Introducing the specificity score: a measure of causality beyond P value

苗旺

北京大学概率统计系，统计科学中心

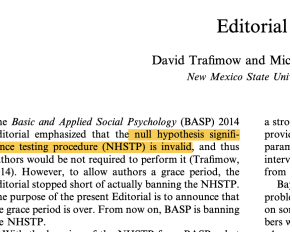
mwfy@pku.edu.cn

May 13, 2023

https://www.math.pku.edu.cn/teachers/mwfy

Wang Miao (2023). Specificity analysis for causal inference in observational studies.

The P-value crisis

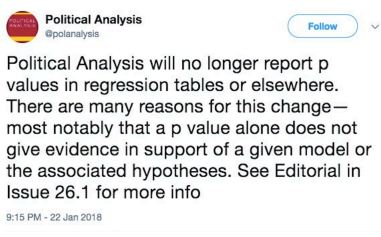


Editorial

David Trafimow and Mic
New Mexico State Uni

The *Basic and Applied Social Psychology* (BASP) 2014 Editorial emphasized that the null hypothesis significance testing procedure (NHSTP) is invalid, and thus authors would be not required to perform it (Trafimow, 2014). However, to allow authors a grace period, the Editorial stopped short of actually banning the NHSTP. The purpose of the present Editorial is to announce that the grace period is over. From now on, BASP is banning the NHSTP.

a stro
provi
paran
interv
from
Ba
probl
on so
bers v



Political Analysis
@polanalysis

Follow

Political Analysis will no longer report p values in regression tables or elsewhere. There are many reasons for this change—most notably that a p value alone does not give evidence in support of a given model or the associated hypotheses. See Editorial in Issue 26.1 for more info

9:15 PM - 22 Jan 2018

nature

Explore content ∨ About the journal ∨ Publish with us ∨ Subscribe

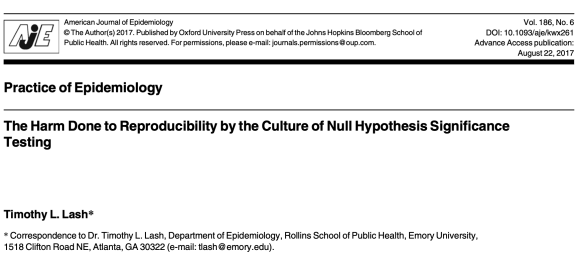
nature › comment › article

COMMENT | 20 March 2019

Scientists rise up against statistical significance

Valentin Amrhein, Sander Greenland, Blake McShane and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects.

Valentin Amrhein ✉, Sander Greenland & Blake McShane



American Journal of Epidemiology
© The Author(s) 2017. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

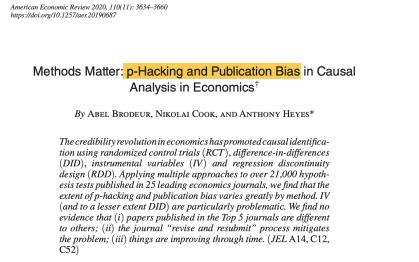
Vol. 186, No. 6
DOI: 10.1093/aje/kwx261
Advance Access publication: August 22, 2017

Practice of Epidemiology

The Harm Done to Reproducibility by the Culture of Null Hypothesis Significance Testing

Timothy L. Lash*

* Correspondence to Dr. Timothy L. Lash, Department of Epidemiology, Rollins School of Public Health, Emory University, 1518 Clifton Road NE, Atlanta, GA 30322 (e-mail: tlash@emory.edu).



American Economic Review 2020, 110(11): 3634–3660
https://doi.org/10.1257/aer.20190687

Methods Matter: p-Hacking and Publication Bias in Causal Analysis in Economics†

By Abel Brodeur, Nikolai Cook, and Anthony Heyes*

The credibility revolution in economics has promoted causal identification using randomized control trials (RCT), difference-in-differences (DID), instrumental variables (IV) and regression discontinuity design (RDD). Applying multiple approaches to over 21,000 hypothesis tests published in 25 leading economics journals, we find that the extent of p-hacking and publication bias varies greatly by method. IV (and to a lesser extent DID) are particularly problematic. We find no evidence that (i) papers published in the Top 5 journals are different to others; (ii) the journal "revise and resubmit" process mitigates the problem; (iii) things are improving through time. (JEL A14, C12, C52)

BASP, Political Analysis, Nature against statistical significance. AER (Brodeur et al., 2020), AJE(Lash, 2017) worried about publication bias and reproducibility crisis.

