# Assessing Racial Disparities in Healthcare Expenditures Using Causal Path-Specific Effects

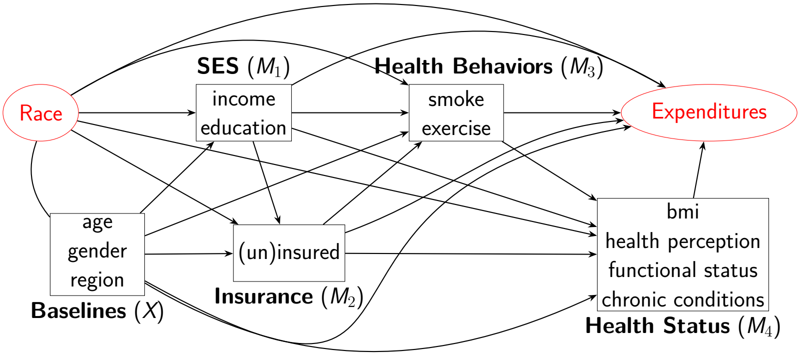
Xiaxian Ou

In collaborations with: Xinwei He, David Benkeser, Razieh Nabi

Department of Biostatistics and Bioinformatics
Emory University

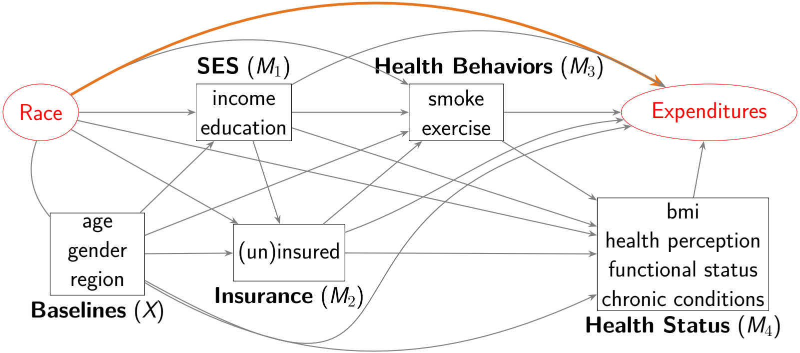
May 15th 2025

- **Racial disparities in healthcare expenditures** have been widely documented, which reflect **inequitable utilization** of healthcare services.
- A multitude of interrelated factors complicates analysis: **socioeconomic status (SES), access to insurance, health behaviors, health status**.



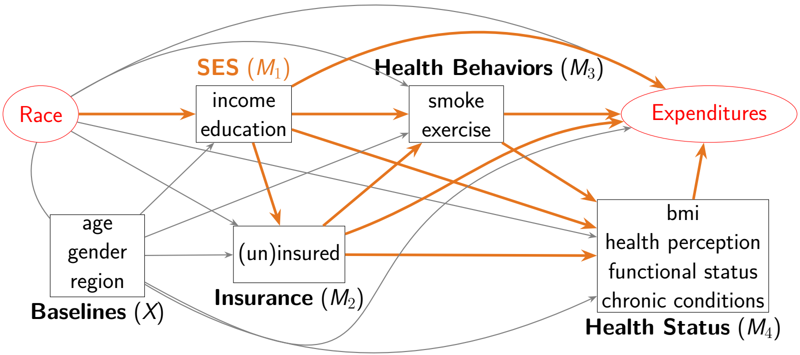
# ★ Effects:

**Race $\rightarrow$ Expenditures**; Race $\rightarrow$ SES $\rightsquigarrow$ Expenditures ; Race $\rightarrow$ Insurance $\rightsquigarrow$ Expenditures; Race $\rightarrow$ Health Behaviors $\rightsquigarrow$ Expenditures ; Race $\rightarrow$ Health Status $\rightarrow$ Expenditures



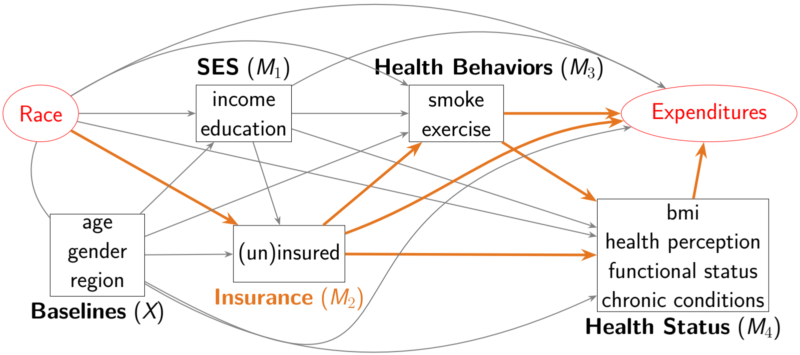
# ★ Effects:

