When are Causal Inference Methods Needed to Answer Causal Questions?

Yale School of Public Health
Data Science and Data Equity
October 12, 2026

Donna Spiegelman, Sc.D.
Susan Dwight Bliss Professor of Biostatistics and
Director, Center for Methods in Implementation and Prevention Science
Yale School of Public Health
https://ysph.yale.edu/cmips/

donna.spiegelman@yale.edu



Classical Assumptions

- Rubin's SUTVA (1986, 1980): Rubin DB. Comment: Which ifs have causal answers. *Journal of the American Statistical Association.* 1986;81(396):961–962.
  - (1) one version of treatment,
  - (2) no interference
  - (3) NO CAUSATION WITHOUT MANIPULATION?
- Robins and Hernan's causal identification assumptions (1986-2006): **Hernán MA, Robins JM. "Estimating causal effects from epidemiological data."** ***Journal of Epidemiology and Community Health.*** **2006;60(7):578–586.** R **Hernán MA, Robins JM.** ***Causal Inference: What If.*** **Boca Raton: Chapman & Hall/CRC; 2020.** Chapters 2–3 present exchangeability, positivity, and consistency as the three causal identifiability conditionsRobins JM. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling.* 1986;7(9–12):1393–1512. (developed g-formula)
  - (a) consistency,
  - (b) exchangeability,
  - (c) positivity

  Key question: **which assumptions are truly necessary for unbiased/consistent estimation (when) (with which methods)?**

  Unbiased estimation = valid estimation = $E[Y(1)] - E[Y(0)] = E[Y \mid A = 1] - E[Y \mid A = 0]$

  Or

  $$E[Y(1)]/E[Y(0)] = E[Y \mid A = 1] / E[Y \mid A = 0]$$

  - unbiased (rarely) or consistent estimates
  - valid asymptotic inference?
  - meaningful scientific interpretation?

- I claim that these assumptions are mostly not necessary for valid causal estimation and inference in epidemiology, clinical research and implementation science nearly all of the time

Are epidemiologists only interested in associations?

NO!

We have always been interested in causal parameters, although we called them valid or unbiased

Same goes for clinical researchers and

Listing and debunking the assumptions one by one

- no interference
- positivity
- one version of treatment/consistency
- no causation without manipulation
- exchangeability

I will refute the requirements for the first 4 leaving only exchangeability, which is the causal term for what epidemiologists call ‘no confounding’ and have been addressing since the 1950s

Interference: Why Was It Initially Forbidden?

- Interference, contamination – bad words
- In public health, interference is often desirable rather than problematic. What could be better than for beneficial interventions to spread freely through the population?
- Spillover permits efficient intervention dissemination through social and organizational networks
- Potential outcomes become more complex, but design and analysis largely follow standard correlated-data methods:
  - Direct, indirect, spillover, and overall effects are easily estimable when
  - the exposure structure is determined through randomization, as in two stage designs (cite) – first randomize a network to intervention or control; then randomize the pre-determined number of participants in the networks randomized to the intervention to directly receive the intervention.
    - E.g. Buchanan AL, Vermund SH, Friedman SR, Spiegelman D. **Assessing individual and disseminated effects in network-randomized studies.** *American Journal of Epidemiology.* 2018;187(11):2449–2459.
  - Or measured (who is in control group? Who directly received intervention? Who was in intervention group but only received intervention indirectly at most
  - Standard inference for correlated data has existed since Liang & Zeger 1986 GEE theory (Liang KY, Zeger SL. **Longitudinal data analysis using generalized linear models.** *Biometrika.* 1986;73(1):13–22.)

A sample of citations providing methods for estimation of causal effects in the presence of interference

General theory and analysis methods

Halloran ME, Struchiner CJ. Study designs for dependent happenings. Epidemiology. 1991;2(5):331–338.

Hudgens MG, Halloran ME. Toward causal inference with interference. Journal of the American Statistical Association. 2008;103(482):832–842.

Tchetgen Tchetgen EJ, VanderWeele TJ. On causal inference in the presence of interference. Statistical Methods in Medical Research. 2012;21(1):55–75.

Forastiere L, **Airoldi EM, Mealli F.** Identification and estimation of treatment and interference effects in observational studies on networks. Journal of the American Statistical Association. 2021;116(534):901–918.

Chao A, **Spiegelman D, Buchanan A,** Forastiere L. Estimation and inference for causal spillover effects in egocentric-network randomized trials in the presence of network membership misclassification. Biostatistics. 2025;26(1):kxaf009.

Halloran ME, Hudgens MG. Dependent happenings: a recent methodological review. Current Epidemiology Reports. 2016;3(4):297–305.

Benjamin-Chung J, Arnold BF, Berger D, Luby SP, Miguel E, Colford JM Jr, Hubbard AE. Spillover effects in epidemiology: parameters, study designs and methodological considerations. International Journal of Epidemiology. 2018;47(1):332–347.



35 years later Claudio Struchiner, International AIDS Conference, Rio de Janeiro, Brazil, 2026

Study Design

- Fang J, **Spiegelman D, Buchanan AL,** Forastiere L**.** Design of egocentric network-based studies to estimate causal effects under interference. Statistical Methods in Medical Research. 2025;34(10):2034–2052.

the tail wags the dog



What has happened

Public health research leveraging spillover to strengthen the effectiveness and cost-effectiveness of interventions was impeded

the dog wags the tail



What should happen

Embrace and leverage spillover whenever it makes sense to do so, which is often!

Listing and debunking the assumptions one by one

- ~~no interference~~
- positivity
- one version of treatment/consistency
- no causation without manipulation
- exchangeability

I will refute the requirements for the first 4 leaving only exchangeability, which is the causal term for what epidemiologists call ‘no confounding’ and have been addressing since the 1950s

Positivity: Necessary for Causal Inference or a Method-Dependent Technical Requirement?

- Classical positivity: everyone exposed must have non-zero probability of being unexposed, and vice versa

  The simplest form, for a binary treatment A with confounders L, theoretical positivity requires that $\boldsymbol{P(A = a \mid L = l) > 0}$ **∀ a &** $\boldsymbol{\forall\, l}$ **with** $\boldsymbol{P(L = l) > 0}$

  That is, within every covariate stratum observed, there is some chance of receiving each treatment level.

  - empirical positivity requires that each treatment level is observed at each adjustment-covariate level

- Section 6.4 of Hernán and Robins: "Exchangeability, positivity, and consistency are conditions required for causal inference via standardization or IP weighting. If any of these conditions does not hold, the numbers arising from the data analysis may not be appropriately interpreted as measures of causal effect." (Hernán MA, Robins JM. *Causal Inference: What If.* Boca Raton: Chapman & Hall/CRC; 2020.)
  - 'via standardization or IP weighting, or universally' – agreed but only for technical reasons
  - The second sentence is not qualified by via standardization or IP weighting, implies universality.
- Positivity problems arise from inverse-probability of treatment weighting and resulting instability of MSMs – a specific limitation of MSM/IPW methods and nothing more
- Corollary: positivity is not universally required for causal interpretation

Positivity: Necessary for Causal Inference or a Method-Dependent Technical Requirement?

- Alternative view of positivity (Miettinen, Poole): what matters is being at risk for the outcome, not risk for exposure (Miettinen OS. *Theoretical Epidemiology: Principles of Occurrence Research in Medicine.* New York: John Wiley & Sons; 1985; **Poole C.** Critical appraisal of the exposure-potential restriction rule. *Am J Epidemiol.* 1987;125(2):179–183.)
  - Groups never exposed can still serve as valid comparison groups, at least for relative risk estimation if their baseline outcome risks are reasonably assumed to be exchangeable with those in the exposed group, should they not have been exposed, conditional on confounders
  - For example, causal inference can be obtained from a study with men as the unexposed group looking at the relationship between oral contraceptives use and venous thrombosis, assuming men have the same risk as unexposed women, conditional on other confounders (Rozemeijer K, Le Cessie S, Van Hylckama VA, et al. Exposure opportunity: the advantages of including men in analyses of female-related risk factors. Am J Epidemiol. 2017;185(10):965-973.)
  - **(Those not at risk for outcome should never be included in study groups)**

- "Positivity … are sometimes presented as general validity requirements. Positivity, however, is not a general requirement for causal inference and does not address the more demanding practical needs for adequate numbers of observations." (Greenland S, Cole SR. Negatives about positivity and consistency as conditions for causal inference. Am J Epidemiol. 2026 Sep 3;195(9):2416-2422.)

