Estimating the Causal Effects of Adolescent and Adult Smoking Behavior on DNA Methylation Signatures

Lauren Schmitz[1]; Robel Alemu[2], Patrick Turley[3], David Cesarini[4], and Dan Benjamin[5]

[1]University of Wisconsin - Madison; [2]Tufts University; [3]Broad Institute of Harvard and MIT; [4]New York University; [5]University of California Los Angles

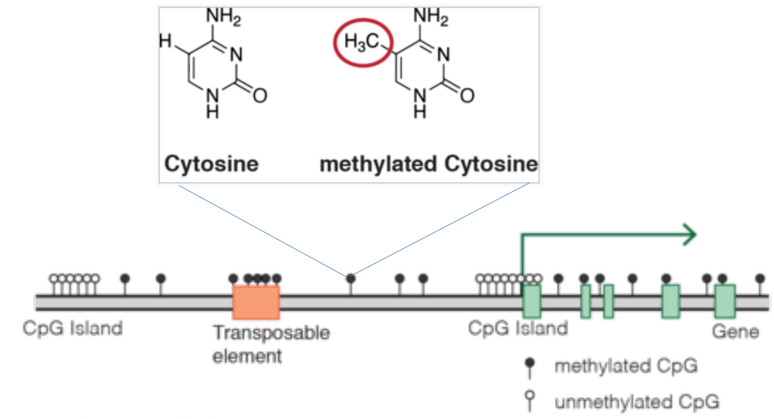
Background

- Smoking is one of the leading causes of mortality & morbidity – 7M deaths/ year globally (WHO, 2017)
- Confers long-term disease risk e.g. cancer, respiratory and CVDs
- Mechanisms behind long-term effects not well understood
- DNA methylation (DNAm) at suggested as one possible explanation
- DNAm is the addition of methyl groups to the DNA molecule at cytosine bases; usually represses gene transcription when located in a gene promoter (does not alter DNA sequence)
- DNAm is essential for normal development - associated with several processes e.g. carcinogenesis, aging, genome imprinting, etc.
- Epigenome Wide Association Studies (EWAS) have shown reproducible association between smoking & altered DNAm at thousands of CpGs (Joehanes et al., 2016; Sikdar et al., 2019)
- However, it is not clear whether differential DNAm are:
  - Caused by smoking
  - Downstream consequences of related co-morbidities
  - Caused by environmental factors highly correlated with smoking
- Research is needed to refine & validate smoking DNAm signatures
- Smoking DNAm signatures may serve as biomarkers of lifetime smoking exposures, shed light on molecular mechanisms & used as predictors of smoking-related diseases
- This study aims to estimate the causal effect of smoking in adolescence and adulthood on DNAm by exploiting state-year variation in cigarette taxes and data.

Data

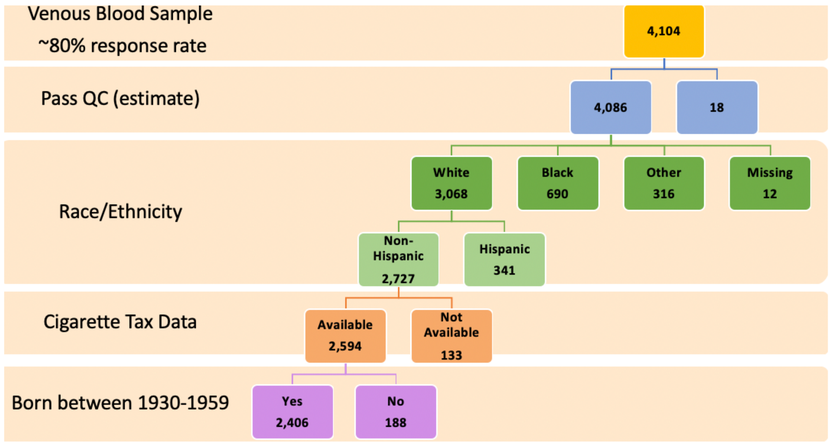
- **Health and Retirement Study (HRS):**
  - All available waves (1992-2016) ~4,000 DNAm samples (pop. representative); ~2,406 European ancestry DNAm samples
  - Smoking initiation (SI) as extensive margin and cigarettes per day (CPD) as an intensive margin measure of smoking
  - For SI, CpGs as reported by Sikdar et al., 2019
  - For CPD, CpGs reported by Joehannes et al., 2016
  - We'll exclude CpGs unique to newborns from maternal smoking & those overlapping between newborns & adults
- **Cigarette Tax Data:**
  - Annual excise taxes per cigarette pack from The Tax Burden on Tobacco (Orzechowski & Walker, 2016) adjusted for inflation
  - Cigarette tax data span all available years the HRS cohorts were in adolescence or adulthood (1940-2016)
  - Average of state and federal cigarette taxes respondents were exposed to:
    - Ages10-18 (1940-1977) to capture adolescent exposure
    - Average of HRS survey years (1992-2016) for adult exposure
- **Tobacco Control Policies Data (controls):**
  - State-level tobacco control policies extracted from *ImactTeens Project* & the Non-Smokers' Rights Foundation (ANRF) databases
  - Key variables to be used are youth access laws & smoke-free air laws