Race $\rightarrow$ Expenditures; **Race $\rightarrow$ SES $\rightsquigarrow$ Expenditures**; Race $\rightarrow$ Insurance $\rightsquigarrow$ Expenditures; Race $\rightarrow$ Health Behaviors $\rightsquigarrow$ Expenditures ; Race $\rightarrow$ Health Status $\rightarrow$ Expenditures



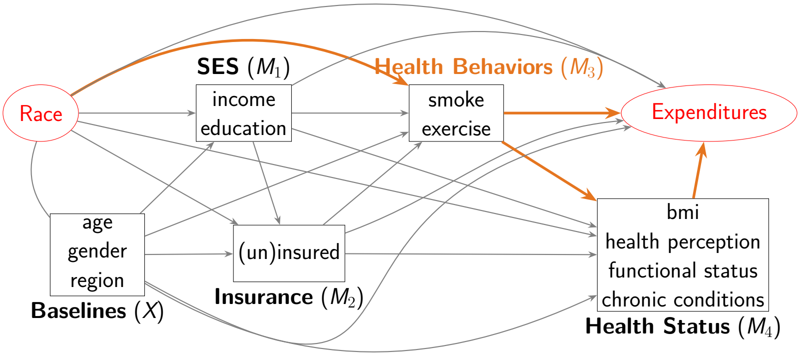
# ★ Effects:

Race $\rightarrow$ Expenditures; Race $\rightarrow$ SES $\rightsquigarrow$ Expenditures; **Race $\rightarrow$ Insurance $\rightsquigarrow$ Expenditures**; Race $\rightarrow$ Health Behaviors $\rightsquigarrow$ Expenditures ; Race $\rightarrow$ Health Status $\rightarrow$ Expenditures



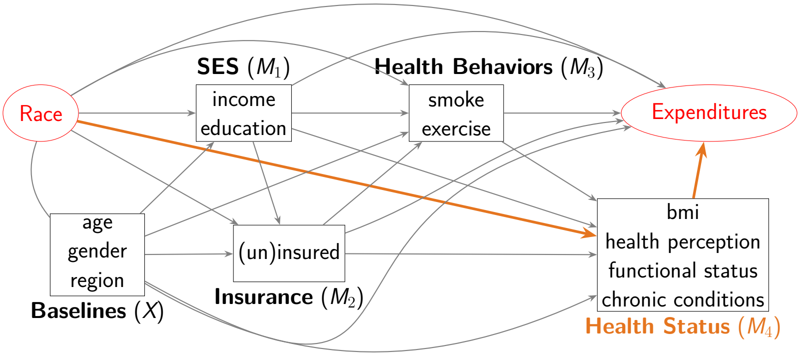
# ★ Effects:

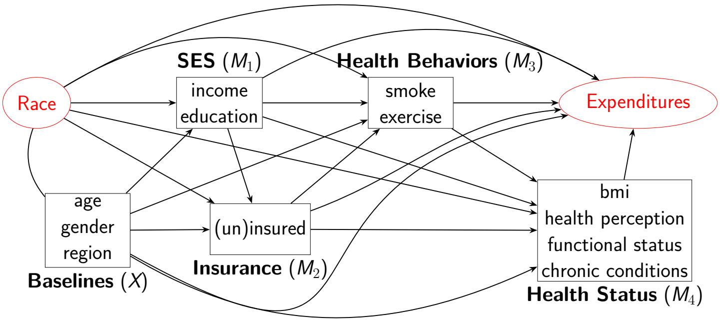
Race → Expenditures; Race → SES ⇝ Expenditures; Race → Insurance ⇝ Expenditures;
**Race → Health Behaviors ⇝ Expenditures** ; Race → Health Status → Expenditures



# ★ Effects:

Race $\rightarrow$ Expenditures; Race $\rightarrow$ SES $\rightsquigarrow$ Expenditures; Race $\rightarrow$ Insurance $\rightsquigarrow$ Expenditures; Race $\rightarrow$ Health Behaviors $\rightsquigarrow$ Expenditures ; **Race $\rightarrow$ Health Status $\rightarrow$ Expenditures**






**A nested potential outcome**: $r_0 \in \{0,1\}$ and $\mathbf{r} := (r_1, r_2, r_3, r_4) \in \{0,1\}^4$.

$$Y(r_0, \mathbf{r}) := Y\Big(r_0, \underbrace{M_1(r_1)}_{:=M_1^c}, \underbrace{M_2(r_2, M_1^c)}_{:=M_2^c}, \underbrace{M_3(r_3, M_1^c, M_2^c)}_{:=M_3^c}, M_4(r_4, M_1^c, M_2^c, M_3^c)\Big) \ ,$$