- No one is giving mathematical proofs of these claims or counter-claims

A solution to the technical problem in IPW methods: overlap weights

- Positivity violations and near positivity violations wreak havoc with the finite sample properties of IPW/standardization-based estimators – this is a practical concern, not a theoretical one
  - IPW methods are key methods for valid causal effect estimation in studies with time-varying treatments/exposures with time-varying confounders
  - Standardization methods are essential for computation of externally transportable effect estimates when effect modification is a concern
- Overlap weights is the leading method to address this problem and has been found to work well

Citations and software for overlap weights

- Li F, Morgan KL, Zaslavsky AM. Balancing covariates via propensity score weighting. *J Am Stat Assoc* 2018;113(521):390–400. doi:10.1080/01621459.2016.1260466. This introduces the balancing-weights framework and overlap weights (OW). It shows OW minimizes asymptotic variance among balancing weights and gives exact mean balance when the PS comes from logistic regression.
- Li F, Thomas LE, Li F. Addressing extreme propensity scores via the overlap weights. *Am J Epidemiol* 2019;188(1):250–257. doi:10.1093/aje/kwy201. It compares OW with IPW and trimming by simulation and gives a closed-form sandwich variance estimator.
- Li F, Li F. Propensity score weighting for causal inference with multiple treatments. *Ann Appl Stat* 2019;13(4):2389–2415. doi:10.1214/19-AOAS1282. It extends OW to more than two treatments
- Cheng C, Li F, Thomas LE, Li F. Addressing extreme propensity scores in estimating counterfactual survival functions via the overlap weights. *Am J Epidemiol* 2022;191(6):1140–1151.
- Cao Z, Ghazi L, Mastrogiacomo C, Forastiere L, Wilson FP, Li F. Using overlap weights to address extreme propensity scores in estimating restricted mean counterfactual survival times. *Am J Epidemiol* 2025;194(8):2402–2411.
- Zhao Z, Yang C, Li F. Propensity score weighted Cox regression for survival outcomes in observational studies with multiple or factorial treatments. arXiv:2601.22572 (2026), preprint

R package Pssurvival for survival data analysis, and Psweight for generalized regression models

the tail wags the dog



What has happened

Positivity has become an axiom in causal inference, and an enormous amount of work has been invested in developing methods to address it

the dog wags the tail



What should happen

Use overlap weights when IPW/standardization is used to adjust for time-varying confounding bias in effects of time-varying exposures. Consider simple methods that use valid unexposed groups.

Listing and debunking the assumptions one by one

- ~~no interference~~
- ~~positivity~~
- one version of treatment/consistency
- no causation without manipulation
- exchangeability

Consistency and Real-World Treatments

- Is the concept of consistency similar to the Stable Unit Treatment Value Assumption? Seems so

$$\text{Consistency: } Y_i^{a = A_i} = Y_i^{obs} \text{, for all } i$$

- An alternative definition, based on studies drawn from super-populations defines consistency in terms of expectations over the super-population; this perspective is not in line with the conditional inference approach of epidemiology and clinical research
  - Cannot be true if there is variation in the treatment, that is, if $A_i = 1$ varies by dose, duration, frequency between treated subjects
- **Real-world treatments always vary: non-adherence, dose variation, measurement error, heterogeneous implementation**
- Strict consistency is therefore unrealistic as a universal requirement, and doing so implies causal inference is always impossible
  - Exception: Intent to treat analysis in trials (but not of substantive interest)
- **Causal contrasts should reflect the average treatment experienced in the target population** (Hernán MA, VanderWeele TJ. Compound treatments and transportability of causal inference. *Epidemiology.* 2011;22(3):368–377.)
  - Average treatment delivery may vary from one study population to another. To the extent this is the case, unless the variations have no impact on the outcomes, transportability of results will be limited to those populations with similar treatment variation distribution.

the tail wags the dog



What has happened

Causal inference has insisted on the unrealistic and impracticable requirement of a single version of treatments/interventions, thereby making itself irrelevant

the dog wags the tail



What should happen

Methods for addressing the real-world settings of multiple versions of treatment should be expanded and scaled out. Methods for external generalizability can be a good starting point.

Listing and debunking the assumptions one by one

- ~~no interference~~
- ~~positivity~~
- ~~one version of treatment/consistency~~
- no causation without manipulation
- exchangeability

Model mis-specification bias, and the turn towards non-parametric estimation for causal inference

- Correct model specification matters, but multiplicative models and GEE methods are frequently robust in practice
- **Multiplicative models rarely show evidence of the presence of interaction terms**
  - Treatment heterogeneity is rare on the multiplicative scale
  - Spiegelman D, VanderWeele TJ. Evaluating Public Health Interventions: 6. Modeling Ratios or Differences? Let the Data Tell Us. *American Journal of Public Health.* 2017;107(7):1087–1091;
  - Spiegelman D, Khudyakov P, Wang M, VanderWeele TJ. Evaluating Public Health Interventions: 7. Let the Subject Matter Choose the Effect Measure: Ratio, Difference, or Something Else Entirely. *American Journal of Public Health.* 2018;108(1):73–76.

"Although more formal quantitative work is needed, there is extensive anecdotal evidence that suggests that often in public health the multiplicative models fit the data well. The first author (me) has published nearly 650 peer-reviewed scientific publications in a diverse range of substantive areas including chronic disease epidemiology, HIV/AIDS, and environmental health, and, for nearly all of these, her primary contribution to the research was that of the study statistician. In almost all of these, the multiplicative model fit the data consistently and extraordinarily well. "

"The second author (VanderWeele), who has devoted much of his methodological work to interaction, has numerous examples of interaction on additive scales but, in more than 230 articles, can think of only two examples of multiplicative interaction that replicated across studies (e.g., VanderWeele et al.[10])"

"These impressions are further confirmed by other senior researchers: our department chair (the Harvard T.H. Chan School of Public Health Epidemiology Department chair), Albert Hofman, has informed us that he cannot think of a single important multiplicative modifier uncovered during his long research career among his more than 2000 scientific articles. Similarly, Walter Willett, former chair of Harvard's nutrition department and one of the world's most highly cited scientists, could think of four multiplicative modifiers among more than 1700 scientific publications.

- Corollary: Since effect modification on the risk ratio scale is so rare, effect modification on the risk difference scale will almost always be present
  - Risk differences are difficult to interpret and disfavored as the parameter of interest in epidemiologic and clinical research

"

Corollary: The estimands framework may be a new impedance to knowledge generation from public health and clinical research

- Estimands are identical in the absence of effect modification (e.g. facility-, patient-, visit- level causal estimates)
- Power may be lost when calculating these higher level estimators
- As previously asserted, effect modification is rare and these new classes of estimators can be generating unnecessary confusion
- Proposal: Before presenting estimators at higher levels, demonstrate that there is substantively important effect modification on that level, e.g. treatment by facility interactions, treatment by patient characteristic interactions
  - As we teach in epidemiology, when effect modification is present, it's best to present results by sub-groups and use the data to explain and understand the effect modification

Model mis-specification bias, and the turn towards non-parametric estimation for causal inference

- GEE models correct inference for correlated data of arbitrary structures
- Nonlinearity beyond monotonicity is exceedingly rare in epidemiologic data (except for seasonality in some infectious disease research)
  - Monotone trends are reasonably well approximated by linear ones
- Model forms can be empirically verified, and non-linearity and effect modification incorporated in the unusual cases when needed
- **Non-parametric estimators lead to an unacceptable loss of power in consequential studies in public health and clinical medicine, and should be avoided in primary analysis**

Further work investigating this point would be very helpful

the tail wags the dog



What has happened

Biostatistics has turned increasingly to development of methods that lead with the assumption of treatment effect heterogeneity, and journals are forcing authors to present results on multiple levels without subject matter justification

the dog wags the tail



What should happen

Longstanding empirical methods should be used to identify the presence of effect modification. If not there, multiple low-powered estimators, non-parametric and at different levels of clustering, should not be required or presented. This will nearly always be the case.

No causation without manipulation?

- Does trisomy 21 cause Down's syndrome? (necessary and sufficient)
- Does being old cause us to die? (sufficient)
- Does menopause cause accelerated bone loss? (sufficient)
- Does the Moon's gravity cause tides? (sufficient)
- Does Earth's tilt cause seasons? (necessary and sufficient)

…hardly

- No causation without manipulation is an absurd assumption or requirement for valid (causal) learning from epidemiologic and clinical data
  - This requirement violates our lived experiences
  - Glymour MM, Spiegelman D. Evaluating Public Health Interventions: 5. Causal Inference in Public Health Research - Do Sex, Race, and Biological Factors Cause Health Outcomes? Am J Public Health. 2017 Jan;107(1):81-85.
  - Plus several other papers arguing for and against this point

the tail wags the dog



What has happened

The application of causal inference theory and methods to the entire discipline of the social determinants of health is cancelled

the dog wags the tail



What should happen

Continue to investigate and document the impact of social determinants of health including, importantly, those that are not manipulable, such as age, race, gender and sexual orientation, using causal inference theory and methods whenever useful.

