Climate Change Made East Africa's Drought 100 Times as Likely, Study Says

The findings starkly show the misery that the burning of fossil fuels, mostly by rich countries, inflicts on societies that emit almost nothing by comparison.

NY Times, April 27 2023

A water well near the town of Kelafo in Ethiopia, one of the nations hit hardest by the

U Ottawa March 8 2024

Human-induced climate change increased drought severity in Horn of Africa

World Weather Attribution, April 27

1. Joyce Kimutai, *Kenya Meteorological Department, Nairobi, Kenya*
2. Clair Barnes, *Grantham Institute, Imperial College London, UK*
3. Mariam Zachariah, *Grantham Institute, Imperial College, London, UK*
4. Sjoukje Philip, *Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands*
5. Sarah Kew, *Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands*
6. Izidine Pinto, *Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands*
7. Piotr Wolski, *Climate System Analysis Group, University of Cape Town, Cape Town, South Africa*
8. Gerbrand Koren, *Copernicus Institute of Sustainable Development, Utrecht University, Utrecht, the Netherlands*
9. Gabriel Vecchi, *Department of Geosciences, Princeton University, Princeton, NJ 08544, USA, High Meadows Environmental Institute, Princeton University, Princeton, NJ 08544, USA*
10. Wenchang Yang, *Department of Geosciences, Princeton University, Princeton, NJ 08544, USA*
11. Sihan Li, *Department of Geography, University of Sheffield*
12. Maja Vahlberg, *Red Cross Red Crescent Climate Centre, The Hague, the Netherlands*
13. Roop Singh, *Red Cross Red Crescent Climate Centre, The Hague, the Netherlands*
14. Dorothy Heinrich, *Red Cross Red Crescent Climate Centre, The Hague, the Netherlands*
15. Carolina Marghidan Pereira, *Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, the Netherlands; Red Cross Red Crescent Climate Centre, The Hague, the Netherlands*
16. Julie Arrighi, *Red Cross Red Crescent Climate Centre, The Hague, the Netherlands; Global Disaster Preparedness Center, Washington DC, USA; University of Twente, The Netherlands*
17. Lisa Thalheimer, *United Nations University, Institute for Environment and Human Security, Bonn, Germany*
18. Cheikh Kane, *Red Cross Red Crescent Climate Centre, The Hague, the Netherlands; Institut de Recherche pour le Développement, U01000/99AA01, Marseille, France*
19. Friederike E. L Otto, *Grantham Institute, Imperial College London, UK*

Main findings

Want to age healthily? Focus on plant protein w

LESLIE BECK

OPINION

FOOD FOR THOUGHT

Protein plays a key role in healthy aging.

The vital nutrient is needed to maintain strong bones, preserve muscle strength and physical mobility and support the immune system.

Research has also hinted that getting enough protein helps maintain cognitive function in older age.

Not all types of dietary protein are considered equal, though.

New research from Tufts University and the Harvard T.H. Chan School of Public Health gives plant protein top marks when it comes to healthy aging.

According to the findings, women who ate more plant protein in their 40s and 50s increased their odds of aging healthily. Here's what to know.

Research shows consuming an additional 12 to 15 grams of plant protein each day in midlife is linked with a 46-per-cent greater likelihood of healthy aging. Plant protein in older age has also been found to protect against muscle loss, hip fracture and frailty. RYAN LIEBE/NYT

Globe & Mail, Feb 5

The American Journal of Clinical Nutrition 119 (2024) 271–282

ASN BE THE DIFFERENCE

American Society for Nutrition

Excellence in Nutrition Research and Practice

The American Journal of CLINICAL NUTRITION

journal homepage: https://ajcn.nutrition.org/

Original Research Article

Dietary protein intake in midlife in relation to healthy aging – results from the prospective Nurses' Health Study cohort

Andres V Ardisson Korat [1,2,*], M Kyla Shea [1], Paul F Jacques [1], Paola Sebastiani [3], Molin Wang [4,5,6], A Heather Eliassen [4,5,7], Walter C Willett [4,5,7], Qi Sun [4,5,7]

[1] USDA Human Nutrition Research Center on Aging, Tufts University, Boston, MA, United States; [2] Tufts University School of Medicine, Tufts

tochemicals, found in plant foods but not animal foods, may have contributed to the strong protective effects of plant protein.

PLANT PROTEIN DIETARY SWAPS

Applications

Example 3

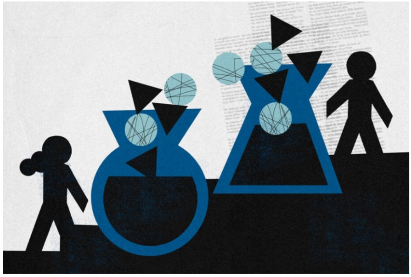
The Washington Post

Democracy Dies in Darkness

BUSINESS

Female scientists don't get the credit they deserve. A study proves it.

By Julianne McShane
June 22, 2022 at 11:00 a.m. EDT



Washington Post illustration)

The Washington Post

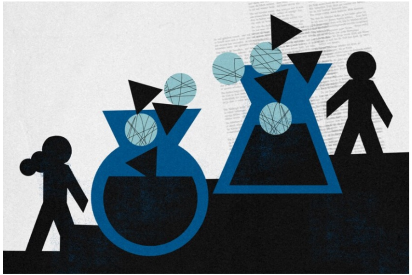
Democracy Dies in Darkness

BUSINESS

Female scientists don't get the credit they deserve. A study proves it.

By Julianne McShane

June 22, 2022 at 11:00 a.m. EDT



Washington Post illustration

nature

Explore content · About the journal · Publish with us

nature › articles › article

Article | Open Access | Published: 22 June 2022

Women are credited less in science than men

Matthew B. Ross, Britta M. Glennon, Raviv Murciano-Goroff, Enrico G. Berkes, Bruce A. Weinberg & Julia I. Lane

Nature **608**, 135–145 (2022) | Cite this article

97k Accesses | **57** Citations | **3111** Altmetric | Metrics

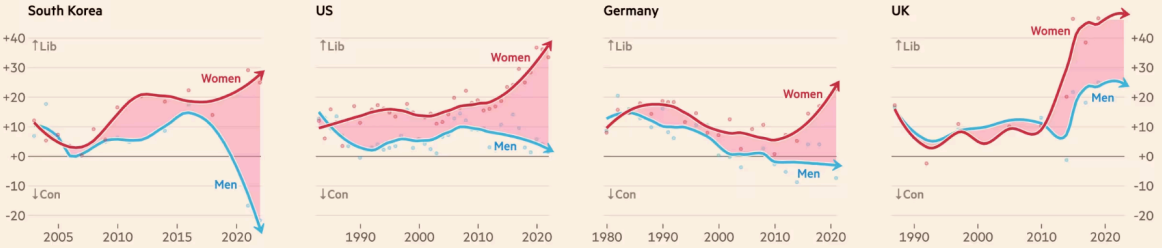
Abstract

There is a well-documented gap between the observed number of works produced by women and by men in science, with clear consequences for the retention and promotion of women[1]. The gap might be a result of productivity differences[2,3,4,5], or it might be owing to women's contributions not being acknowledged[6,7]. Here we find that at least part of this gap is the result of unacknowledged contributions: women in research teams are significantly less likely

Financial Times Jan 26

A wide ideology gap is opening up between young men and women in countries across the world

Political ideology of 18-29s (% liberal minus % conservative), by sex



Sources: Daniel Cox, Survey Center on American Life; Gallup Poll Social Series; FT analysis of General Social Surveys of Korea, Germany & US and the British Election Study. US data is respondent's stated ideology. Other countries show support for liberal and conservative parties
All figures are adjusted for time trend in the overall population
FT graphic: John Burn-Murdoch / @jburnmurdoch

Linking theory with practice

1. Drought Climate change attribution
   climate models, subgroup analyses, predictions
2. Protein Public Health
   observational study, missing data, confounding, multiple testing, causality
3. Science Co-authorship and gender
   linear regression, binary outcome, confounding
4. Ideology Political leanings by gender
   data sources, smoothing, visualization

Applications

Drought

CBC NEWS

Science

Deadly African drought not possible without climate change, study finds

Warming climate made long rains twice as likely to underdeliver, World Weather Attribution calculates

Thomson Reuters · Posted: Apr 27, 2023 8:43 AM EDT | Last Updated: April 27

ALJAZEERA

News ⌄ Ukraine war Features Economy Opinion Video

News | Drought

Global warming made Horn of Africa drought possible: WWA study

World Weather Attribution scientists say rising greenhouse gas emissions made the nearly 3-year drought at least 100 times more likely

Climate Change Made East Africa's Drought 100 Times as Likely, Study Says

The findings starkly show the misery that the burning of fossil fuels, mostly by rich countries, inflicts on societies that emit almost nothing by comparison.