**Natural path-specific effects**:

$$\gamma_{R\to Y} := \mathbb{E}[Y(1,\mathbf{0})] \ , \quad \gamma_{R\to M_k \rightsquigarrow Y} := \mathbb{E}[Y(0,\mathbf{1}_k)] \ , \quad \gamma_{\text{ref}} = \mathbb{E}[Y(0,\mathbf{0})] \ .$$

$$\rho_{R\to Y} := \gamma_{R\to Y} - \gamma_{\text{ref}} \ , \qquad \rho_{R\to M_k \rightsquigarrow Y} := \gamma_{R\to M_k \rightsquigarrow Y} - \gamma_{\text{ref}} \ .$$

**Assumptions**: (a) Consistency; (b) Positivity; (c) Conditional ignorability:

for any $\overline{m}_k, r, r_0, r_k$,

- $Y(r_0, \overline{m}_4), \underline{M}_4(r_4, \overline{m}_3) \perp R \mid X$
- $Y(r_0, \overline{m}_4), \underline{M}_{k+1}(r_{k+1}, \overline{m}_k) \perp M_k(r, \overline{m}_{k-1}) \mid \overline{M}_{k-1}, R, X$

**Identification formulae**:

$$\rho_{R\to M_k \rightsquigarrow Y} = \int y\Big\{ dP(y \mid \overline{m}_4, R=0, x) \prod_{k=1}^{K} dP(m_k \mid \overline{m}_{k-1}, r_k, x) - dP(y \mid R=0, x)\Big\} dP(x),$$

$$\rho_{R\to Y} = \int y\Big\{ dP(y \mid \overline{m}_4, R=1, x) \prod_{k=1}^{K} dP(m_k \mid \overline{m}_{k-1}, R=0, x) - dP(y \mid R=0, x)\Big\} dP(x).$$

## One-step corrected plug-in estimators

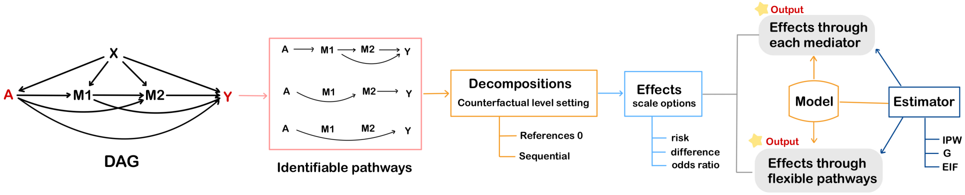
- $\gamma^{\text{plug-in}}(\hat{Q}) = \gamma(Q) - P\Phi(\hat{Q}) + R_2(\hat{Q}, Q)$ (first order bias)
- $\gamma^{+}(\hat{Q}) = \gamma^{\text{plug-in}}(\hat{Q}) + P_n\Phi(\hat{Q})$ ($\Phi$ denotes the influence function of the parameter)
  - ★ Robustness against model misspecification
  - ★ Flexible estimation
  - ★ Asymptotic normality

$$\gamma^{+}_{R\to Y}(\hat{Q}) = \frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{R_i}{1-\hat{\pi}(X_i)}\,\frac{1-\hat{g}_4(\overline{M}_{4,i},X_i)}{\hat{g}_4(\overline{M}_{4,i},X_i)}\,\big\{Y_i - \hat{\mu}_4(\overline{M}_{4,i},R=1,X_i)\big\}$$
$$+\frac{1-R_i}{1-\hat{\pi}(X_i)}\{\hat{\mu}_4(\overline{M}_{4,i},R=1,X_i) - \hat{\mathcal{C}}_{\mu_4}(R=0,X_i)\} + \hat{\mathcal{C}}_{\mu_4}(R=0,X_i)\Big\},$$