Listing and debunking the assumptions one by one

- ~~no interference~~
- ~~positivity~~
- ~~one version of treatment/consistency~~
- ~~no causation without manipulation~~
- Exchangeability = no confounding

Definitions: Exchangeability = No unmeasured confounding conditional on measured confounders

- **Marginal exchangeability** guaranteed by randomization but not generally necessary

  $Y^a \perp\!\!\!\perp A$, which implies $E[Y^0 | A = 1] = E[Y^0 | A = 0]$, that is, the potential outcome is independent of the treatment actually received.

- **Conditional exchangeability** (observational, adjusting for L):

  $$Y^a \perp\!\!\!\perp A \mid L$$

- **Weaker and nearly always sufficient — mean conditional exchangeability:**

  $$E[Y^a \mid A = a, L] = E[Y^a \mid L], \text{ implying}$$
  $$E[Y^0 | A = 1, L] = E[Y^0 | A = 0, L]$$

- Each of these defines conditions for no confounding
- Mean conditional exchangeability remains the standard identifying assumption for causal inference, that is, no confounding

To what extent is confounding, the only remaining criterion for causal identification, an issue in epidemiologic, clinical and public health research?

- In our attempt to identify specific causes of the large and rapid changes in NCD rates in space and time, the concern has been repeatedly raised that these efforts are difficult to interpret due to confounding.
- The primary hypothesized confounder in nutritional and environmental epidemiology is **'healthy lifestyle'.**
- What is healthy lifestyle?
  - Healthy diet, regular physical activity, not smoking, moderate or no alcohol intake, absence of traumatic events, including physical, sexual and emotional abuse in childhood and adulthood, high quality regular preventive health care, what else?

The Limits of Confounding



- It has been suggested that adjusting for a proxy for healthy lifestyle can address at least some of the confounding due to this mysterious variable.
  - Aka negative exposure controls (Lipsitch and others)
  - Proximal causal inference (Tchetgen Tchetgen and others)

- Suggested proxies for healthy lifestyle in relation to cancer and other NCD outcomes that have been studied or proposed include sunscreen use, agreeing to be in an RCT, adherence to placebo and/or active regimen in an RCT, seatbelt use, adherence to CRC or BrCA screening guidelines, adherence to annual health check-ups

‘Healthy lifestyle’ proxies

- In the Kaiser Permanente case-control study of hormone replacement therapy and vascular disease, a 50% (95% confidence interval (CI) 32%-57%) reduction in vascular disease risk was found among **sunscreen users** after adjustment for other risk factors .

- Sesso et al. reported a 36% (95% CI 18%-49%) reduction in the five year risk of cardiovascular mortality among physicians who agreed to be randomized (66%) compared with those who declined, **after adjusting for many behavioral and clinical risk factors** .

- Adherence to a placebo regimen in a double-blinded trial of myocardial infarction (91%) was associated with a 60% reduction in risk of mortality, most likely due to other healthy lifestyles to which the adherents follow .

- In a randomized, double-blind trial, good adherers to both the treatment and to placebo regimens (66%) were observed to be 40% and 50% less likely to die, respectively, again presumably due to other healthy lifestyle factors associated with good adherence behavior.

Pettitti, D.B., *Cardiovascular disease and hormone therapy: can we agree on what the women's health initiative randomized trial teaches us?* . Am J Epidemiol, 2005. **161**: p. S97.

What do we see in Harvard's Nurses Health Study and Health Professionals Follow-up Study with a healthy lifestyle proxy?

- Describe NHS and HPFS
- Outcome – CVD incidence
- Proxy – Sunscreen use
- Confounders – all measured baseline and time-varying determinants of CVD risk
- Claim: If sunscreen use is protective for CVD risk, assuming that it itself has no causal effect on CVD risk, it is a proxy for healthy lifestyle.
- If adjustment by measured confounders makes its effect go away, it suggests that there is no residual confounding by healthy lifestyle (at least not that correlated with sunscreen use)

What do we see in NHS and HPFS with a healthy lifestyle proxy?

- HPFS: Adjusted for age in months, and baseline and time-varying measurements of the following CHD risk factors: body mass index(<21, 21-22.9, 23-24.9, 25-26.9, 27-29.9, 27-29.9, 30-32.9, 33-34.9, 35+ kg/m^2), smoking status (never, past, current[1-14,15-24,25+] cigs/day),physical activity(Mets hrs/week, quintiles),alcohol consumption(0-4.9, 5-9.9, 10-14.9, 15-29.9, 30+g/day), healthy diet score(quintiles), hypertension, hypercholesterolemia, diabetes, baseline family history of MI and aspirin intake(yes, no) .
- NHS: Adjusted for age in months, and baseline and time-varying values of the following CHD risk factors: body mass index(<21, 21-22.9, 23-24.9, 25-26.9, 27-29.9, 27-29.9, 30-32.9, 33-34.9, 35+ kg/m2), smoking status (never, past, current[1-14,15-24,25+] cigs/day),physical activity(hrs/week, quintiles),alcohol consumption(0, 0.1-4.9, 5-14.9, 15+g/day), dietary score (quintiles), hypertension, hypercholesterolemia, diabetes, family history of MI and aspirin intake(yes, no) .
- Prevalence of regular sunscreen use was 41% in HPFS and 25% in NHS during the follow-up period: 1992-2002 in HPFS, and 1980-2002 in NHS

Table 1 Baseline Characteristics of Risk Factors According to Sunscreen Use in Men (1992) and Women (1980)

	Sunscreen use in 1992 (HPFS) (% frequency of time when in the sun)			Sunscreen use in 1980 (NHS)		
Risk factors	0%	25-50%	>75%	Never outdoors	Outdoors but little or no protection	Outdoors and usually use
BMI (kg/m^2)	26.2	25.7	25.5	24.9	24.3	23.9
Alcohol (gm/day)	10.8	11.0	10.2	5.3	6.4	7.0
Physical activity*	38.6	39.2	38.9	2.6	4.0	4.3
Calories (Kcal/day)	1971	1962	1916	1514	1581	1560
Energy-adjusted Cereal fiber (gm/day)	6.3	6.7	7.3	2.6	2.5	2.5
Energy-adjusted Polyunsaturated Fat (gm/day)	13.1	13.2	13.0	9.5	9.4	9.1
Energy-adjusted Saturated fat (gm/day)	24.1	23.3	22.1	28.4	28.2	27.2
Energy-adjusted Trans fat (gm/day)	3.61	3.47	3.26	4.06	4.07	3.82
Energy-adjusted Glycemic Load	128	129	133	85.6	85.4	86.1
Energy-adjusted Omega 3 fat (gm/day)	0.30	0.35	0.37	0.08	0.08	0.09
Dietary score**	14.2	15.0	15.4	14.8	14.8	15.3
Current smoking (%)	11	6	5	28	29	29
Hypertension (%)	25	25	25	17	15	15
Hypercholesterolemia(%)	32	34	34	6	5	5
Diabetes (%)	4	3	3	3	2	2
Family history of CHD (%)	31	32	32	20	20	21
Aspirin use (%)	28	31	32	46	46	48
Current hormone use among postmenopausal women (%)	----	----	----	32	29	33

- Prevalence of regular sunscreen use was 41% in HPFS and 25% in NHS

Table 2. The relative risks (95% CIs) for sunscreen use in 1992 in relation to CHD incidence 1992-2002 in HPFS, and in relation to CHD incidence 1980-2002 in NHS

	HPFS (n=27,341)			NHS (n=87,659)		
Sunscreen exposure	0%	25-50%	>75%	Never outdoors	Outdoors but little or no sunscreen use	Outdoors and usually use sunscreen
Number of events	360	251	348	466	1323	616
Incidence rate/100K PY	456	337	328	165	122	136
Adjusted for age in 5 year groups only[1]	1.0	0.88(0.75,1.04)	0.85(0.73,0.99)	1.0	0.85(0.76,0.94)	0.87(0.77,0.98)
Adjusted for age in months only	1.0	0.88(0.74,1.03)	0.84(0.72,0.98)	1.0	0.86(0.77,0.95)	0.87(0.77,0.98)
Adjusted for age in months, and main effects of baseline covariates[2 3]	1.0	0.98(0.82,1.15)	0.96(0.82,1.12)	1.0	0.96(0.86,1.07)	1.04(0.92,1.17)
Adjusted for age in months, and main effects of all baseline and time-varying covariates[4 5]	1.0	0.97(0.82,1.15)	0.96(0.82,1.12)	1.0	0.95(0.86,1.06)	1.02(0.90,1.15)
Adjusted for age in months, main effects of all baseline and time-varying covariates[4 5], and all two-way interactions of baseline and time-varying covariates, including age in five year groups, selected at p=0.20	1.0	0.98(0.83,1.15)	0.98(0.84,1.13)	1.0	0.94(0.85,1.05)	0.98(0.86,1.11)

1 Age in five-year age groups (<50, 50-54.9,55-59.9, 60-64.9, 70+)

The Limits of Confounding: What do we see in NHS and HPFS with a healthy lifestyle proxy?