The American Statistician 2019 Special Issue (43 papers) discussed "Statistical Inference in the 21st Century: A World Beyond $p < 0.05$".

Gelman and Loken (2016) described as "The Statistical Crisis in Science".

Responses to the crisis

- Banish the P Value.
- ASA statement (Wasserstein and Lazar, 2016): 6 principles for the proper use and interpretation of the P value. This is the first time that the 177-year-old ASA has made explicit recommendations on such a foundational matter in statistics.
- Supplement and replacement approaches: confidence, credibility, or prediction intervals; Bayesian methods; alternative measures of evidence, e.g., Bayes Factors; and false discovery rates.
- P-value thresholds lowered to 0.005 for the social and biomedical sciences (Chawla, 2017, Nature Human Behaviour).

Critics on P value

- Misuse and misunderstand.
- P-values are not adjusted for the effects of the search: P-hacking, model mining, publication bias, and reproducibility crisis.
- Statistical significance does not mean causality.
- The significance depends on sample size.

Gill (2018, Political Analysis Editorial): Most notably that in isolation a P value simply does not give adequate evidence in support of a given model or the associated hypotheses.

Causality, bias and chance

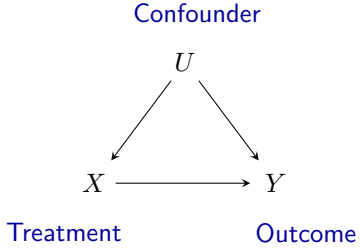
P-value: The probability of the observed data (or data showing a more extreme departure from the null hypothesis) when the null hypothesis is true.

$$\text{Data law} \Leftarrow \underbrace{\text{Causality (Truth)} + \text{Bias}}_{\text{Association (Spurious)}} + \text{Chance}$$

Freedman (1999): "P is about sampling error, not bias."

Bias in causal inference

Confounding bias: Unmeasured common causes of the treatment and outcome; Simpson (1951) paradox; endogeneity.



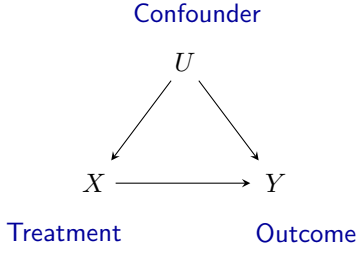
$$Y = \beta X + U + \varepsilon; \quad U \not\perp\!\!\!\perp X.$$

Examples: Genes, cancer, batch effects; education, income, ability;

Other bias: Selection bias; misspecification bias; measurement error;

Bias in causal inference

Confounding bias: Unmeasured common causes of the treatment and outcome; Simpson (1951) paradox; endogeneity.



$$Y = \beta X + U + \varepsilon; \quad U \not\perp\!\!\!\perp X.$$

Examples: Genes, cancer, batch effects; education, income, ability;

Other bias: Selection bias; misspecification bias; measurement error;

Pearl and Mackenzie (2018):"...the complete solution of the confounding problem one of the main highlights of the Causal Revolution because it ended an era of confusion that has probably resulted in many wrong decisions in the past."

Causal criteria

What aspects should we especially consider in order to separate causation from association?

Sir Austin Bradford Hill (1897-1991)

Sir Austin Bradford Hill

Austin Bradford Hill was a British epidemiologist best known for his collaborative research with Richard Doll which linked smoking with cancer and other serious diseases. He was also acknowledged by his peers as the world's leading medical statistician and was a pioneer in the use of randomised clinical trials.

A pilot during the First World War, Hill contracted tuberculosis while on service. The illness and his long recovery from it put paid to his intended medical career and he studied economics instead. In the interwar period he was able to combine these two interests through research in epidemiology and statistics, originally with the Industry Fatigue Research Board. This work included important studies on the health of printers, cotton workers and other occupational groups. At the end of the 1930s he also produced a book,

50 years ago in JRSM

The Royal Society of Medicine

Journal of the Royal Society of Medicine; 2015, Vol. 108(1) 32–37
DOI: 10.1177/0141076814562718

The environment and disease: association or causation?