Give this article

A water well near the town of Kelafo in Ethiopia, one of the nations hit hardest by the drought. Eduardo Soteras/Agence France-Presse — Getty Images

Human-induced climate change increased drought severity in Horn of Africa

Link

1. Joyce Kimutai, *Kenya Meteorological Department, Nairobi, Kenya*
2. Clair Barnes, *Grantham Institute, Imperial College London, UK*
3. Mariam Zachariah, *Grantham Institute, Imperial College, London, UK*
4. Sjoukje Philip, *Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands*
5. Sarah Kew, *Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands*
6. Izidine Pinto, *Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands*
7. Piotr Wolski, *Climate System Analysis Group, University of Cape Town, Cape Town, South Africa*
8. Gerbrand Koren, *Copernicus Institute of Sustainable Development, Utrecht University, Utrecht, the Netherlands*
9. Gabriel Vecchi, *Department of Geosciences, Princeton University, Princeton, NJ 08544, USA, High Meadows Environmental Institute, Princeton University, Princeton, NJ 08544, USA*
10. Wenchang Yang, *Department of Geosciences, Princeton University, Princeton, NJ 08544,*

The data

- observational data, 3 sources
  1. global daily rainfall & temperature — $0.5^\circ \times 0.5^\circ$, 1979 –
  2. daily rainfall — infra-red, "SoA", 1981 –
  3. monthly rainfall — 1981–2014

- 4-year smoothed mean surface temperature — proxy for anthropogenic climate change

- climate modelling data, 4 sources
  1. combine 12 global and 8 regional climate models: resolution 0.44° — 29 sims
  2. combine 5 global and 4 regional climate models: resolution 0.22° — 10 sims
  3. atmosphere-ocean coupled GCMs (two) — 10/3 simulations
  4. sea-surface temperature forced ensemble, high resolution — 11 simulations

- thanks to Whitney Huang, Clemson U, for many clarifications

climate models projected to 2100, "RCP8.5"

The analysis

- response is $\log_{10}$(monthly rainfall) in 2021 and 2022

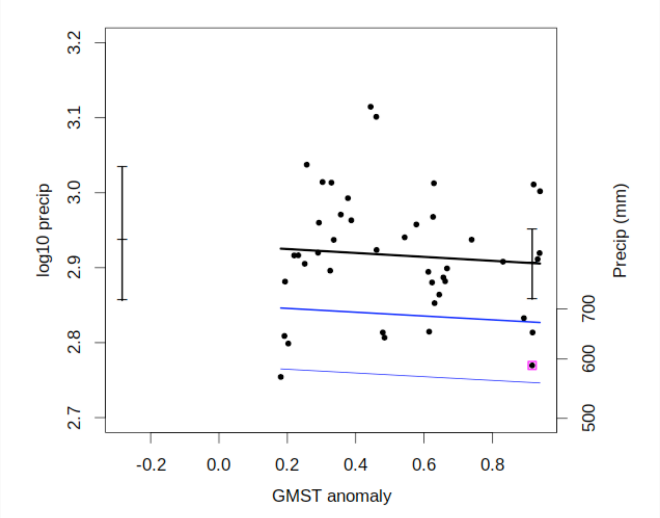
  and $\log_{10}$(PET) — potential evapotranspiration
- covariates are global temperature anomaly, and ENSO index

  El Nino-Southern Oscillation
- "As a measure of anthropogenic climate change we use smoothed GMST"

  Global Mean Surface Temperature
- "Methods for observational and model analysis … and synthesis are used according to the World Weather Attribution Protocol" Philip et al. 2020

1. trend using observational data
2. find climate models consistent with 1.
3. compare predictions from 1. and 2.
4. synthesize results in 3. to provide conclusions

The results



- rainfall decreasing with increasing temperature

  but not much

- 2022 rainfall is about a 1 in 20 year event

- 2022 drought about **2 times** more likely under climate change

- uncertainty 0.1 — 360

2 times more likely?

Kimutai et al. 2023

Climate Change Made East Africa's Drought 100 Times as Likely, Study Says

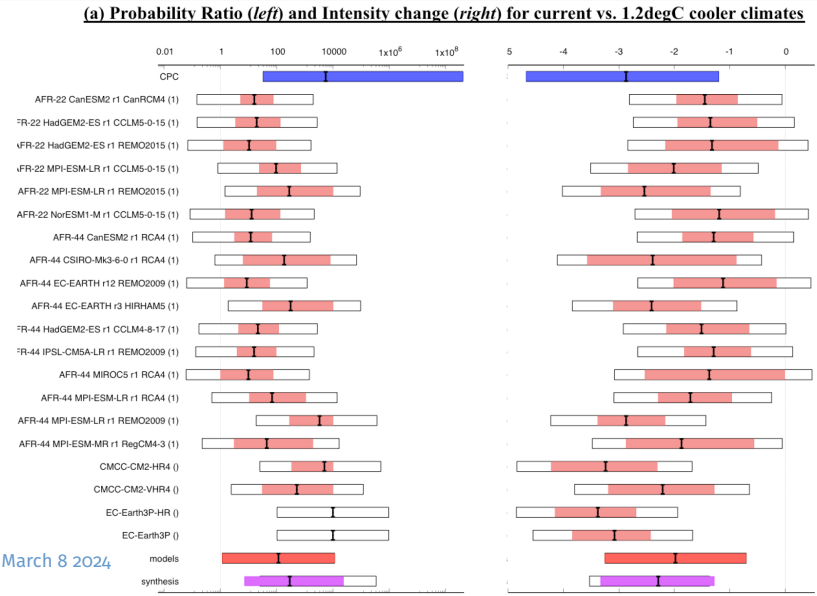
The findings starkly show the misery that the burning of fossil fuels, mostly by rich countries, inflicts on societies that emit almost nothing by comparison.

Give this article

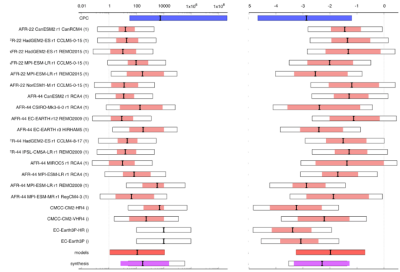
- change the response to SPEI
  rainfall adjusted for evaporation
- 2022 drought now **5500** times more likely
  uncertainty 32 to 4×10^8
- consider 'long rains' and 'short rains' separately MAM, OND
- combine model simulation results with observational data

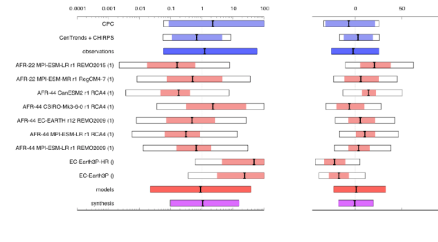
Combining climate simulations and data

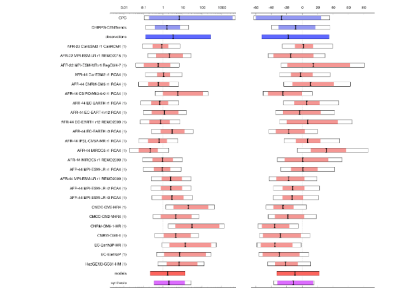
Kimutai et al. 2023

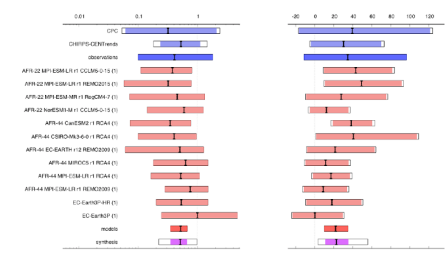


precipitation adjusted for evaporation









forest plots

- extrapolation beyond observations
  extreme value modelling — Stein *Statist. Sci.* 2019
- assigning uncertainty to combined results
  sources of uncertainty — Senn 2020
- ratios of estimated probabilities
  nearly unbounded confidence intervals — Cox & Hinkley 1974
- selection of events for analysis
- joint modelling of precipitation and evapotranspiration
  tail copula modelling