- Conclusion – appreciable residual confounding by healthy lifestyle in well annotated epi studies where risk factors (aka potential confounders) are repeatedly measured with high quality **is unlikely.**

The Limits of Confounding: Theoretical results

- Bias due to confounding, including residual confounding, is limited by the minimum of 3 factors (Observation and Inference, Alec Walker, 1991) :
  1. The strength of the association of the confounder (U) with the outcome of interest (D)
  2. The strength of the association of the confounder (U) with the exposure of interest (E)
  3. The prevalence of the confounder

  Apologies for change in notation: D is Y, E is A, and U is L

- As prevalence gets smaller, bias goes to 0; as $Corr(E, U)$ gets smaller, bias goes to 0, as $RR(D|U)$ goes to 0, bias goes to 0, no matter what the values of the other 2 parameters. (e.g. Schlesselman, 1979)

The E-value (Vanderweele and Ding, 2017)

- The descendant of earlier work starting with Cornfield (1959), through Bross (1965), Schlesselman (1979) and many others, the E-value represents the minimum strength of association on the risk ratio scale, that an unmeasured confounder would need to have with **both** the exposure and the outcome, conditional on the measured covariates, to explain away the observed effect.

- The E-value assumes the prevalence of the unmeasured confounder is 50%, which is the value at which confounding bias is at its maximum with respect to this factor. Thus, the E-value will usually be an over-estimate of the extent to which residual confounding can explain observed impact estimates.

The E Value (Vanderweele and Ding, 2017)

- For an observed risk ratio, RR_{ED}^{obs}

$$\text{E-value} = RR_{ED}^{obs} + \sqrt{RR_{ED}^{obs} \times \left(RR_{ED}^{obs} - 1\right)}$$

  - The formula applies to a risk ratio greater than 1; for a risk ratio less than 1, one first takes the inverse of the observed risk ratio and then applies the formula.
- E-value is very easy to calculate, and applies very generally
- Applies to all sorts of models, outcomes, exposures

VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. Ann Intern Med. 2017 Aug 15;167(4):268-274.

The E -Value

- Since prevalence of the confounder matters – greatly we will see –
- Back to healthy lifestyle as the unmeasured confounder responsible for associations in observational epidemiology of NCDs…
- How common is a healthy lifestyle? Answer: not common at all

How Common is Healthy Lifestyle?

- We looked at this in Harvard's Health Professionals Follow-up Study (HPFS) and Nurses' Health Study (NHS)
- Among the 6 modifiable risk factors for colorectal cancer measured in HPFS:

1. Body mass index- an indicator for obesity, calculated as weight in kilograms divided by the square heigh in meters
2. Physical activity- Metabolic equivalent task hours per week derived from time engaged in leisure time activities
3. Folic acid intake- Multivitamin or a specific supplement
4. Alcohol consumption-Calculated from daily servings of beer, wine, or liquor
5. Smoking before age 30- Expressed in pack years and calculated from the number of cigarettes smoked daily in each decade of life
6. Red meat intake- Servings a day of beef, pork, or lamb, as a main dish

We then calculated a risk score that was the sum across the six risk factors of the values of 1 (healthiest level) to 5 (unhealthiest level)

How Common is Healthy Lifestyle?

- In the person-years of follow-up among 47,927 men in HPFS between 1986-1996
  - 18% had a risk score $\leq$ 14
  - 6% had a risk score $\leq$ 11
  - 3% had no risk factors
    - PAR% for CRC was from these risk factors was 71% (95% CI = 33-92%).

  - Similar proportions were found in the 1,724,234 person-years of follow-up among 81,092 women in NHS between 1986-2010 NHS, where 20% of the follow-up corresponded to a risk score of 0 or 1 (Erdrich et al. Proportion of colon cancer attributable to lifestyle in a cohort of US women. Cancer Causes Control, 2015)

Survey of E-values in Nutritional Epidemiology

- Survey of E-values in nutritional epidemiology (Ludovic Trinquart et al., American Journal of Epidemiology, 2019)
- Surveyed 100 nutritional epidemiology studies with significant findings
- The median relative risk reported was 1.33 (referent matters)

Survey of E-values in Nutritional Epidemiology

- The median E-value corresponding to these studies was 2.
  - **If 50% of the population had an unhealthy lifestyle**, and, approximately,
  - If unhealthy lifestyle increased the risk of outcomes by at least two-fold **AFTER** adjusting for measured health lifestyle related risk factors, **AND**
  - If those in the high-risk exposure group were twice as likely to also have an unhealthy lifestyle than those in the low-risk group **AFTER** adjusting for the association of the measured risk factors with the exposure
  - In general, the bias due to confounding is bounded by $\frac{RR_{DU}\,RR_{EU}}{RR_{DU}+RR_{EU}-1}$
  - Then, the entire effect of the exposure would be due to residual confounding.

Survey of E-values in Nutritional Epidemiology

- Authors concluded that little to moderate unmeasured confounding could explain away most of the observed associations.

- I don’t agree with this conclusion

The E Value (Vanderweele and Ding, 2017)

- For an observed risk ratio, RR_{ED}^{obs}

$$\text{E-value} = RR_{ED}^{obs} + \sqrt{RR_{ED}^{obs} \times \left(RR_{ED}^{obs} - 1\right)}$$

  - The formula applies to a risk ratio greater than 1; for a risk ratio less than 1, one first takes the inverse of the observed risk ratio and then applies the formula.
- E-value is very easy to calculate, and applies very generally
- Applies to all sorts of models, outcomes, exposures

VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. Ann Intern Med. 2017 Aug 15;167(4):268-274.

The E-Value+: considering confounder prevalence

- We can use Schlesselman's (1979) formula to calculate the E-value+ without making the assumption of 50% confounder prevalence

$$Relative\ Bias(RR_{ED}^{obs}) = \frac{RR_{DU}\,RR_{EU}\,p_U + (1 - RR_{EU}\,p_U)}{RR_{DU}\,p_U + (1 - p_U)}$$, where

  - RR_{DU} = relative risk for the outcome, D, in relation to the confounder, U
  - RR_{EU} = the probability of the exposure, E, given $U = 1$, over the same when $U = 0$, quantifying the magnitude of the confounder-exposure association
  - p_U = the confounder prevalence

- This formula is similar to Vanderweele and Ding's E-value formula for binary outcome, confounder, and exposure, when there is no effect modification, and when confounder prevalence is 50%.
- This formula can be used to write the E-value in terms of RR_{EU}, RR_{DU}, and p_U. We will call this E-value+.