Sir Austin Bradford Hill
Professor Emeritus of Medical Statistics, University of London, UK

This article was first published by JRSM in Volume 58 issue 5, May 1965. Our full back archive is available online at jrs.sagepub.com.

Among the objects of this newly founded Section of Occupational Medicine are: first, 'to provide a means, not readily afforded elsewhere, whereby physicians and surgeons with a special knowledge of the relationship between sickness and injury and conditions of work may discuss their problems, not only with each other, but also with colleagues in other fields, by holding joint meetings with other Sections of the Society'; and second, 'to make available information about the physical, chemical and psychological hazards of occupation, and in particular about those that are rare or not easily recognized'.

I have no wish, nor the skill, to embark upon a philosophical discussion of the meaning of 'causation'. The 'cause' of illness may be immediate and direct, it may be remote and indirect underlying the observed association. But with the aims of occupational, and almost synonymously preventive, medicine in mind, the decisive question is whether the frequency of the undesirable event B will be influenced by a change in the environmental feature A. How such a change exerts that influence may call for a great deal of research. However, before deducing 'causation' and taking action, we shall not

Hill (1965) nine criteria: Strength; Consistency; Specificity; Temporality; Biological gradient; Plausibility; Coherence; Experiment; Analogy.

These criteria are enormously influential, frequently taught in epidemiology (even more than P value), and widely invoked as a checklist approach to assess causation,

Causal criteria

- Highly subjective and qualitative, without a comprehensive and rigorous causal theory.
- There is considerable debate and counterexamples about the criteria in epidemiology, statistics and medicine. (e.g., Hill, 1965; Rothman and Greenland, 2005; Weiss, 2002; Ioannidis, 2016; Höfler, 2005).
- The most questioned one—Specificity: if the association with an exposure is specific to a particular outcome then this may render a causal conclusion more plausible.

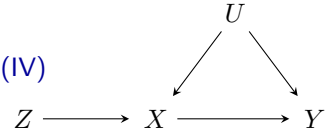
  One of the main means for the US Surgeon General's Advisory Committee (1964) and Doll and Hill (1950) to conclude that smoking causes lung cancer.

  Rothman and Greenland (2005):"the criterion is useless and misleading."

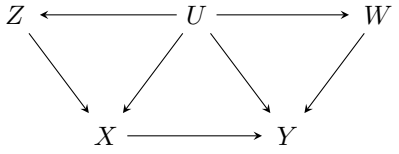
Confounding bias adjustment

- Instrumental variable (Wright, 1928; Goldberger, 1972; Angrist et al., 1996). Nobel Prize in Economic Sciences 2021.

(IV)



(NCE)



(NCO)

- Negative controls/proximal inference (Miao and Tchetgen Tchetgen, 2017; Miao et al., 2018, 2023; Cui et al., 2023; Shi et al., 2020; Li et al., 2022; Ying et al., 2022; Shi et al., 2021; Luo et al., 2022). Rousseeuw Prize for Statistics (1 million dollars).

The principle—the exclusion restriction assumption.

Invoke complete and exact knowledge about the validity of auxiliary variables.

It is more plausibly to have a rough idea on the extent of causal associations.

Notation and setting

$X = (X_1, \ldots, X_K)^{\mathrm{T}}$ a vector of K treatment; $Y = (Y_1, \ldots, Y_P)^{\mathrm{T}}$ a vector of P outcomes; U an unobserved confounder

$X_{\bar{k}}$ the remaining treatments except for X_k

The primary goal is to test the effect of X_1 on Y_1, i.e.,

$$\mathbb{H}_0 : X_1 \perp\!\!\!\perp Y_1 \mid U, X_{\bar{1}}.$$

Note that U is not observed.