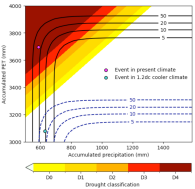


Figure 3: Joint distribution of 24-month precip and PET with corresponding SPEI drought classification (CPC dataset). The solid contours indicate return periods under the joint distribution in the current climate, while the dashed contours indicate the same return periods in a 1.2°C cooler climate. The shaded contours represent different levels of drought severity. The magenta point indicates the 2022 drought event in the current climate, with a joint return period of 26 years (uncertainty: 25-37 years), while the turquoise point shows an event of the

August 2023

world weather attribution

Home About Analyses News Peer reviewed research

On average, wildfires burn about 2.5 million hectares in Canada each year. In 2023, wildfires have already burned nearly 14 million hectares. Photo by Audrey Marcoux, SOPFEU.

Home » Wildfire » Climate change more than doubled the likelihood of extreme fire weather conditions in Eastern Canada

Climate change more than doubled the likelihood of extreme fire weather conditions in Eastern Canada

22 August, 2023

During May and June 2023 Canada witnessed exceptionally extreme fire-

Full study

- Download the full study: Climate change more than doubled the likelihood of extreme fire weather conditions in Eastern Canada (26 pages, 1.8MB)

… August 2023

Main findings

- Fire weather is one important condition driving wildfires, although changes in vegetation (wildfire fuel), ignition factors, and fire management strategies also contribute to future wildfire risk.
- In today's climate, intense fire weather like that observed in May-July 2023 is a moderately extreme event, expected to occur once every 20-25 years. This means in any given year such an event is expected with 4-5% probability.
- Climate change made the cumulative severity of Québec's 2023 fire season
- to the end of July around 50% more intense, and seasons of this severity at least seven times more likely to occur. Peak fire weather (FWI7x) like that experienced this year is at least twice as likely, and the intensity has increased by about 20% due to human-induced climate change.
- Observed changes are typically larger than in the models.
- As expected, likelihood and intensity are projected to increase further in a 2°C warmer world.
- Changes in fire weather are associated with an increase in temperature and decrease in humidity, both of which are driven by human-induced warming; the effect was compounded in 2023 by unusually low precipitation
- The extent, magnitude, and location of concomitant wildfires posed significant challenges for wildfire management which largely focused on disaster response and wildfire containment to limit the impact on lives and infrastructure.
- The wildfires had disproportionate impacts on indigenous, fly-in, and other remote communities who were particularly vulnerable due to lack of services and barriers to response interventions.
- The consequences from the wildfires reached far beyond the burned areas with displaced impacts due to air pollution threatening health, mobility, and economic activities of people across North America.
- As fire weather risks increase, changes in fire management strategies and increased resources will be required to meet the increased challenges.

"peak fire weather …
at least twice as likely"

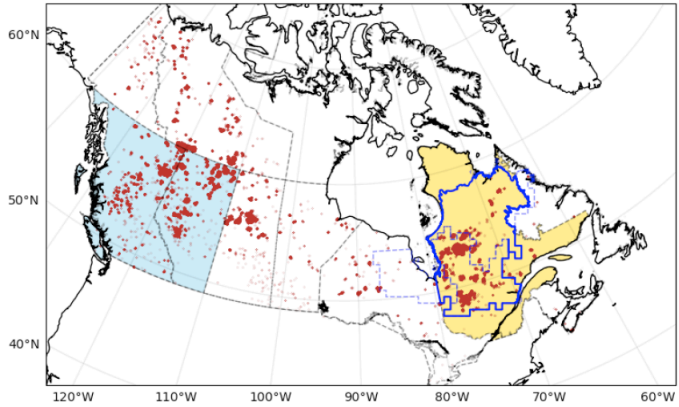


Figure 1: *Red dots mark active fires from January-July 2023 identified with high confidence (>80%) in the MCD14DL MODIS Active Fire and Thermal Anomalies product. The heavily impacted provinces of Alberta and*

… August 2023

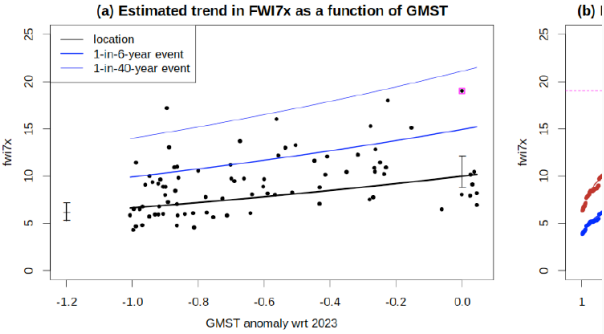
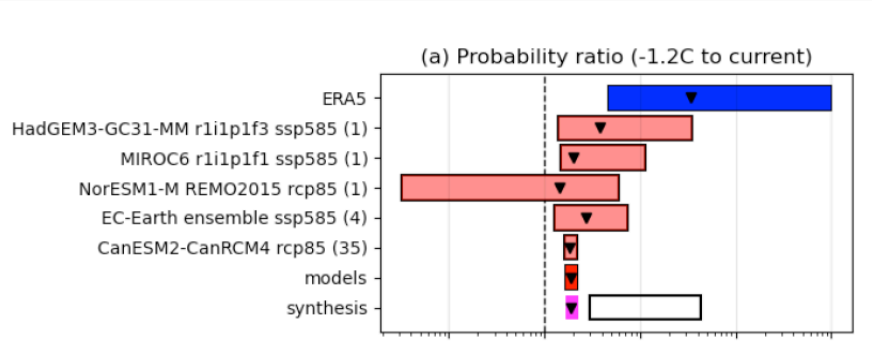


***Figure 7**: (a) Linear trend in ERA5 FWI7x as a function of GMST. parameter of the fitted distribution, and the blue lines show estimat lines show the 95% confidence interval for the location parameter*

… August 2023



"data": 34 times as likely (4.7, ∞)

models: 1.9 times as likely (1.66, 2.24)

Applications

Protein

Want to age healthily? Focus on plant protein w

LESLIE BECK

OPINION

FOOD FOR THOUGHT

Protein plays a key role in healthy aging.

The vital nutrient is needed to maintain strong bones, preserve muscle strength and physical mobility and support the immune system.

Research has also hinted that getting enough protein helps maintain cognitive function in older age.

Not all types of dietary protein are considered equal, though.

New research from Tufts University and the Harvard T.H. Chan School of Public Health gives plant protein top marks when it comes to healthy aging.

According to the findings, women who ate more plant protein in their 40s and 50s increased their odds of aging healthily. Here's what to know.

Research shows consuming an additional 12 to 15 grams of plant protein each day in midlife is linked with a 46-per-cent greater likelihood of healthy aging. Plant protein in older age has also been found to protect against muscle loss, hip fracture and frailty. RYAN LIEBE/NYT

The American Journal of Clinical Nutrition 119 (2024) 271–282

ASN BE THE DIFFERENCE

American Society for Nutrition

Excellence in Nutrition Research and Practice

The American Journal of CLINICAL NUTRITION

journal homepage: https://ajcn.nutrition.org/

Original Research Article

Dietary protein intake in midlife in relation to healthy aging – results from the prospective Nurses' Health Study cohort

Andres V Ardisson Korat [1,2,*], M Kyla Shea [1], Paul F Jacques [1], Paola Sebastiani [3], Molin Wang [4,5,6], A Heather Eliassen [4,5,7], Walter C Willett [4,5,7], Qi Sun [4,5,7]

[1] USDA Human Nutrition Research Center on Aging, Tufts University, Boston, MA, United States; [2] Tufts University School of Medicine, Tufts

tochemicals, found in plant foods but not animal foods, may have contributed to the strong protective effects of plant protein.