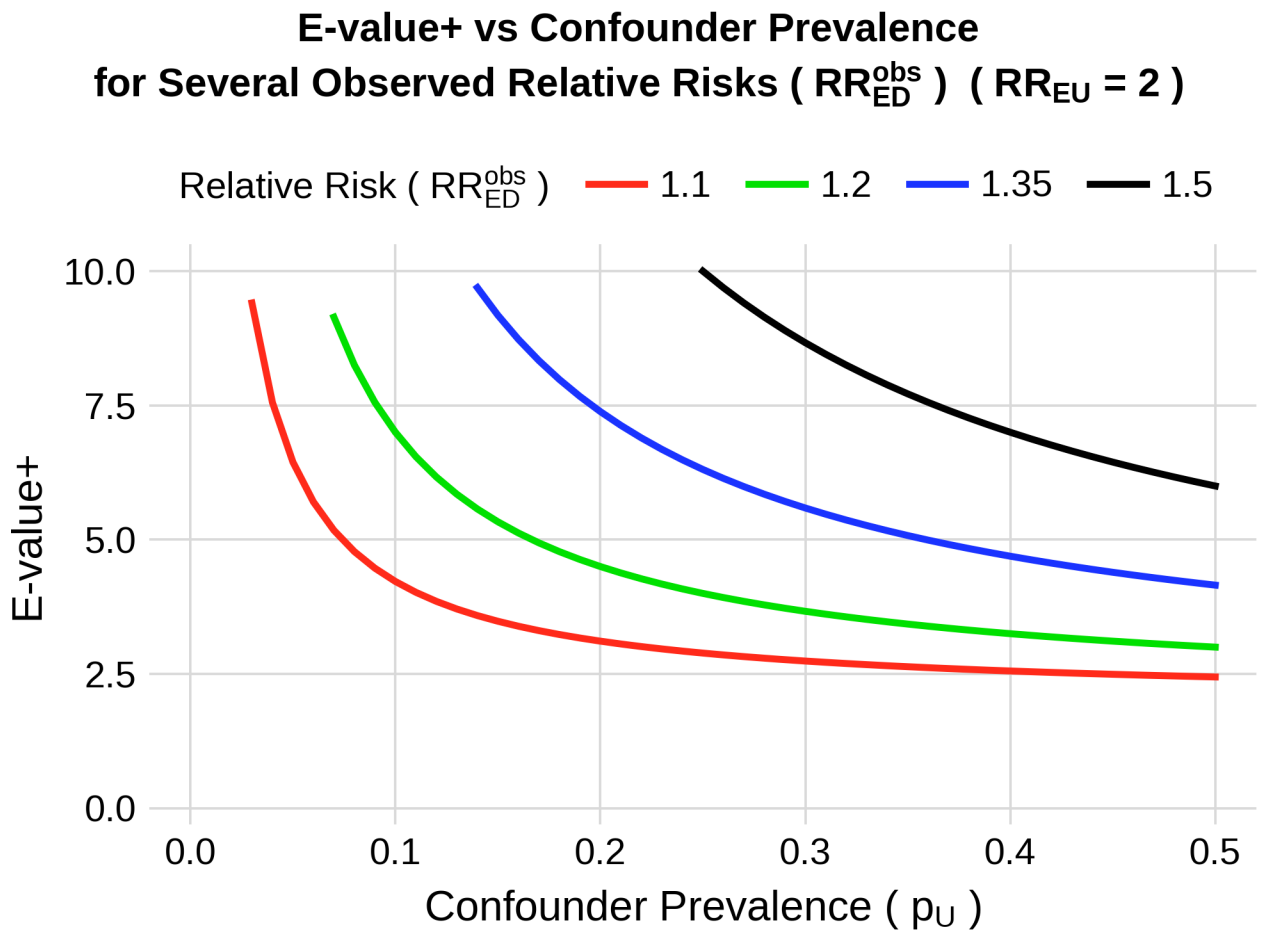
The E-value+ Considering confounder prevalence

- We used this expression to solve for the value of $RR_{DU} \times RR_{EU}$ that arises from the E-value+ to explain away the RR_{ED}^{obs} for a given p_U, and then compare this to the extent to which the E-value underestimates this.

- At confounder prevalence of 10% and $RR_{EU} = 2$, using the E-value+ formula on the previous slide, the confounder-outcome association, RR_{DU}, would have to be at least 6.4 to see an observed RR_{ED} of 1.35. The E-value+ for this scenario is 12.8 and the E-value is 2.04.

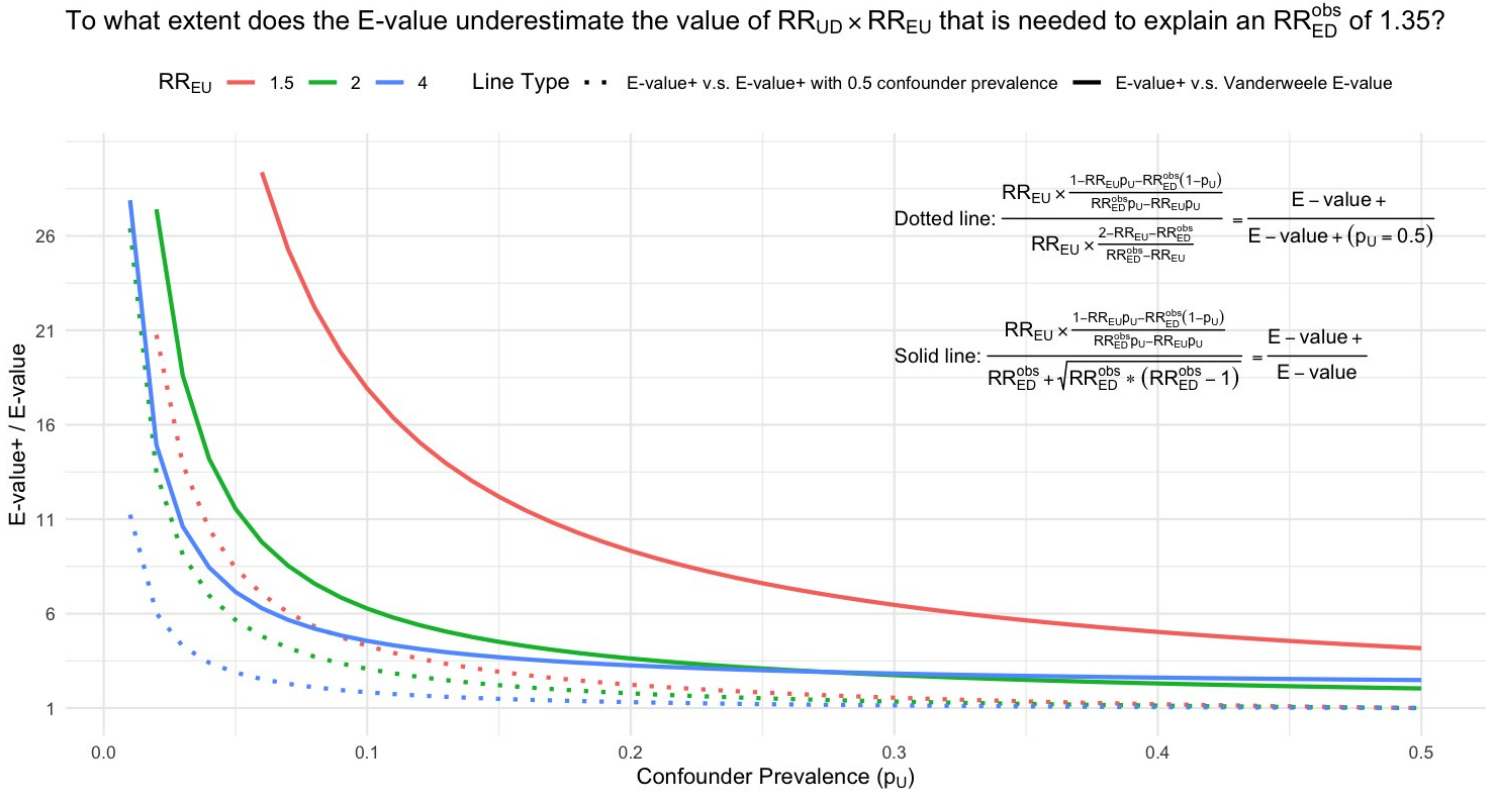
- At confounder prevalence of 20% and $RR_{EU} = 2$, using the E-value+ formula on the previous slide, the confounder-outcome association, RR_{DU}, would have to be at least 3.7 to see an observed RR_{ED} of 1.35. The E-value+ for this scenario is 7.4 and the E-value is 2.04.

The E-Value+ – Considering confounder prevalence



- The E-value+ has to be bigger and bigger as confounder prevalence gets smaller to explain away observed associations
- As the observed RR increases, the E-value+ has to be larger to explain away observed associations

To what extent is does the E-value under-estimates the true associations needed to explain away confounding?



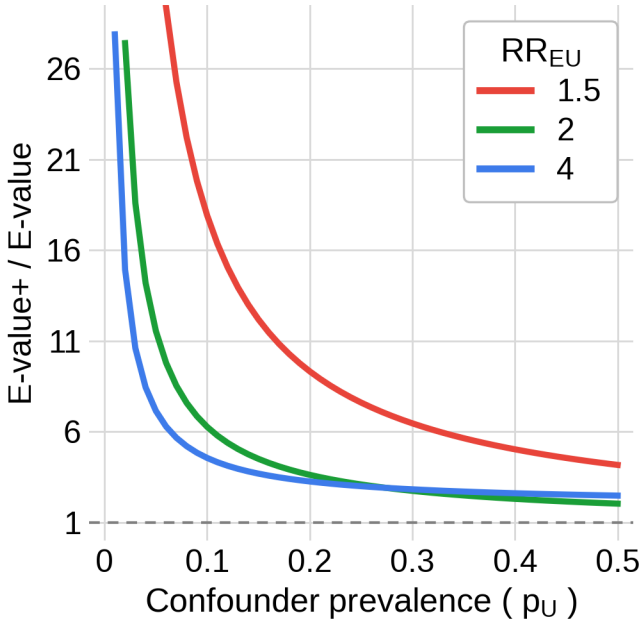
The E-value under-estimates the E-value+ dramatically as confounder prevalence goes towards 0. falsely suggesting the observed association can be explained away by residual and unmeasured confounding at much lower levels of association than is really required.