Notation and setting

I will use the linear model to ground ideas.

$$\begin{gathered} E(Y \mid X, U) = \beta^{\mathrm{T}} X + \alpha^{\mathrm{T}} U, \\ \beta = (\beta_{kp})_{K \times P}, \quad \alpha = (\alpha_p)_{1 \times P}, \quad \alpha_1 = 1 \text{ and } \alpha_p \neq 0 \text{ for all } p, \end{gathered} \tag{1}$$

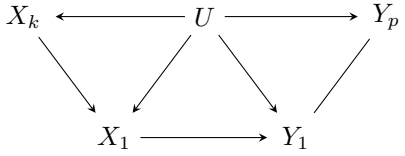
We only have from the observed data

$$\Gamma = (\Gamma_{kp})_{K \times P} = E(XX^T)^{-1} E(XY^{\mathrm{T}}).$$

$$\Gamma = \beta + \delta\alpha, \quad \delta = E(XX^{\mathrm{T}})^{-1} E(XU).$$

Note that (β, α, δ) are unknown.

Suppose we know the specificity



(X_k, Y_p) a pair of negative controls for (X_1, Y_1). (e.g., Miao et al., 2018)

$\Lambda_{kp} = |\Gamma_{k1}\Gamma_{kp}^{-1}\Gamma_{1p}| = |\delta_1\alpha_1|$ retrieves the confounding effect between (X_1, Y_1).

Causal specificity assumption

$\mathcal{X}_p = \{X_k : X_k \not\perp\!\!\!\perp Y_p \mid (U, X_{\bar{k}})\}$ the set of causes of Y_p
$\mathcal{Y}_k = \{Y_p : X_k \not\perp\!\!\!\perp Y_p \mid (U, X_{\bar{k}})\}$ the set of outcomes of X_k
$|\mathcal{X}_p|, |\mathcal{Y}_k|$ the cardinality of $\mathcal{X}_p$ and $\mathcal{Y}_k$

Let $K^* = \max_p(|\mathcal{X}_p|), P^* = \max_k(|\mathcal{Y}_k|)$

Assumption 1 (Causal specificity for testing).

(i) Specificity of causes: $|\mathcal{X}_1| \leq K^*$ for a given $K^* < K - 1$.
(ii) Specificity of effects: for any $k \neq 1$, $|\mathcal{Y}_1 \cup \mathcal{Y}_k| \leq P^*$ for a given $P^* < P - 1$;

- Assumption 1 only requires an upper bound for the number of active causal effects; K^*, P^* represents the strictness of the assumption.
- A sufficient condition for (ii): each treatment can directly affect at most $P^*/2$ outcomes;
- Differences from Hill's specificity criterion.

Testable implications of causal specificity

The causal specificity leads to the following result.

Theorem 1. Let $\Lambda_{kp} = |\Gamma_{k1}\Gamma_{kp}^{-1}\Gamma_{1p}|$ for $k \neq 1, p \neq 1$. Under Model (1) and Assumption 1, if $\mathbb{H}_0$ is correct then

at least $(K-1-K^*)(P-1-P^*)$ entries of Λ_{kp} are equal to $|\Gamma_{11}|$.

- under $\mathbb{H}_0$ the crude effect Γ_{11} is completely attributed to confounding;
- Therefore, one can test the null hypothesis by assessing how far away Λ_{kp} departs from $|\Gamma_{11}|$.

The specificity score

q_1 the proportion of Λ_{kp}'s strictly larger than $|\Gamma_{11}|$; q_2 the proportion of Λ_{kp}'s strictly smaller than $|\Gamma_{11}|$.

The specificity score for testing the causal effect of X_1 on Y_1 is defined as

$$\mathrm{SPC}_{11} \quad = \quad \max(q_1, q_2). \tag{2}$$

- Roughly, a larger specificity score means that a stronger specificity assumption with smaller (K^*, P^*) that observed data could admit under $\mathbb{H}_0$;
- Thus, a larger specificity score means a higher credibility to reject $\mathbb{H}_0$, i.e., the credibility for the causal effect of X_1 on Y_1.

Properties of the specificity score

By definition, $0 \leq \mathrm{SPC}_{11} \leq 1$; Dimensionless. A nontrivial upper bound:

Theorem 2. Under Model (1) and Assumption 1, letting

$$\tau = 1 - \frac{(K - 1 - K^*)(P - 1 - P^*) - 1}{(K - 1)(P - 1)}, \tag{3}$$

if the null hypothesis $\mathbb{H}_0$ is correct, then $\mathrm{SPC}_{11} < \tau$.