PLANT PROTEIN DIETARY SWAPS

Globe & Mail, Feb 5

The American Journal of Clinical Nutrition 119 (2024) 271–282

American Society for Nutrition
Excellence In Nutrition Research and Practice

The American Journal of CLINICAL NUTRITION

journal homepage: https://ajcn.nutrition.org/

Original Research Article

Dietary protein intake in midlife in relation to healthy aging – results from the prospective Nurses' Health Study cohort

Andres V Ardisson Korat [1,2,*], M Kyla Shea [1], Paul F Jacques [1], Paola Sebastiani [3], Molin Wang [4,5,6], A Heather Eliassen [4,5,7], Walter C Willett [4,5,7], Qi Sun [4,5,7]

[1] *USDA Human Nutrition Research Center on Aging, Tufts University, Boston, MA, United States;* [2] *Tufts University School of Medicine, Tufts*

The data

- Nurses Health Study
  - launched in 1976
  - 121,700 registered female nurses 30-55 y — married
  - follow-up every two years
  - initially focussed on oral contraception, later expanded to include diet and physical activity, biological samples, genetic information, ...
  - NHS II launched in 1989; NHS 3 in 2010 — younger ages, Canada, male
- detailed food frequency questionnaire administered in 1984, 1986
- treatment: protein intake
  - calculated from 1984 and 1986 questionnaires — self-reported
  - separated into animal, dairy, and plant protein — percentage of total energy intake
- response: healthy aging
  absence of chronic disease, no memory impairment, no limitation in physical function, good mental health — questionnaires 2014, 2016
- many potential confounding variables — education, smoking, race, etc.

- participants: 81,702 returned FFQ in 1984 (baseline)
- exclusions: $\geq$ 60 y, history of chronic diseases at baseline, incomplete FFQs in 1984 or 1986, incomplete or missing information on memory, physical function, depression questionnaires 2014 or 2016
- leaving 48,762 participants for main analyses
- 7.6% achieved "healthy aging"
  - 31% free of chronic disease
  - 47% no memory complaints
  - 15% no physical limitations
  - 37% good mental health

self reports

The analysis

Korat et al 2024

- "we used multivariable logistic regression analysis to calculate the odds ratios and 95% confidence intervals for the association of each dietary protein exposure with the odds of healthy aging"

- "adjusted for age, race, education, marital status, postmenopausal hormone use, smoking status, alcohol intake, physical activity, baseline history of hypertension or hypercholesterolemia, aspirin and multivitamin use, sugar-sweetened beverage intake, and total energy intake"
- "we also adjusted for BMI"
- "in our final regression we further adjusted for type of protein"

- "we used a missing indicator for missing categorical covariate data. No values were missing for continuous covariates"
- additional analyses of substitution effects, several sensitivity analyses

- response $y_i = 1/0$ — healthy aging/not
- treatment $z_i =$ continuous — ? % of total calories from protein
- covariates $\boldsymbol{x}_i$ — confounding variables: age, etc.
- probability model

$$y_i \sim \text{Bernoulli}(p_i), \quad \log\{p_i/(1-p_i)\} = \beta_0 + \gamma z_i + \boldsymbol{x}_i^{\mathrm{T}}\beta$$

- the estimate of $\exp(\gamma)$ is the estimated increase in odds of healthy aging associated with a unit increment in z — all other variables held fixed
- if the 95% confidence interval for $\exp(\gamma)$ excludes 1, association is called "statistically significant" — by many

The results

Korat et al 2024

	Quintile of protein intake					*P*-trend[5]	ORs (95% CI) for 3%-energy increment
	1	2	3	4	5		
Total protein							
Healthy ager/participants (*n*)	723/9752	781/9753	796/9752	748/9753	673/9752		
Median intake (IQR) % energy	14.6 (13.6, 15.2)	16.5 (16.2, 16.9)	17.3 (17.6, 18.3)	19.4 (19.0, 19.9)	21.9 (21.1, 23.3)		
Age-adjusted model[1]	1.00	1.12 (1.00, 1.25)	1.18 (1.06, 1.31)	1.16 (1.03, 1.29)	1.11 (0.99, 1.25)	0.06	1.05 (1.01, 1.09)
Multivariate model 1[2]	1.00	0.96 (0.85, 1.08)	0.99 (0.88, 1.11)	0.95 (0.84, 1.07)	0.89 (0.78, 1.01)	0.07	0.97 (0.93, 1.01)
Multivariate model 2[3]	1.00	0.99 (0.88, 1.11)	1.06 (0.94, 1.20)	1.08 (0.95, 1.21)	1.08 (0.95, 1.23)	0.11	1.05 (1.01, 1.10)
Animal protein							
Healthy ager/participants (*n*)	775/9752	758/9753	783/9752	754/9753	651/9752		
Median intake (IQR) % energy	9.6 (8.6, 10.2)	11.6 (11.2, 12.0)	13.1 (12.7, 13.4)	14.6 (14.2, 15.1)	17.3 (16.3, 18.7)		
Age-adjusted model[1]	1.00	0.99 (0.89, 1.10)	1.03 (0.92, 1.14)	1.01 (0.91, 1.13)	0.94 (0.84, 1.05)	0.38	0.99 (0.94, 1.05)
Multivariate model 1[2]	1.00	0.91 (0.81, 1.01)	0.92 (0.82, 1.03)	0.90 (0.80, 1.00)	0.81 (0.72, 0.91)	0.001	0.94 (0.91, 0.98)
Multivariate model 2[3]	1.00	0.93 (0.83, 1.04)	0.99 (0.89, 1.11)	1.02 (0.91, 1.15)	0.99 (0.88, 1.12)	0.64	1.01 (0.98, 1.05)
Multivariate model 3[4]	1.00	0.98 (0.87, 1.10)	1.07 (0.95, 1.20)	1.12 (0.995, 1.27)	1.13 (0.99, 1.29)	0.01	1.07 (1.02, 1.11)
Dairy protein							
Healthy ager/participants (*n*)	693/9752	729/9753	799/9752	750/9753	750/9752		
Median intake (IQR) % energy	1.7 (1.4, 2.0)	2.6 (2.4, 2.8)	3.4 (3.2, 3.6)	4.3 (4.0, 4.6)	5.9 (5.4, 6.8)		
Age-adjusted model[1]	1.00	0.97 (0.87, 1.09)	1.11 (0.99, 1.24)	1.05 (0.94, 1.17)	1.17 (1.05, 1.31)	0.002	1.12 (1.06, 1.20)
Multivariate model 1[2]	1.00	0.93 (0.83, 1.04)	0.99 (0.89, 1.11)	0.93 (0.83, 1.04)	1.04 (0.92, 1.17)	0.42	1.05 (0.98, 1.12)
Multivariate model 2[3]	1.00	0.95 (0.85, 1.07)	1.03 (0.91, 1.15)	0.97 (0.86, 1.09)	1.06 (0.94, 1.20)	0.26	1.06 (0.99, 1.13)
Multivariate model 3[4]	1.00	0.97 (0.87, 1.09)	1.06 (0.95, 1.20)	1.02 (0.91, 1.15)	1.17 (1.03, 1.33)	0.009	1.14 (1.06, 1.23)
Plant protein							
Healthy ager/participants (*n*)	611/9752	750/9753	726/9752	796/9753	838/9752		
Median intake (IQR) % energy	3.7 (3.4, 4.0)	4.4 (4.2, 4.5)	4.8 (4.7, 4.9)	5.2 (5.1, 5.4)	6.0 (5.7, 6.4)		
Age-adjusted model[1]	1.00	1.26 (1.13, 1.42)	1.26 (1.12, 1.41)	1.47 (1.31, 1.65)	1.77 (1.58, 1.99)	<0.0001	1.82 (1.64, 2.03)
Multivariate model 1[2]	1.00	1.12 (0.996, 1.26)	1.10 (0.98, 1.24)	1.22 (1.08, 1.37)	1.42 (1.26, 1.60)	<0.0001	1.46 (1.30, 1.65)
Multivariate model 2[3]	1.00	1.12 (0.99, 1.26)	1.08 (0.96, 1.22)	1.17 (1.03, 1.32)	1.32 (1.16, 1.49)	<0.0001	1.31 (1.16, 1.48)
Multivariate model 3[4]	1.00	1.15 (1.02, 1.29)	1.12 (0.99, 1.27)	1.23 (1.08, 1.40)	1.41 (1.24, 1.62)	<0.0001	1.38 (1.24, 1.54)

BMI, body mass index; CI, confidence interval; IQR, interquartile range; MET, metabolic equivalent; OR, odds ratio; SSB, sugar-sweetened beverage.