To what extent does the E-value underestimate the value of $RR_{UD} \times RR_{EU}$ needed to explain an RR_{ED}^{obs} of 1.35?

A. vs. VanderWeele E-value

$$\frac{RR_{EU} \times \frac{1 - RR_{EU}p_U - RR_{ED}^{obs}(1-p_U)}{RR_{ED}^{obs}p_U - RR_{EU}p_U}}{RR_{ED}^{obs} + \sqrt{RR_{ED}^{obs}(RR_{ED}^{obs}-1)}}$$

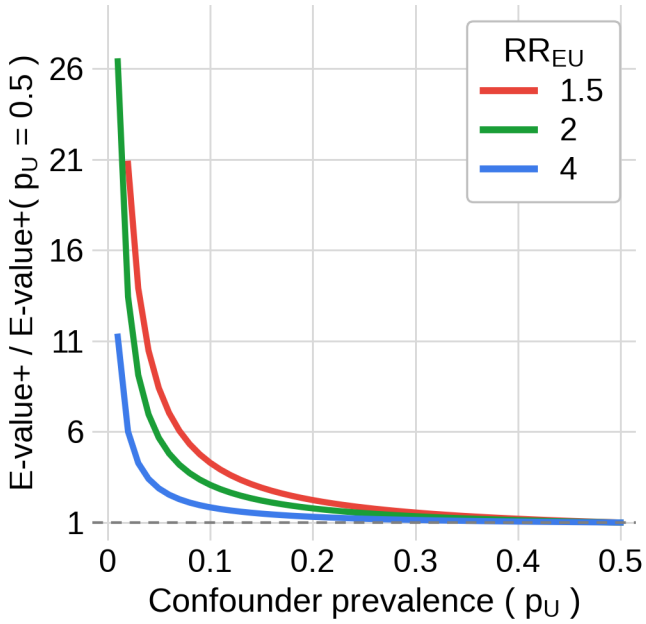
= E-value+ / E-value



B. vs. E-value+ at $p_U = 0.5$

$$\frac{RR_{EU} \times \frac{1 - RR_{EU}p_U - RR_{ED}^{obs}(1-p_U)}{RR_{ED}^{obs}p_U - RR_{EU}p_U}}{RR_{EU} \times \frac{2 - RR_{EU} - RR_{ED}^{obs}}{RR_{ED}^{obs} - RR_{EU}}}$$

= E-value+ / E-value+ ($p_U = 0.5$)



The E-value under-estimates the E-value+ dramatically as confounder prevalence goes towards 0, falsely suggesting the observed association can be explained away by residual and unmeasured confounding at much lower levels of association than is really required.

We have established

- Empirical evidence shows that a plausible surrogate for residual confounding by lifestyle has no effect on CVD incidence, suggesting little residual confounding by lifestyle
- Tight theoretical bounds on bias due to residual confounding makes it implausible that residual confounding could explain away diet-disease associations
  - The most likely residual confounder, healthy lifestyle, is uncommon, thereby dramatically increasing the properly calculated E-value
  - Even an E-value of 2 corresponds to an E-value+ of 6 or more with prevalence of a joint residual confounder of 20% or lower.
  - If the magnitude of baseline confounding is tightly bounded, residual confounding by time-varying confounders conditional on their baseline values is going to be even more tightly restricted. This is why 'causal methods' to account for time-varying confounding in the presence of time-varying exposures rarely materially change scientific findings.
- Confounding is usually not a major source of bias in epidemiologic research

the tail wags the dog



What has happened

Standard methods for multivariate adjustment for confounding have been replaced with ‘causal’ methods such as non-parametric CATEs and ATEs, propensity-score matched estimators, and inverse-probability of exposure/treatment weighted estimators. These offer no advantage for causal inference in the absence of time-varying treatment with time-varying confounders or unless effect modification is present (CATE, ATE only), and result in a loss of efficiency relative to standard multivariate adjustment. Plus, a huge loss of transparency.

the dog wags the tail



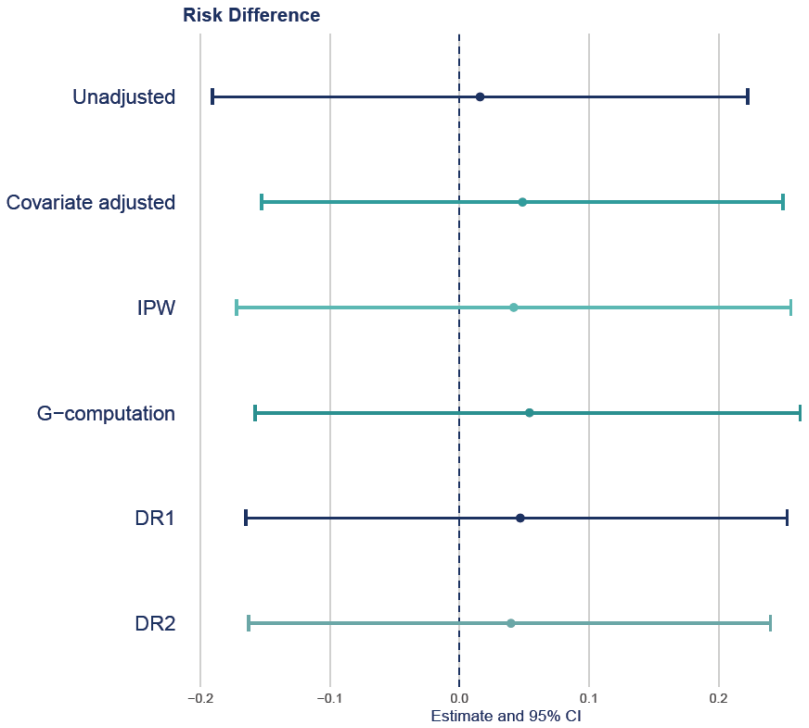
What should happen

Use standard methods for multivariate adjustment unless there is a time-varying intervention with time-varying confounders. Even then, expect very little change in results from the standard methods which are formally biased **in this context alone.**

Causal inference methods for analysis add no value unless there is time-varying treatments and time-varying confounding

Spiegelman D, Zhou X. Evaluating Public Health Interventions: 8. Causal Inference for Time-Invariant Interventions. *Am J Public Health*. 2018 Sep;108(9):1187–90.

- This is why epidemiology nearly always gets the right answer!



Listing and debunking the assumptions one by one

- ~~no interference~~
- ~~positivity~~
- ~~one version of treatment/consistency~~
- ~~no causation without manipulation~~
- ~~exchangeability = no confounding~~

So what's going on?

There are two additional sources of bias discussed in fundamental epidemiology texts: selection bias and information bias.

If there are major biases lurking in epidemiologic, public health and clinical studies, it must be due to violations of one of these assumptions.

Selection bias unlikely to be a major source of bias in observational research

- In multiplicative models, bias cancels unless the selection probabilities depend jointly and non-multiplicatively on exposure and outcome, after controlling for each one separately (**Kleinbaum DG, Kupper LL, Morgenstern H.** *Epidemiologic Research: Principles and Quantitative Methods*. Belmont, CA: Lifetime Learning Publications; 1982, Chapter 11)
- In multiplicative models, selection factors can be controlled for just as confounders
- In the absence of effect modification, which I've already shown is rare, results are automatically transportable to any other study population
- Additive measures of effect, such as risk differences, will rarely be transportable because they confound baseline risk with the intervention effect.
- These points can be further elaborated for sure

Measurement error!

- Boundless
- Pervasive
- Non-adherence in randomized studies is a type of measurement error (‘random’ non-adherence currently overlooked in the causal inference community)
- A huge toolkit of methods exist to adjust for measurement error bias in all sorts of statistical models using main study/validation study and main study/reliability study designs
- Software is available although needs strengthening

The boundlessness of measurement error

Random measurement error

$$Z = X + e \qquad \rho = Corr(X, Z)$$

$$\text{True OR} = (\text{Observed OR})^{1/\rho}$$

Expected true OR for selected observed OR values (for 1 SD change in X)

Observed OR	True OR at ρ = 0.3	True OR at ρ = 0.5	True OR at ρ = 0.7	True OR at ρ = 0.9
1.1	1.37	1.21	1.15	1.11
1.2	1.84	1.44	1.30	1.22
1.5	3.86	2.25	1.78	1.57
2.0	10.08	4.00	2.69	2.16

Armstrong BK, White E, Saracci R. Principles of Exposure Measurement in Epidemiology. Oxford University Press, 1992.