This result establishes the basis for constructing valid hypothesis tests given (P^*, K^*).

- The critical value for rejecting $\mathbb{H}_0$ with the specificity score is τ;
- The critical value decreases when K, P increase or K^*, P^* decrease;
- τ approximates zero and the specificity test exhibits a higher power when K^* and P^* are much smaller than K and P, e.g., genetic studies;
- a most conservative critical value $\tau^{\max} = 1$—the premise is that there exist at least one pair of negative controls

Properties of the specificity score

- A formal, qualitative, and objective approach.
- Dimensionless.
- Either small or large association can lead to large SPC; Strength; Publication bias.
- Does not depend on sample size.

Specificity score estimation, specificity test, and specificity map

$$\hat{q}_1 = \frac{\sum_{k,p} \mathbb{1}\{|\hat{\Lambda}_{kp}| > |\hat{\Gamma}_{11}| + (\log(n)/n)^{1/2}\}}{(K-1)(P-1)},$$
$$\hat{q}_2 = \frac{\sum_{k,p} \mathbb{1}\{|\hat{\Lambda}_{kp}| < |\hat{\Gamma}_{11}| - (\log(n)/n)^{1/2}\}}{(K-1)(P-1)}, \tag{4}$$
$$\widehat{\mathrm{SPC}}_{11} = \max(\hat{q}_1, \hat{q}_2).$$

Test: $T_n = 1$ if $\widehat{\mathrm{SPC}}_{11} \geq \tau$ and $T_n = 0$ otherwise. (5)

Specificity score estimation, specificity test, and specificity map

Theorem 3. Suppose $\hat{\Gamma}_{kp}$ and $\hat{\Lambda}_{kp}$ are $n^{1/2}$-consistent, then

(i) Estimation consistency:
$\hat{q}_1, \hat{q}_2, \widehat{\mathrm{SPC}}_{11}$ are consistent estimators of $q_1, q_2, \mathrm{SPC}_{11}$, respectively;

(ii) Test conservativeness:
under Model (1) and Assumption 1, $f(T_n = 1) \to 0$ if $\mathbb{H}_0$ is correct;

(iii) Test power:
under Model (1) and Assumption 1, $f(T_n = 1) \to 1$ if $|\beta_{11}| > 2 \max_{k \neq 1, p \neq 1} (\Lambda_{kp})$.

As a rule of thumb, one can apply the bootstrap method in practice.

Analogously for other causal effects with $\widehat{\mathrm{SPC}}_{kp}$.

Specificity score estimation, specificity test, and specificity map

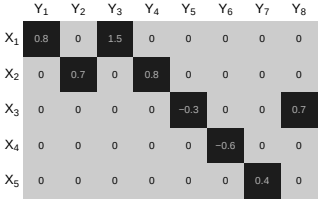
A visualization approach in a universal and efficient manner using a heatmap to show magnitude of the specificity scores.

Example 1. $n = 5000$ samples of five treatments and eight outcomes generated from

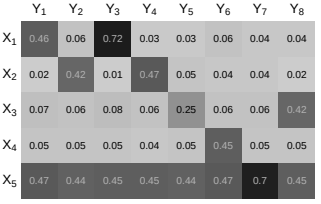
$$U \sim N(0,1), \quad X \sim N(\delta^{\mathrm{T}} U, I_5), \quad Y \sim N(\beta^{\mathrm{T}} X + \alpha^{\mathrm{T}} U, I_8),$$
$$\alpha = (1, \ldots, 1)^{\mathrm{T}}, \quad \delta = (0.4, \ldots, 0.4, 2),$$

The value of β is specified as in (a). $P^* = 4, K^* = 1$ and $\tau = 0.714$.