[1] Logistic model adjusted for baseline age.

[2] Multivariate logistic model 1 was adjusted for baseline age (continuous), race (White, other), education (registered nurse, bachelor, or graduate), marital status (married, other), postmenopausal hormone use (premenopausal; never, past user, current user), smoking status (never smoked; former smoker, 0.1–14.9, 15.0–29.9, >30 pack-y), alcohol intake (0, 0.1–4.9, 5.0–14.9, >15.0 g/d), physical activity (<3, 3–8.9, 9–17.9, 18–26.9, >27 MET/wk), baseline history of hypertension or hypercholesterolemia (yes, no), aspirin use (never, past, current), multivitamin use (yes, no), SSB intake, and total energy intake (kcal/d, quintiles).

[3] Multivariate model 2 included the covariates in multivariate model 1 and was additionally adjusted for BMI (averaged 1984 and 1986; <22.5, 22.5–24.9, 25.0–27.5, 27.5–30.0, 30.0–34.9, >35.0).

[4] Multivariate model 3 included covariates in multivariate model 2 with additional mutual adjustment of animal and plant protein. The models for dairy protein were adjusted for plant protein and other animal

[5] Korat et al.

The American Journal of Cli

... the results

Korat et al 2024

	Plant Protein intake (quintiles)					
	1	2	3	4	5	ORs (95% CI)
Healthy/Total	611/9752	750/9753	726/9752	796/9753	838/9752	(per 3% ↑)
Median Intake (%)	3.7	4.4	4.8	5.2	6.0	
Model 1	1.00	1.12 (0.996, 1.26)	1.10 (0.98, 1.24)	1.22 (1.08, 1.37)	1.42 (1.26, 1.60)	1.46 (1.30, 1.64)
Model 2	1.00	1.12 (0.99, 1.26)	1.08 (0.96, 1.22)	1.17 (1.03, 1.32)	1.32 (1.16, 1.49)	1.31 (1.16, 1.48)
Model 3	1.00	1.15 (1.02, 1.29)	1.12 (0.99, 1.27)	1.23 (1.08, 1.40)	1.41 (1.24, 1.62)	1.38 (1.24, 1.54)

The results

Korat et al 2024

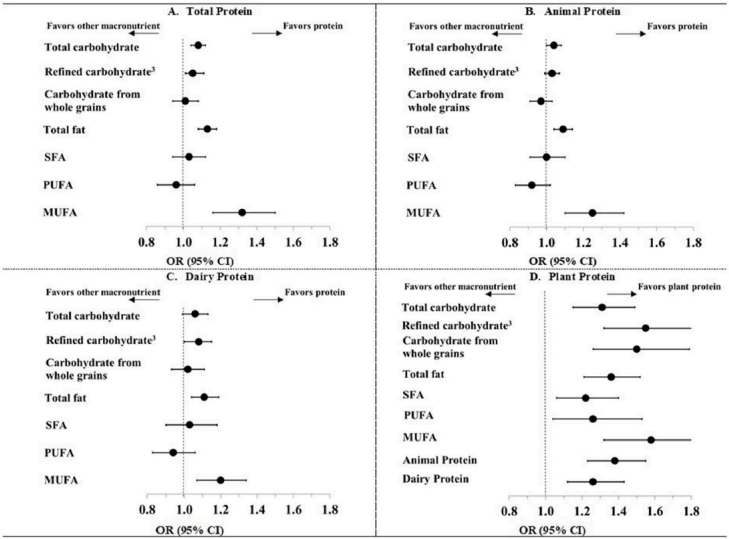


FIGURE 1. Odds ratios (ORs) (95% confidence intervals) of healthy aging (n = 3721) associated with isocaloric substitution of protein (total, animal, dairy,

Some questions

- most variables were categorized, including protein intake
- does this lose information?

- several dozen confidence intervals and *p*-values
- should these be adjusted for multiple comparisons? how?
  "which may lead to false positive findings"
- plant protein intake was small, and didn't vary much — does this matter?
- observational study, although prospective "association is not causation"

- Discussion: "dietary protein intake, especially … plant protein, in middle-aged female nurses, may be related to higher odds of healthy aging"

- Abstract: "Dietary protein intake, especially plant protein, in midlife, is associated with higher odds of healthy aging … in a large cohort of female nurses."

... Results

	Plant Protein intake (quintiles)					
	1	2	3	4	5	ORs (95% CI)
Healthy/Total	611/9752	750/9753	726/9752	796/9753	838/9752	(per 3% ↑)
Median Intake (%)	3.7	4.4	4.8	5.2	6.0	
Mean Intake (%)	4.7	4.9	4.9	4.9	4.8	
Model 1	1.00	1.12 (0.996, 1.26)	1.10 (0.98, 1.24)	1.22 (1.08, 1.37)	1.42 (1.26, 1.60)	1.46 (1.30, 1.64)
Model 2	1.00	1.12 (0.99, 1.26)	1.08 (0.96, 1.22)	1.17 (1.03, 1.32)	1.32 (1.16, 1.49)	1.31 (1.16, 1.48)
Model 3	1.00	1.15 (1.02, 1.29)	1.12 (0.99, 1.27)	1.23 (1.08, 1.40)	1.41 (1.24, 1.62)	1.38 (1.24, 1.54)

Some theory

- Bayesian approaches increasingly common in applied work
- how would this be implemented here?
- model

$$f(\boldsymbol{y}; \beta, \gamma) = \prod_{i=1}^{n} p_i(\boldsymbol{\beta}, \gamma)^{y_i} \{1 - p_i(\boldsymbol{\beta}, \gamma)\}^{1-y_i}$$

- prior ??

$$\pi(\boldsymbol{\beta}, \gamma)$$

- posterior

$$\pi(\boldsymbol{\beta}, \gamma \mid \boldsymbol{y}) \propto f(\boldsymbol{y}; \boldsymbol{\beta}, \gamma)\pi(\boldsymbol{\beta}, \gamma)$$

- marginal posterior

$$\pi(\gamma \mid \boldsymbol{y}) = \int_{\mathbb{R}^p} \pi(\boldsymbol{\beta}, \gamma) d\boldsymbol{\beta}$$

Applications

Authorship

nature

Article | Open Access | Published: 22 June 2022

Women are credited less in science than men

Matthew B. Ross, Britta M. Glennon, Raviv Murciano-Goroff, Enrico G. Berkes, Bruce A. Weinberg & Julia I. Lane ✉

Nature **608**, 135–145 (2022) | Cite this article

The data

Ross et al 2022 Nature

- “finding ‘what isn’t there’ from ‘what is there’ is a fundamental problem in statistics”

- analytic data
  - 118 campuses send deidentified data to U Michigan + survey + qualitative analysis
  - tracks spending on personnel for each research project payroll, all funding sources
  - 57 campuses with complete data 2013–2016
  - identify teams: PI, faculty, PDF, PhD, UGrad, Research Staff weighting of 1 for each person
  - identify publications Web of Science
  - identify gender, job titles, scientific fields, patents, ...