Common Approaches to the Measurement Error Problem in Epidemiology

1. Sensitivity Analysis
   - Non-differential error bias toward the null - but not always

2. Validation of exposure measurement
   - Reporting a statistically significant association (through p-value testing) between true and observed does not imply no bias due to measurement error.

3. Quantitative correction of effect estimate
   i. DESIGN: Main study/ **reliability or validation study**
   ii. ANALYSIS: measurement error & misclassification methods

Epidemiologists have conducted hundreds of validation and reliability studies

[56]M. G. Jain, T. E. Rohan, C. L. Soskolne, and N. Kreiger. Calibration of the dietary questionnaire for the Canadian Study of Diet, Lifestyle and Health cohort. Public Health Nutr, 6(1):79–86, 2003.
[57]B. R. Khani, W. Ye, P. Terry, and A. Wolk. Reproducibility and validity of major dietary patterns among Swedish women assessed with a food-frequency questionnaire. J Nutr, 134(6):1541–5, 2004.
[58]M. Messerer, S. E. Johansson, and A. Wolk. The validity of questionnaire-based micronutrient intake estimates is increased by including dietary supplement use in Swedish men. J Nutr, 134(7):1800–5, 2004.
[59]R. G. Munger, A. R. Folsom, L. H. Kushi, S. A. Kaye, and T. A. Sellers. Dietary assessment of older Iowa women with a food frequency questionnaire: nutrient intake, reproducibility, and comparison with 24-hour dietary recall interviews. Am J Epidemiol, 136(2):192–200, 1992.
[60]M. L. Neuhouser, A. R. Kristal, D. McLerran, R. E. Patterson, and J. Atkinson. Validity of short food fre- quency questionnaires used in cancer chemoprevention trials: results from the Prostate Cancer Prevention Trial. Cancer Epidemiol Biomarkers Prev, 8(8):721–5, 1999.
[61]R. E. Patterson, A. R. Kristal, L. F. Tinker, R. A. Carter, M. P. Bolton, and T. Agurs-Collins. Measurement characteristics of the Women's Health Initiative food frequency questionnaire. Ann Epidemiol, 9(3):178– 87, 1999.
[62]P. Pietinen, A. M. Hartman, E. Haapa, L. Rasanen, J. Haapakoski, J. Palmgren, D. Albanes, J. Virtamo, and J. K. Huttunen. Reproducibility and validity of dietary assessment instruments. I. A self-administered food use questionnaire with a portion size picture booklet. Am J Epidemiol, 128(3):655–66, 1988.
[63]E. Riboli, P. Toniolo, R. Kaaks, R. E. Shore, C. Casagrande, and B. S. Pasternack. Reproducibility of a food frequency questionnaire used in the New York University Women's Health Study: effect of self-selection by study subjects. Eur J Clin Nutr, 51(7):437–42, 1997.
[64]S. Salvini, D. J. Hunter, L. Sampson, M. J. Stampfer, G. A. Colditz, B. A. Rosner, and W. C. Willett. Food- based validation of a dietary questionnaire: the effects of week-to-week variation in food consumption. Int J Epidemiol, 18(4):858–67, 1989.
[65]S. Sasaki, M. Kobayashi, S. Tsugane, and Jphc. Validity of a self-administered food frequency question- naire used in the 5-year follow-up survey of the JPHC Study Cohort I: comparison with dietary records for food groups. J Epidemiol, 13(1 Suppl):S57–63, 2003.
[66]X. O. Shu, G. Yang, F. Jin, D. Liu, L. Kushi, W. Wen, Y. T. Gao, and W. Zheng. Validity and reproducibility of the food frequency questionnaire used in the Shanghai Women's Health Study. Eur J Clin Nutr, 58(1):17– 23, 2004.
[67]D. O. Stram, J. H. Hankin, L. R. Wilkens, M. C. Pike, K. R. Monroe, S. Park, B. E. Henderson, A. M. Nomura,
M. E. Earle, F. S. Nagamine, and L. N. Kolonel. Calibration of the dietary questionnaire for a multiethnic cohort in Hawaii and Los Angeles. Am J Epidemiol, 151(4):358–70, 2000. PMCID: PMC4482461.
[68]A. F. Subar, F. E. Thompson, V. Kipnis, D. Midthune, P. Hurwitz, S. McNutt, A. McIntosh, and S. Rosenfeld. Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires : the Eating at America's Table Study. Am J Epidemiol, 154(12):1089–99, 2001.

....

Empirical correction for measurement error in exposures and confounders

Good news:

– with these validation or reliability data, hundreds of publications provide a robust, flexible set of methods to empirically correct for measurement error in exposures of interest and confounders simultaneously.

Empirical correction for measurement error in exposures and confounders

- Three key papers from our group:
  - Rosner, B., Willett, W.C. and Spiegelman, D. (1989), Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. Statist. Med., 8: 1051-1069.
  - Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. Am J Epidemiol. 1990 Oct;132(4):734-45.
  - Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. Am J Epidemiol. 1992
- **1,786** citations as of 4 months ago, ~84% of citations in non-statistical journals
- Other work covers survival data analysis, population attributable risks, principles of modeling, design of MS/VS, misclassification, risk prediction, changepoint analysis, and a wide range of other topics.
  - Other major author groups include Ray Carroll, Ross Prentice, Grace Yi, and their colleagues and academic descendants
- Software available here https://ysph.yale.edu/cmips/research/software/; R library is growing (see Molin Wang's website)

How Much Cancer Is Attributable to Environmental Causes?

Variation by country at the same time:

Population:	Median attributable fraction
Compared with the 8 sites where Nigeria is lowest: Lung, corpus uteri, colon, rectum, larynx, prostate, breast, pancreas	
US Black	81%
US White	82.5%
US overall	**81%**
Compared with the 7 sites in Japan is lowest: Colon, rectum, lung, prostate, breast, corpus uteri, ovary	
Hawaii Japanese	74%
Hawaii Caucasians	81%

Population attributable risk of modifiable risk factors for colon cancer

	PAR % (95% CI)
Women · Nurses' Health Study (1,127 cases / 81,092; 24 y)	
Seven measured modifiable risk factors: BMI < 25 kg/m^2; physical activity ≥ 21 MET per week; alcohol ≤ 30 g/day; < 10 pack-years of smoking before age 30; current multivitamin use ≥ 15 years; total calcium ≥ 700 mg/day. [b] Two tablets of aspirin per week for six or more years.	43% (14-65%)
Men · Health Professionals Follow-up Study	
≥ 1 risk factor vs. none (3.1%)	**71 (33–92)**

Doll and Peto's (1981) conclusion: roughly 75–80% of US cancer is avoidable through elimination of modifiable causes

Sources. Doll R, Peto R. The causes of cancer. *J Natl Cancer Inst*. 1981;66(6):1191–1308 (Tables 3, 4).
Erdrich J, Zhang X, Giovannucci E, Willett W. Proportion of colon cancer attributable to lifestyle in a cohort of US women. *Cancer Causes Control*. 2015;26(9):1271–79.
Platz EA, Willett WC, Colditz GA, Rimm EB, Spiegelman D, Giovannucci E. Proportion of colon cancer risk that might be preventable in a cohort of middle-aged US men. *Cancer Causes Control*. 2000;11(7):579–88.

Do we know what we need to know to nearly fully prevent excess colon cancer? Yes!

1 Low-risk definitions in two US cohorts

Factor	Women · NHS *81,092 women · 1,127 cases* *1986–2010 (24 y)*	Men · HPFS *47,927 men · 411 cases* *ages 40–75 · 1986–96 (10 y)*
Body mass index	< 25 kg/m²	≤ 25 kg/m²
Physical activity	≥ 21 MET-h/week	≥ 15 MET-h/week
Alcohol	≤ 30 g/day	< 15 g/day, not former drinker[b]
Smoking, early adult	< 10 pack-yrs before age 30	≤ 3 pack-years
Multivitamin use	current use, ≥ 15 years	—
Total calcium intake	≥ 700 mg/day	—
Red meat	—	≤ 2 servings/week
Folic acid, supplemental	—	≥ 100 µg/day
Regular aspirin[a]	≥ 2 tablets/wk, ≥ 6 years	not in score
Observed PAR	**43%** (14–65) all 7 factors	**71%** (33–92) ≥ 1 risk factor vs. none[c] 39–55% by score percentile

2 Correcting the PAR for score validity

$$AF^* = 1 - (1 - AF)^{1/\rho}$$ (Levin, 1953)

AF observed PAR · **ρ** validity: corr(measured score, true exposure) · **AF*** corrected PAR. Reference = no exposure, so the prevalence term drops out.