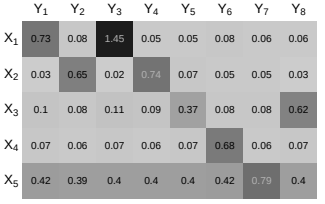
Specificity score estimation, specificity test, and specificity map



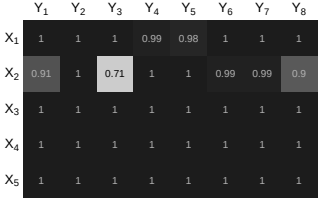
(a) Value of β



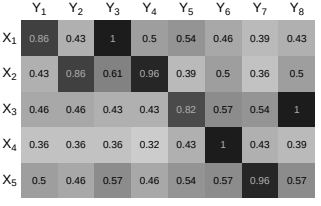
(b) Partial correlation



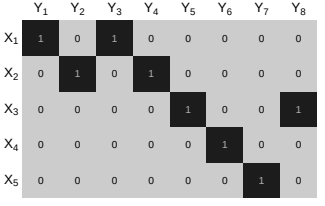
(c) Regression coefficients



(d) $1-$P value



(e) Specificity score



(f) Specificity test

Figure 1

Sensitivity analysis for the specificity test

Hume's Skepticism.

For $0 \leq \eta < 1$, $\mathcal{X}_p(\eta) = \{X_k : |\beta_{kp}/\Gamma_{kp}| > \eta\}$ and $\mathcal{Y}_k(\eta) = \{Y_p : |\beta_{kp}/\Gamma_{kp}| > \eta\}$.

Assumption 1S (η-violation to causal specificity).

(i) specificity of causes: $|\mathcal{X}_1(\eta)| \leq K^*$ for a given $K^* < K - 1$;

(ii) specificity of effects: for each $k \neq 1$, $|\mathcal{Y}_1(\eta) \cup \mathcal{Y}_k(\eta)| \leq P^*$ for a given $P^* < P - 1$.

No perfect negative controls. But causal effects are not all equally large.

The sensitivity parameter η characterizes the extent to which Assumption 1 is violated.

Sensitivity analysis for the specificity test

$q_1(\eta)$ the proportion of Λ_{kp}'s that satisfies (6), $q_2(\eta)$ be the proportion of Λ_{kp}'s that satisfies (7)

$$|\Gamma_{11}| > \Lambda_{kp}\frac{(1+\eta)^2}{1-\eta}. \quad (6)$$

$$|\Gamma_{11}| < \Lambda_{kp}\frac{(1-\eta)^2}{1+\eta}, \quad (7)$$

$$\mathrm{SPC}_{11}(\eta) = \max\{q_1(\eta), q_2(\eta)\} \quad (8)$$

Proposition 1. Under Model (1) and Assumption 1S, if $\mathbb{H}_0$ is correct then (7) and (6) hold together for at least $(K-1-K^*)(P-1-P^*)$ pairs of (k,p) and $\mathrm{SPC}_{11}(\eta) < \tau$.

One can test whether $\mathrm{SPC}_{11}(\eta) < \tau$ by varying the value of η to assess the correctness of $\mathbb{H}_0$ and the robustness of the specificity test.

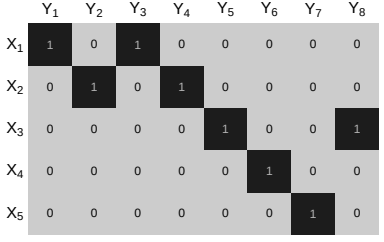
Nonparametric model, identification and estimation

- The following model is referred to as the nonparametric outcome model:

$$Y_p = m_p(X, U, \varepsilon_p), \quad \varepsilon_p \perp\!\!\!\perp (X, U), \quad p = 1 \ldots, P. \tag{9}$$

  Confounding bridge proposed by Miao et al. (2018) is used to characterize associations and confounding effects.