Typical Mammalian DNAm Landscape

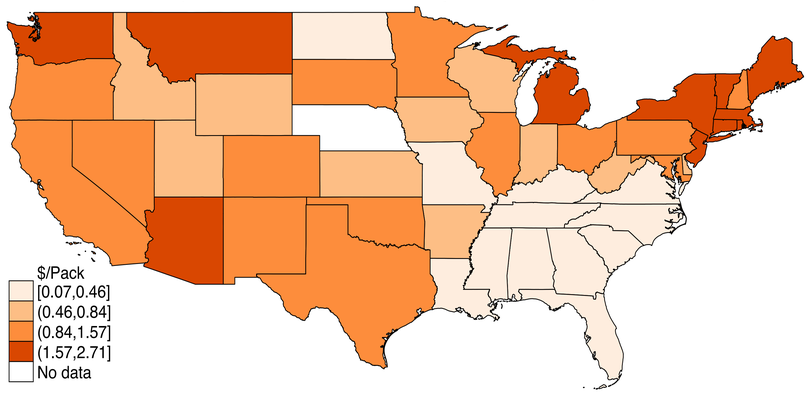


Source: https://en.wikipedia.org/wiki/DNA_methylation

HRS DNAm Sample (approx., data will be released soon)



Spatial Variation in the Inflation Adj. State Excise Tax, 2007



Empirical Strategy

- Exploit state-year variation in cigarette taxes as an instrumental variable
- Use 2-Stage Least Square (2SLS) model:

First stage:

$$\boldsymbol{SmokingPhen_{isbt}} = \boldsymbol{\pi_0} + \boldsymbol{\pi_1 CigTax_{st}} + \boldsymbol{X'_{isbt}}\,\boldsymbol{\pi_2} + \boldsymbol{S_s} + \boldsymbol{B_b} + \mathrm{P}_{st} + \boldsymbol{\varepsilon_{isbt}} \quad (1)$$

- $SmokingPhen_{isbt}$ is smoking phenotypes of interest for individual *i* in state s born in year *b* & interviewed in time *t*
- $CigTax_{st}$ is the average state cigarette tax during adulthood or adolescence
- X_{isbt} is a vector of individual characteristics including sex, race & education
- S_s *and* B_b are fixed effects that control for time-invariant state & cohort characteristics
- P_{st} is a vector of state-level tobacco control policies

Second stage:

$$\boldsymbol{CpG_{isbt}} = \boldsymbol{\delta_0} + \boldsymbol{\delta_1 SmokingPhen_{it}} + \boldsymbol{X'_{isbt}\delta_2} + \boldsymbol{S_s} + \boldsymbol{B_b} + P_{st} + \boldsymbol{\varepsilon_{isbt}} \quad (2)$$

- CpG_i is a single *CpG* site or average (aggregate) methylation beta value for *CpG* sites linked with a single known gene

- Additionally, we'll examine whether mQTLs or PGSs for smoking moderate the relationship between cigarette taxes and DNAm
- Moreover, we'll validate our results by testing for reverse causality using a Mendelian Randomization approach – use mQTLs as IV for DNAm & test whether DNAm influences smoking behavior

- **Power Calculation:**
  - Approximate effect size of *CpG* site *p* from an "IV EWAS" as:

$$\boldsymbol{R^2_{p\ IV\ EWAS}} = \boldsymbol{R^2_{p\ 1st\ STAGE}} \times \boldsymbol{R^2_{p\ EWAS}} \quad (3)$$

  - $R^2_{p\ EWAS}$ taken from most recent smoking EWAS (Sikdar, Joehanes, Joubert, Xu, Vives-Usano, et al., 2019)
  - $R^2_{p\ 1st\ STAGE}$ estimated for smoking outcomes using incremental R^2 of the $\pi_1 CigTax_{st}$ term in equation (1)

$$\boldsymbol{\Phi}\left(\boldsymbol{\Phi}^{-1}\left(\frac{\boldsymbol{\alpha}}{\boldsymbol{k}}\right) - \sqrt{\boldsymbol{N R^2_{p\ IV\ EWAS}}}\right) + \mathbf{1} - \boldsymbol{\Phi}\left(\boldsymbol{\Phi}^{-1}\left(\mathbf{1} - \frac{\boldsymbol{\alpha}}{\boldsymbol{k}}\right) - \sqrt{\boldsymbol{N R^2_{p\ IV\ EWAS}}}\right) \quad (4)$$

  - Φ is *CDF* of the standard normal distribution and Φ^{-1} denotes its inverse function
  - 80% power to detect effects in the European ancestry sample (N=2,406) when $R^2_{p\ IV\ EWAS} > 0.003$
  - Powered to detect effects at 386 differentially methylated *CpGs* b/n never & current smokers (Sikdar et al., 2019).
  - For CPD, well powered to detect effects at 70 differentially methylated *CpGs* (Joehanes et al., 2016)

Acknowledgements

This work was supported by a grant to Schmitz from the National Institute on Aging (K99/R00 AG056599). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. World Health Organization. WHO Report on the Global Tobacco Epidemic, 2017external icon. Geneva: World Health Organization, 2017 [accessed 2019 Jan 31].
2. Joehanes, R., Just, A. C., … London, S. J. (2016). Epigenetic Signatures of Cigarette Smoking. Circulation: Cardiovascular Genetics, 9(5), 436–447. https://doi.org/10.1161/CIRCGENETICS.116.001506.
3. Orzechowski, W., & Walker, R. (2016). The Tax Burden on Tobacco Volume 51, 1970-2016 (Vol. 51). Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.
4. Sikdar, S., Joehanes, R., …Aslibekyan, S. (2019). Comparison of smoking-related DNA methylation between newborns from prenatal exposure and adults from personal smoking. Epigenomics, 11(13), 1487–1500. https://doi.org/10.2217/epi-2019-0066