- 9800 teams with 129,000 team members
- 39,000 articles; 18m ‘potential authorships’, 367,000 actual authorships

scientific articles

- response attribution rate $= \dfrac{\#\text{ actual authorships}}{\#\text{ potential authorships}} = \text{pr}(\text{ attribution })$
- covariates date of publication, number of days worked in the team, calendar time, position in the team, team's PI
- model

$$P(\text{ named }) = \beta_0 + \beta_1 woman + \beta^T \text{ covariates } + \text{ error}$$

Empirical strategy

The empirical approach was to estimate linear regressions using a model of the form

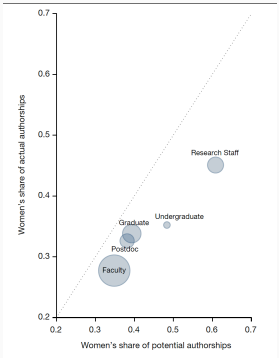
$$P[\text{named}_{i,t,e,l}|\ldots = \beta_0 + \beta_1 \text{woman}_{i,e} + X_{i,e} + M_{i,t} + \text{O}_{i,e} + \text{Team}_{i,l} + \mu_{i,t,e,l} \quad (1)$$

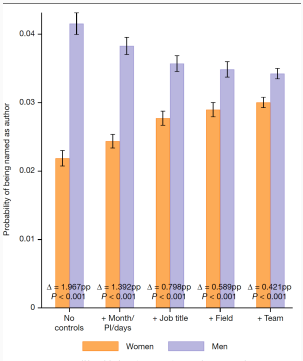
The results

overall attribution rate 3.1%;

attribution rate for men 4.23%; attribution rate for women 2.12% includes patents

difference smaller when covariates included but still statistically significant





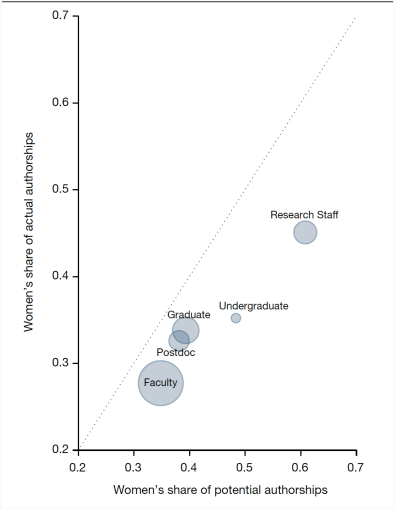


Figure 1

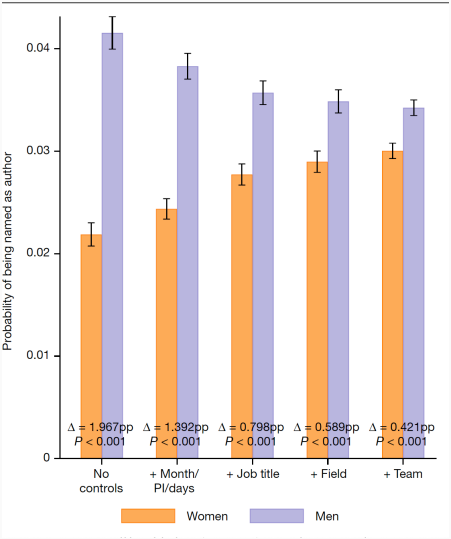


Figure 2

The theory

- linear regression but the response is a proportion
- logistic function is pretty linear for $p \in (0.2, 0.8)$
- but these p's $\in (0.01, 0.04)$ also many t-tests comparing ps
- there's a paper for that!

  On the linear in probability model for binary data Battey, Cox & Jackson 2019

  least squares estimate is more robust, coefficient directly interpretable

  less efficient, incorrect for observations out of range
- possibly more concerning: what is the unit of observation?

  potential authorship? article? team?

  are the standard errors correct?

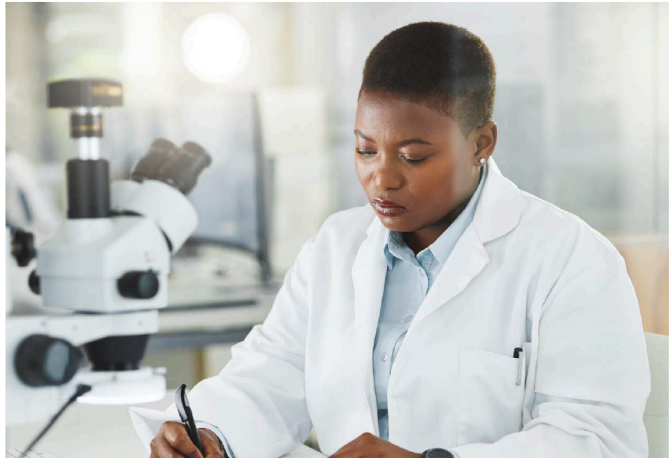
This just in

The international journal of science / 7 March 2024

nature

Too few women are being published in *Nature* — we must change that

This journal will double down on efforts to diversify the pool of corresponding authors.



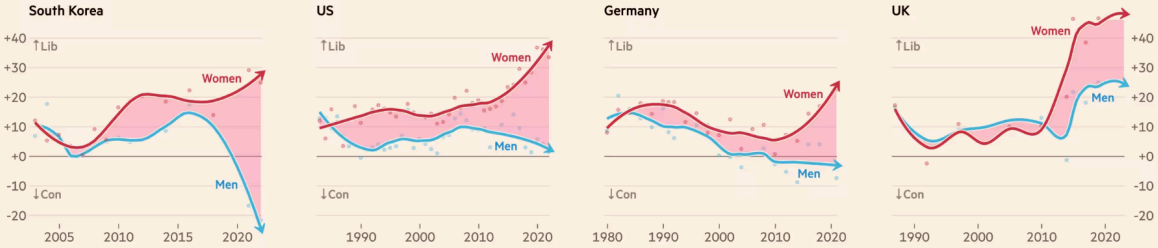
Researchers submitting original research to *Nature* over the past year will have noticed an extra question, asking them to self-report their gender. Today, as part of our commitment to helping to

Applications

Ideology

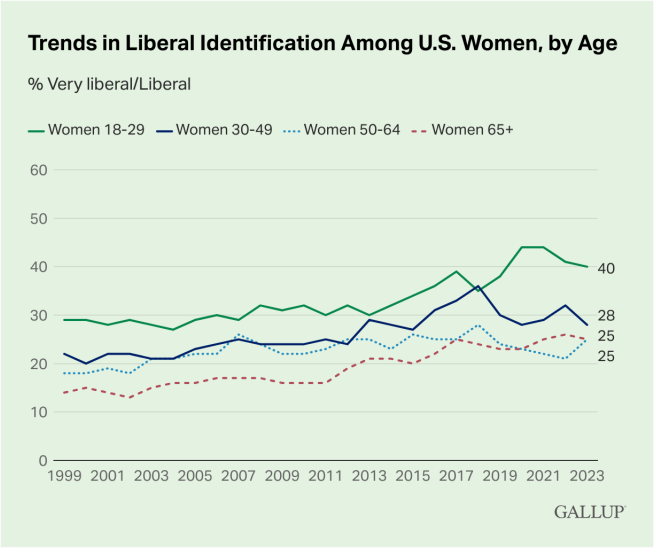
A wide ideology gap is opening up between young men and women in countries across the world

Political ideology of 18-29s (% liberal minus % conservative), by sex



Sources: Daniel Cox, Survey Center on American Life; Gallup Poll Social Series; FT analysis of General Social Surveys of Korea, Germany & US and the British Election Study. US data is respondent's stated ideology. Other countries show support for liberal and conservative parties
All figures are adjusted for time trend in the overall population
FT graphic: John Burn-Murdoch / @jburnmurdoch

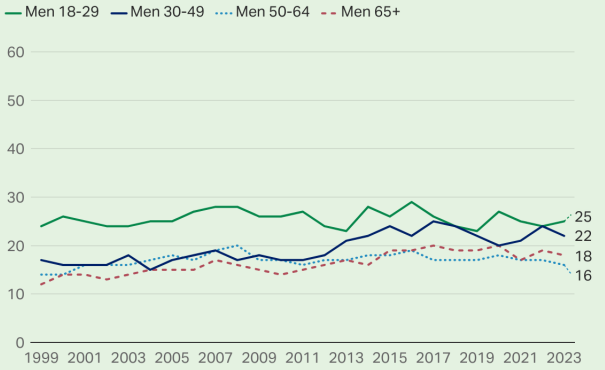
Financial Times Jan 24 "% liberal minus % conservative"



Gallup Report “% Very liberal/Liberal”

Trends in Liberal Identification Among U.S. Men, by Age

% Very liberal/Liberal

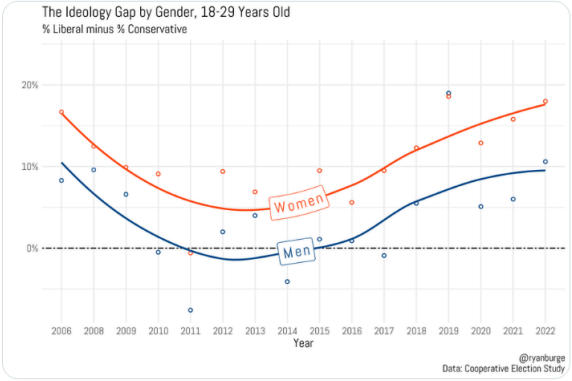


GALLUP

Ryan Burge 📊
@ryanburge

The finding that young women are becoming a lot more liberal while young men are becoming a lot more conservative DOES NOT REPLICATE in the Cooperative Election Study.