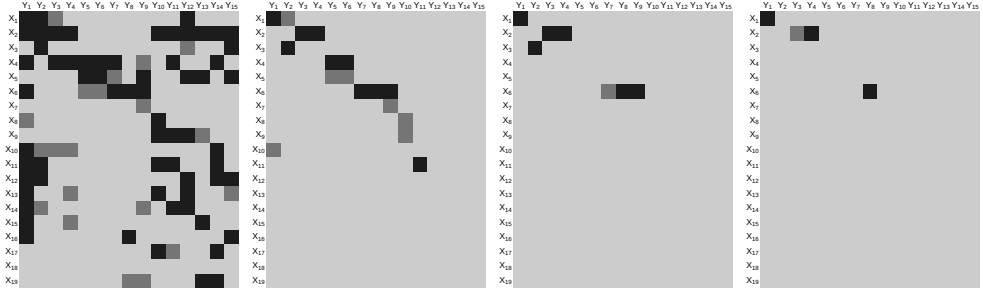
- Stepwise identification: Start with the negative controls identified with specificity test; iteratively identify causal effects and update negative control sets.



- Simultaneous identification: achievable under a stronger causal specificity assumption.

The identification strategies work for both linear and nonparametric models.

Application



(a) P value based test (b) BSPC, $\eta = 0$ (c) BSPC, $\eta = 0.1$ (d) BSPC, $\eta = 0.2$

Figure 2: Significance of the estimates.

Application

Some of our findings are consistent with previous biological studies.

- Overexpression of SOCS2 can inhibit the expression of Leptin receptor (Zhang et al., 2020);
- Fat absorption stimulates intestinal Apoa4 synthesis and secretion, serves as a satiety factor in response to ingestion of dietary fat (Tso et al., 2004);
- Igfbp2 protects against the development of obesity and insulin resistance (Wheatcroft et al., 2007);
- Body adiposity index captures the difference in body composition between males and females (Bergman et al., 2011);
- Irx3 is associated with lifestyle changes (Schneeberger, 2019)

Specificity analyses show that these above genes–traits associations cannot be completely attributed to confounding.

Although, false discoveries may present due to violation of assumptions, misspecification of models, reversion of causes and effects, etc.

Discussion

- Summary: Characterize the causal specificity assumption;
  Establish the specificity analysis including specificity score, specificity test, specificity map, specificity estimation;
  Has several favorable properties;
  Gauge effect size, bias and chance in a unified way;
- Extensions: handling other bias;
  high-dimensional setting;
  determination of the number of confounders;
  elaborate theory of the specificity score estimation and test;
  Causal diagram learning with unmeasured confounding;
- Limitations: single treatment setting;
  technical difficulties for the nonparametric model;
  potential publication bias;
  untestable assumptions.

Ambition for the causal big model

Hume (1988) Skepticism (18th century): there is no proof of causation.

The discovery of causes required an intellectual process going "beyond the impression of our senses." $\Leftarrow$ untestable assumptions

I (humbly) advocate the notion of the causal big model: Causal specificity is almost surely achieved in a big and elaborate causal system.

谢谢!

References

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association 91*, 444–455.

Bergman, R. N., D. Stefanovski, T. A. Buchanan, A. E. Sumner, J. C. Reynolds, N. G. Sebring, A. H. Xiang, and R. M. Watanabe (2011). A better index of body adiposity. *Obesity 19*(5), 1083–1089.

Brodeur, A., N. Cook, and A. Heyes (2020). Methods matter: P-hacking and publication bias in causal analysis in economics. *American Economic Review 110*, 3634–3660.

Chawla, D. S. (2017). Big names in statistics want to shake up much-maligned p value. *Nature 548*(7665).

Cui, Y., H. Pu, X. Shi, W. Miao, and E. Tchetgen Tchetgen (2023). Semiparametric proximal causal inference. *Journal of the American Statistical Association*, in press.

Doll, R. and A. B. Hill (1950). Smoking and carcinoma of the lung. *British medical journal 2*, 739.

Freedman, D. (1999). From association to causation: some remarks on the history of statistics. *Statistical Science 14*, 243–258.

References

Gelman, A. and E. Loken (2016). The statistical crisis in science. *The best writing on mathematics (Pitici M, ed)*, 305–318.

Gill, J. (2018). Comments from the new editor. *Political Analysis 26*, 1–2.

Goldberger, A. S. (1972). Structural equation methods in the social sciences. *Econometrica 40*, 979–1001.

Hill, A. B. (1965). The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine 58*, 295–300.

Höfler, M. (2005). The bradford hill considerations on causality: a counterfactual perspective. *Emerging themes in epidemiology 2*, 11.