Validity ρ	1/ρ	NHS · observed 43%		HPFS · observed 71%	
		RR*	PAR*	RR*	PAR*
0.5 *base case*	2.00	3.08	**68%**	11.9	**92%**
0.4	2.50	4.08	**75%**	22.1	**96%**
0.3	3.33	6.51	**85%**	62.0	**98%**

RR de-attenuated RR · PAR* corrected PAR*

ρ drives the correction, not the prevalence. Lowering ρ from 0.5 to 0.3 moves NHS from 68% to 85% and HPFS from 92% to 98%.

NHS reaches Doll & Peto's 75–80% at ρ = 0.4; HPFS exceeds it at every ρ — but ρ is assumed, not estimated: a sensitivity analysis, not a result.

For orientation: NHS FFQ validation gave correlations of 0.63 (total calcium) and 0.84 (alcohol).

Sources. Erdrich J, Zhang X, Giovannucci E, Willett W. Proportion of colon cancer attributable to lifestyle in a cohort of US women. *Cancer Causes Control*. 2015;26(9):1271–1279.
Platz EA, Willett WC, Colditz GA, Rimm EB, Spiegelman D, Giovannucci E. Proportion of colon cancer risk that might be preventable in a cohort of middle-aged US men. *Cancer Causes Control*. 2000;11(7):579–588.
Levin ML. The occurrence of lung cancer in man. *Acta Unio Int Contra Cancrum*. 1953;9(3):531–541.

Wong BHW, Peskoe SB, Spiegelman D. *Stat Med*. 2018;37(8):1259–1275.
Wong BHW, Lee J, Spiegelman D, Wang M. *Biostatistics*. 2021;22(4):805–818.

Do we know what we need to know to nearly fully prevent excess breast and prostate cancer? Yes!

1 Low-risk profiles in two US cohorts

Factor	Breast · NHS women *8,421 postmenopausal cases* *1980–2010*	Lethal prostate · HPFS men *42,701 men · 576 lethal cases* *1986–2010*
Body weight	adult gain < 2 kg[a]	BMI < 30 kg/m²
Physical activity	top quartile, MET-h/week	≥ 3 h/wk vigorous or ≥ 7 h/wk walking
Alcohol	none	—
Hormone therapy	not current use	—
Breastfeeding	ever[b]	—
Smoking	—	never, or quit ≥ 10 years
Tomatoes	—	≥ 7 servings/week
Fatty fish	—	≥ 1 serving/week
Processed meat	—	< 3 servings/week
Observed PAR	**34.6%** (22.7–45.4) ER+ 39.7% · ER− 27.9%	**46.7%** ≥ 5 vs. 0–4 factors US men > 60 (NHANES)[c]
Top single factor	Weight gain **18.7%**	Vigorous activity **33.7%**

2 Correcting the PAR for score validity

$$AF^* = 1 - (1 - AF)^{1/\rho}$$ (Levin, 1953)

AF observed PAR · **ρ** validity: corr(measured score, true exposure) · **AF*** corrected PAR. Reference = no exposure, so the prevalence term drops out.

Validity ρ	1/ρ	Breast · observed 35%		Prostate · observed 47%	
		RR*	PAR*	RR*	PAR*
0.5 *base case*	2.00	2.34	**57%**	3.52	**72%**
0.4	2.50	2.89	**65%**	4.83	**79%**
0.3	3.33	4.12	**76%**	8.16	**88%**

RR de-attenuated RR, not low-risk vs. low-risk · PAR* corrected PAR*

Sources. Tamimi RM, Spiegelman D, Smith-Warner SA, Wang M, Pazaris M, Willett WC, Eliassen AH, Hunter DJ. Population attributable risk of modifiable and nonmodifiable breast cancer risk factors in postmenopausal breast cancer. *Am J Epidemiol*. 2016;184(12):884–893.
Kenfield SA, Batista JL, Jahn JL, et al. Development and application of a lifestyle score for prevention of lethal prostate cancer. *J Natl Cancer Inst*. 2016;108(3):djv329.

Levin ML. The occurrence of lung cancer in man. *Acta Unio Int Contra Cancrum*. 1953;9(3):531–541.

Wong BHW, Peskoe SB, Spiegelman D. *Stat Med*. 2018;37(8):1259–1275.
Wong BHW, Lee J, Spiegelman D, Wang M. *Biostatistics*. 2021;22(4):805–818.

The case is made for implementation science

Take-Home Messages



- Positivity limitations are largely artifacts of specific estimators such as MSMs
- Interference should often be embraced, not excluded
- Consistency should accommodate realistic treatment variation
- Confounding/exchangeability needs to be considered, but randomization and standard multivariate adjustment largely addresses it
- Effect modification (aka treatment heterogeneity) is exceptionally rare on the multiplicative scale – analyses should lead assuming it is not present
- Causes do not have to be manipulable
- Measurement error correction is essential for valid (causal) inference in epidemiology, EHR-based research, public health research, implementation science, and usually in clinical research
- Causal inference theory must remain grounded in substantive scientific reality
  - **Please show results using standard methods and demonstrate your new approach actually adds something**

We know enough right now to prevent and treat the major causes of disease and mortality in the world – no need to find new causes.

Classic and new statistical methods for design and analysis can be pivoted to determine why these interventions are not being used and get them out there, broadly and sustainably.

Thank you to Bhramar Mukherjee and Yale Public Health Data Science and Data Equity for supporting my work and encouraging me to refine and disseminate my thinking



Thank you to my many colleagues, students, and post-docs over the years, from around the world, and especially at Yale and Harvard, all 600+ of you, working with me on all of these amazing projects

the dog wags the tail



What should happen

Let's renew our focus on what matters!

- External generalizability is essential for transporting causal exposure effects beyond that estimated in the study population, and in the presence of effect modification, variation in distributions of modifiers, otherwise not
- External generalizability depends on similarity of effect measure modifier distributions, if any, otherwise results are immediately generalizable

- Purveyors of doubt
- **Closest to "purveyors of doubt" in health**
- **David Michaels,** *Doubt Is Their Product: How Industry's Assault on Science Threatens Your Health* (2008), and *The Triumph of Doubt* (2020). Michaels is an epidemiologist and former head of OSHA, and the 2008 book is probably the most directly on point.
- **Naomi Oreskes and Erik Conway,** *Merchants of Doubt* (2010). It starts with tobacco and extends to acid rain, the ozone hole, and climate change.
- **Robert Proctor,** *Golden Holocaust* (2011), on the tobacco industry, and the edited volume *Agnotology: The Making and Unmaking of Ignorance* (2008, with Londa Schiebinger). Proctor coined "agnotology," the study of culturally produced ignorance.
- **Industry-specific histories**
- **Allan Brandt,** *The Cigarette Century* (2007), on tobacco.
- **Gerald Markowitz and David Rosner,** *Deceit and Denial* (2002) and *Lead Wars* (2013), on the lead and chemical industries.
- **Devra Davis,** *When Smoke Ran Like Water* (2002) and *The Secret History of the War on Cancer* (2007).
- **Marion Nestle,** *Unsavory Truth* (2018), on food-industry-sponsored research.
- **Sheldon Krimsky,** *Science in the Private Interest* (2003), on commercial influence on research.
- **Related**
- **Michael Specter,** *Denialism* (2009), and **Paul Offit,** *Deadly Choices* (2011), on vaccine and medical misinformation.
- **Ben Goldacre,** *Bad Pharma* (2012), on distortions in pharmaceutical research.

- Stochastic vs deterministic potential outcomes
- International variation in rates, vs. low R^2 of models

Consistency revisited, relative to exchangeability

- I don't see how there can be consistency in the sense of

$$Y_i^{a = A_i} = Y_i^{obs}, \text{ for all } i$$

without there being no confounding.

- That is, how can the observed outcome for the observed treatment equal the counterfactual potential outcome for that treatment unless there is no confounding?
  - If you agree, there is a component of the consistency assumption that overlaps the definition of no confounding and is essential.
- No confounding could be a necessary but not sufficient condition to ensure consistency, since multiple versions of the treatment can also lead to the non-equivalence between the observed outcome at a given treatment level and the potential outcome at that level.