In fact, the two lines have run in almost perfect parallel for the last 15 years.



R. Burge: Cooperative Election Study

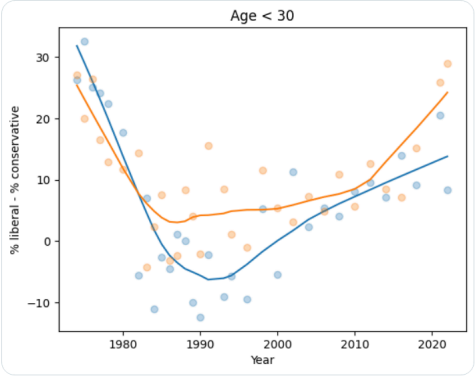
Allen Downey
@AllenDowney

This is interesting if true, but I'm not sure it is.

I tried to replicate the US graph with GSS data and I'm not seeing it.

Whether the gap is growing depends how seriously we take the last two points is a noisy series.

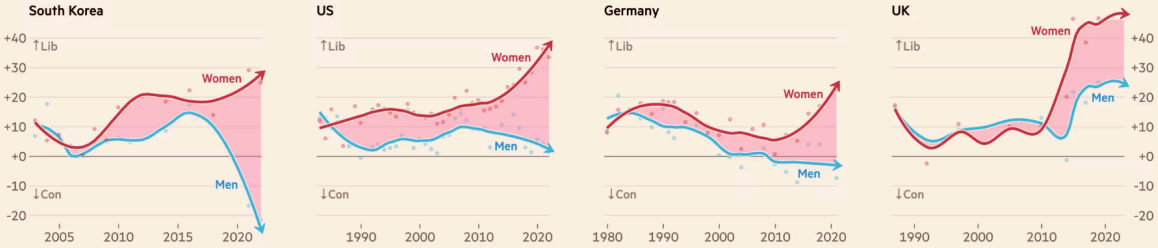
But it's nowhere near 30 points.



A. Downey, General Social Survey

A wide ideology gap is opening up between young men and women in countries across the world

Political ideology of 18-29s (% liberal minus % conservative), by sex



Sources: Daniel Cox, Survey Center on American Life; Gallup Poll Social Series; FT analysis of General Social Surveys of Korea, Germany & US and the British Election Study. US data is respondent's stated ideology. Other countries show support for liberal and conservative parties
All figures are adjusted for time trend in the overall population
FT graphic: John Burn-Murdoch / @jburnmurdoch

Financial Times Jan 24 “% liberal minus % conservative” ??

Back to Theory

Linking theory with practice

1. Drought Climate change attribution
   climate models, subgroup analyses, predictions
2. Protein Diet and Health
   observational study, missing data, confounding, multiple testing, causality
3. Science Co-authorship and gender
   linear regression, binary outcome, confounding
4. Ideology Gender imbalance in liberal/conservative leanings
   data sources, smoothing, visualization

Some observations

- statistical “workflows” seem to be emerging in different disciplines
  - Drought — “A Protocol for probabilistic extreme event attribution analysis ”
    Philip et al 2020, *Adv. Stat. Clim. Met. Ocean*
  - “Writing statistical methods for ecologists” Davis & Kay 2023, *Ecosphere*

- tutorial-type articles in scientific journals
  - *Annals of Thoracic Surgery* — the statistician’s page
  - *J Am Medical Association* — Guide to Statistics and Methods
  - *Nature Methods* — Points of Significance
  - *British Med J* — Statistics Notes

- “open data” continues to be elusive
  - Drought — “Almost all the data are available via the KNMI Climate Explorer”
  - Protein — “Data, code will not be made available due to patient confidentiality ”
  - Women — “datasets … are available at the Virtual Data Enclave Repository”

Communication!

For example…

Discussion: “dietary protein intake, especially … plant protein, in middle-aged female nurses, may be related to higher odds of healthy aging”

Abstract: “Dietary protein intake, especially plant protein, in midlife, is associated with higher odds of healthy aging … in a large cohort of female nurses.”

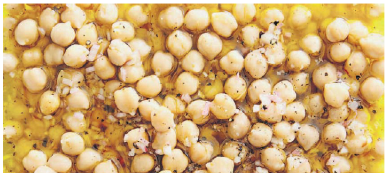
Globe & Mail article: “Protein plays a key role in healthy aging ”

Want to age healthily? Focus on plant protein when you’re younger

LESLIE BECK

OPINION

FOOD FOR THOUGHT



Protein plays a key role in healthy aging.

The vital nutrient is needed to maintain strong bones, preserve muscle strength and physical mobility and support the immune system.

Research has also hinted that getting

ciations hold up in males, ethnically diverse populations or in groups that eat a mainly plant-based diet.

HOW PLANT PROTEIN MAY PROTECT

Dietary protein stimulates muscle protein synthesis, a process that becomes blunted when older. Increasing muscle protein synthesis is associated with improved physical function in older adults. (Resistance training is the most potent stimulus of muscle protein synthesis.)

Plant protein has also been tied to favourable levels of risk factors for chronic

Science

PERSPECTIVE | HEALTH

July 21,2023

Heart failure causes sleepless nights

Cardiac dysfunction triggers immune-mediated loss of pineal gland melatonin release

By Harvey Davis and David Attwell

Science

PERSPECTIVE | HEALTH

July 21,2023

RESEARCH ARTICLE | HEART DISEASE

Immune-mediated denervation of the pineal gland underlies sleep disturbance in cardiac disease

KARIN A. ZIEGLER, ANDREA AHLES, ANNE DUECK, DENA ESFANDYARI, PAULINE PICHLER, KAROLIN WEBER, STEFAN KOTSCHI, ALEXANDER BARTELT, INGA SINICINA, [...], AND STEFAN ENGELHARDT +12 authors Authors Info & Affiliations

SCIENCE · 20 Jul 2023 · Vol 381, Issue 6655 · pp. 285-290 · DOI: 10.1126/science.abn6366

Heart failure causes sleepless nights

Cardiac dysfunction triggers immune-mediated loss of pineal gland melatonin release

By Harvey Davis and David Attwell

Science

PERSPECTIVE | HEALTH

July 21,2023

Heart failure causes sleepless nights

Cardiac dysfunction triggers immune-mediated loss of pineal gland melatonin release

By Harvey Davis and David Attwell

RESEARCH ARTICLE | HEART DISEASE

Immune-mediated denervation of the pineal gland underlies sleep disturbance in cardiac disease