Hume, D. (1988). An abstract of a treatise of human nature, 1740. In *Hume D. An Enquiry Concerning Enquiry Concerning Human Understanding*, pp. 27–43. La Salle, IL: Open Court.

Imbens, G. W. (2021). Statistical significance, p-values, and the reporting of uncertainty. *Journal of Economic Perspectives 35*, 157–74.

Ioannidis, J. (2016). Exposure-wide epidemiology: Revisiting bradford hill. *Statistics in medicine 35*, 1749–1762.

References

Lash, T. L. (2017). The harm done to reproducibility by the culture of null hypothesis significance testing. *American journal of epidemiology 186*, 627–635.

Li, K. Q., X. Shi, W. Miao, and E. Tchetgen Tchetgen (2022). Double negative control inference in test-negative design studies of vaccine effectiveness. *Journal of the American Statistical Association*, in press.

Luo, S., W. Li, W. Miao, and Y. He (2022). Identification and estimation of causal effects in the presence of confounded principal strata.

Miao, W., Z. Geng, and E. Tchetgen Tchetgen (2018). Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika 105*, 987–993.

Miao, W., X. Shi, and E. Tchetgen Tchetgen (2023). A confounding bridge approach for double negative control inference on causal effects.

Miao, W. and E. Tchetgen Tchetgen (2017). Invited commentary: Bias attenuation and identification of causal effects with multiple negative controls. *American Journal of Epidemiology 185*, 950–953.

Pearl, J. and D. Mackenzie (2018). *The Book of Why: The New Science of Cause and Effect*. Basic books.

References

Rothman, K. J. and S. Greenland (2005). Causation and causal inference in epidemiology. *American Journal of Public Health 95*, S144–S150.

Schneeberger, M. (2019). Irx3, a new leader on obesity genetics. *EBioMedicine 39*, 19–20.

Shi, X., W. Miao, M. Hu, and E. Tchetgen Tchetgen (2021). Theory for identification and inference with synthetic controls: A proximal causal inference framework. *Revision invited by Journal of the American Statistical Association*.

Shi, X., W. Miao, J. C. Nelson, and E. Tchetgen Tchetgen (2020). Multiply robust causal inference with double negative control adjustment for categorical unmeasured confounding. *Journal of the Royal Statistical Society: Series B 82*, 521–540.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society: Series B 13*, 238–241.

Tso, P., W. Sun, and M. Liu (2004). Gastrointestinal satiety signals iv. apolipoprotein a-iv. *American Journal of Physiology-Gastrointestinal and Liver Physiology 286*, G885–G890.

Wang, S., N. Yehya, E. E. Schadt, H. Wang, T. A. Drake, and A. J. Lusis (2006). Genetic and genomic analysis of a fat mass trait with complex inheritance reveals marked sex specificity. *PLoS Genetics 2*, e15.

References

Wasserstein, R. L. and N. A. Lazar (2016). The asa statement on p-values: context, process, and purpose.

Weiss, N. S. (2002). Can the specificity of an association be rehabilitated as a basis for supporting a causal hypothesis? *Epidemiology 13*, 6–8.

Wheatcroft, S. B., M. T. Kearney, A. M. Shah, V. A. Ezzat, J. R. Miell, M. Modo, S. C. Williams, W. P. Cawthorn, G. Medina-Gomez, A. Vidal-Puig, J. K. Sethi, and P. A. Crossey (2007). Igf-binding protein-2 protects against the development of obesity and insulin resistance. *Diabetes 56*, 285–294.

Wright, P. G. (1928). *Tariff on Animal and Vegetable Oils*. New York: Macmillan.

Ying, A., W. Miao, X. Shi, and E. Tchetgen Tchetgen (2022). Proximal causal inference for complex longitudinal studies. *Journal of the Royal Statistical Society: Series B*, in press.

Zhang, T., Y. Chen, J. Cai, M. Pan, Q. Sun, J. Zhang, and C. Sun (2020). Socs2 inhibits mitochondrial fatty acid oxidation via suppressing lepr/jak2/ampk signaling pathway in mouse adipocytes. *Oxidative Medicine and Cellular Longevity 2020*.