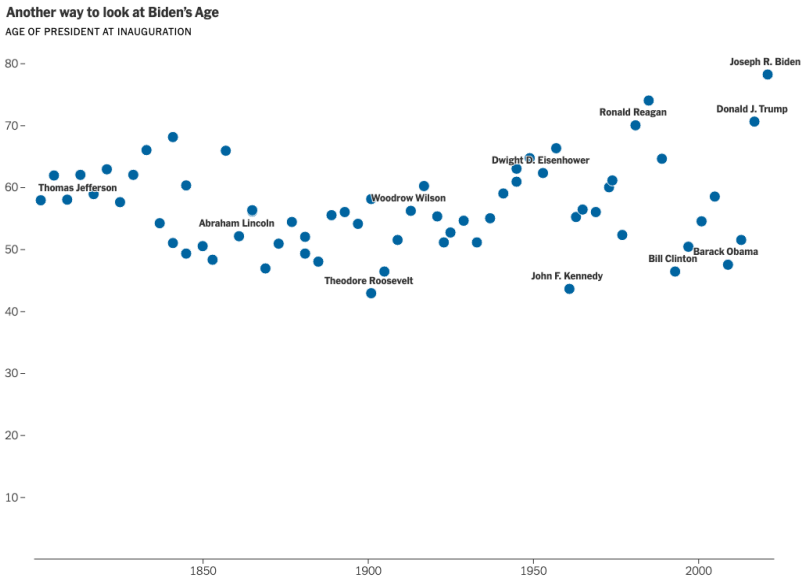
KARIN A. ZIEGLER, ANDREA AHLES, ANNE DUECK, DENA ESFANDYARI, PAULINE PICHLER, KAROLIN WEBER, STEFAN KOTSCHI, ALEXANDER BARTELT, INGA SINICINA, [...], AND STEFAN ENGELHARDT +12 authors Authors Info & Affiliations

SCIENCE · 20 Jul 2023 · Vol 381, Issue 6655 · pp. 285-290 · DOI: 10.1126/science.abn6366

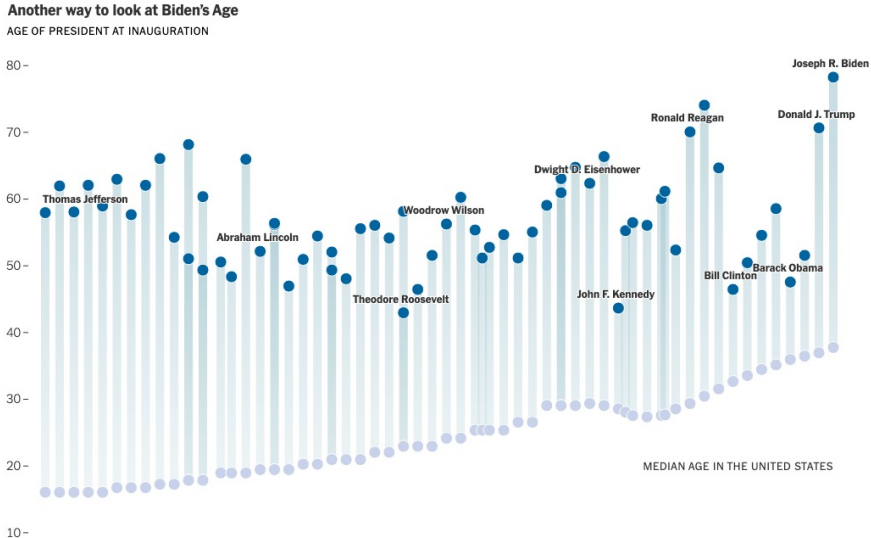
In mice

and in cells

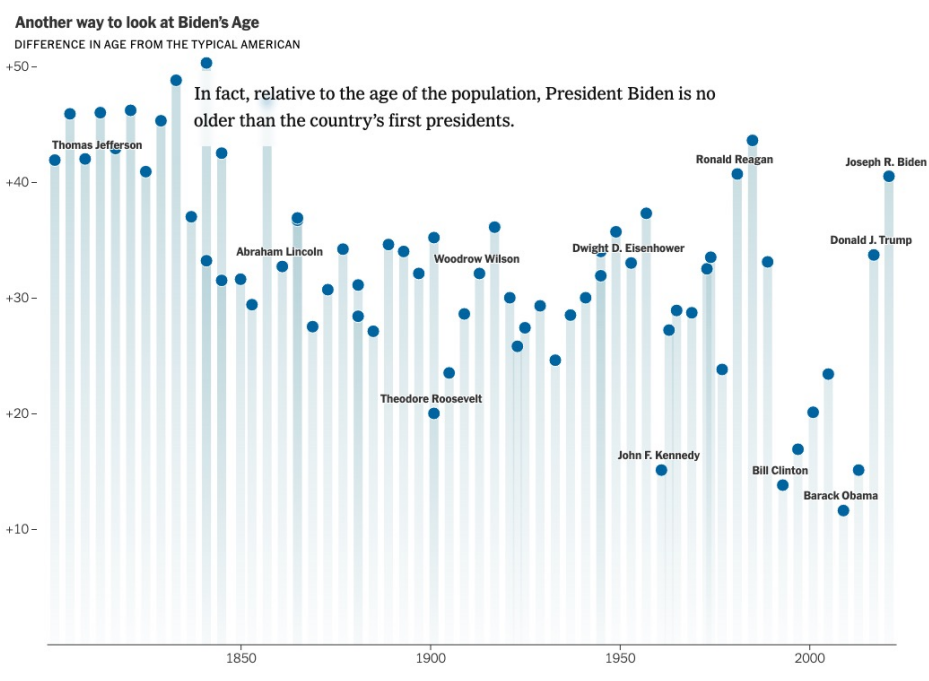
NY Times
March 6



When President Biden was elected in 2020, he became the oldest person to step into the office.



But since the country's founding, the American population also grew much older. The median age in America more than doubled since 1801.


Another way to look at Biden's Age
DIFFERENCE IN AGE FROM THE TYPICAL AMERICAN
In fact, relative to the age of the population, President Biden is no older than the country's first presidents.
+50
+40
+30
+20
+10
Thomas Jefferson
Abraham Lincoln
Woodrow Wilson
Theodore Roosevelt
Dwight D. Eisenhower
John F. Kennedy
Ronald Reagan
Bill Clinton
Barack Obama
Joseph R. Biden
Donald J. Trump
1850
1900
1950
2000

Jane Gentleman

Lynne Billard

Mary Thompson

Thank You!!

References i

Barnes, C. et al. (2023). Climate change more than doubled the likelihood of extreme fire weather conditions in Eastern Canada. World Weather Attribution Report 22 August 2023 https://www.worldweatherattribution.org/climate-change-more-than-doubled-the-likelihood-of-

Battey, H.S., Cox, D.R. and Jackson, M. (2019). On the linear in probability model for binary data. *Royal Society Open Science* **6**, 190067.

Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*. London, Chapman & Hall.

Davis, A.J. and Kay, S. (2023). Writing statistical methods for ecologists. *Ecosphere* 14:e4539. doi:10.1002/ecs2.4539

Fisher, R.A. (1958) *Statistical Methods for Research Workers.* 13th ed. Edinburgh, Oliver & Boyd.

Kimutai, J. et al. (2023). Human induced climate change increased drought severity in Horn of Africa. World Weather Attribution Report. https://www.worldweatherattribution.org/

References ii

human-induced-climate-change-increased-drought-severity-in-southern-horn-of-africa/ accessed August 23, 2023.

Korat, A.V.A. et al. (2024). Dietary protein in midlife in relation to healthy again — results from the prospective Nurses' Health Study cohort. *Am. J. Clin. Nutrition* **119**, 271–282.

Philip, S. et al. (2020). A protocol for probabilistic extreme event analysis. *Adv. Stat. Clim. Meteorol. Oceanogr.* **6**, 177–203.

Reid, N. and Cox, D.R. (2015). On some principles of statistical inference. *Int. Statist. Rev.* **83**, 298–308. doi:10.1111/insr.12067

Ross, M.B. et al. (2022). Women are credited less in science than men. *Nature* **608**, 135–145.

Savage, L. J. (1954). *The Foundations of Statistics*. New York, Dover.

References iii

Senn, S. (2020). “Error point: the importance of knowing how much you don’t know” (Guest Post). January, 20, 2020. https://errorstatistics.com/2020/01/20/s-senn-error-point-the-importance-of-knowing-how-much-you-dont-know-guest-post/

Stein, M.L. (2019). Some statistical issues in climate science. *Statist. Sci.* **35**, 31–41.

Zhong, R. (2023). “Climate change made East African drought 100 times as likely, study says”. *New York Times* April 22, 2023. https://www.nytimes.com/2023/04/27/climate/horn-of-africa-somalia-drought.html accessed August 13, 2023.