$$\gamma^{+}_{R\to M_k \rightsquigarrow Y}(\hat{Q})$$
$$= \frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{1-R_i}{1-\hat{\pi}(X_i)}\,\frac{\hat{g}_k(\overline{M}_{k,i},X_i)}{1-\hat{g}_k(\overline{M}_{k,i},X_i)}\,\frac{1-\hat{g}_{k-1}(\overline{M}_{k-1,i},X_i)}{\hat{g}_{k-1}(\overline{M}_{k-1,i},X_i)}\,\big\{Y_i - \hat{\mu}_k(\overline{M}_{k,i},R=0,X_i)\big\}$$
$$+\frac{R_i}{1-\hat{\pi}(X_i)}\,\frac{1-\hat{g}_{k-1}(\overline{M}_{k-1,i},X_i)}{\hat{g}_{k-1}(\overline{M}_{k-1,i},X_i)}\{\hat{\mu}_k(\overline{M}_{k,i},R=0,X_i) - \hat{\mathcal{B}}_k(\overline{M}_{k-1,i},R=1,X_i)\}$$
$$+\frac{1-R_i}{1-\hat{\pi}(X_i)}\{\hat{\mathcal{B}}_k(\overline{M}_{k-1,i},R=1,X_i) - \hat{\mathcal{C}}_{\mathcal{B}_k}(r_1,X_i)\} + \hat{\mathcal{C}}_{\mathcal{B}_k}(r_1,X_i)\Big\}.$$

# `flexPaths`: https://github.com/xxou/flexPaths

## An R Package for Causal Path-Specific Effect Estimation with Flexible Settings

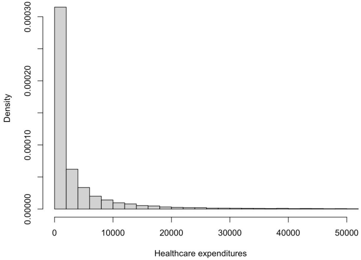


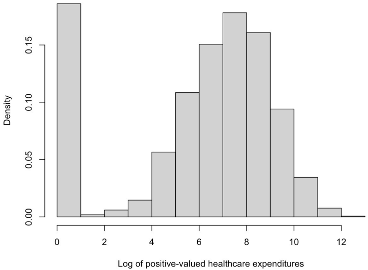

```
Info <- pathsInfo(data = singTreat, A="treat", Y="outcome1", cov_x=c("X1", "X2"),
                   M.list=list(M1="med1", M2=c('med2_1', 'med2_2')), estimation="EIF",
                   model.outcome=~SuperLearner(SL.library=c('SL.glm','SL.mean'),
                                               family=gaussian()),
                   model.treatment=~bart(verbose=FALSE, ndpost=1000))

re0 <- pathsEffect(Info, decomposition="refer0", scale="diff", CI_level=0.95)
##                        Path     Effect         SE   CI.lower  CI.upper P.value
## 1           A->M1->...->Y 0.15573821 0.03233753 0.09235782 0.2191186   0e+00
## 2           A->M2->...->Y 0.08354994 0.02165028 0.04111617 0.1259837   1e-04
## 3                    A->Y 0.49605603 0.06662290 0.36547755 0.6266345   0e+00
## 4 total effect: A->...->Y 0.73534418 0.06847876 0.60112828 0.8695601   0e+00
```

## Medical Expenditures Panel Survey (MEPS) in 2009 and 2016

- A large-scale survey that collects detailed data on healthcare costs, use, and insurance coverage from families, individuals, medical providers, and employers across the United States.
- Whites(9,830 vs. 8,772), Blacks(3,905 vs. 3,584), Asians(1,431 vs. 1,537) and Hispanics(5,150 vs. 5,636)

**! Zero-inflated right-skewed expenditure data**




✗ arithmetic mean ✓ scaled geometric mean

- $\mathbb{E}[\mathbb{I}(Y>0)\log(Y) \mid X] = \underbrace{P(Y>0 \mid X)}_{\text{part 1}} \times \underbrace{\mathbb{E}[\log(Y) \mid Y>0, X]}_{\text{part 2}}$
- $\exp\left(\mathbb{E}[Y(r_0, \mathbf{r}) - Y(0, \mathbf{0})]\right) \approx G_n\left(Y_{\text{pos}}(r_0, \mathbf{r})\right)^{\hat{P}(Y(r_0,\mathbf{r})>0)} / G_n\left(Y_{\text{pos}}(0, \mathbf{0})\right)^{\hat{P}(Y(0,\mathbf{0})>0)}$

**! SuperLearner**: `mean`, `glm`, `gam`, `glmnet`, `earth`, `ksvm`, `xgboost`, `ranger`, `dbarts`

## Summary:

- The effects via **SES, insurance and health status** were significant in five comparisons.
- The **direct effects** were significant in the comparisons between Whites and any minority.

Table: Path-Specific Effects among different race comparison (ratio scale) in **2009**

| Path | Effect(95%CI) | p value |
|---|---|---|
| Whites vs Blacks* | | |
| $R \to M_1 \rightsquigarrow Y$ | 1.114(1.054~1.173) | **0.000** |
| $R \to M_2 \rightsquigarrow Y$ | 1.017(0.984~1.050) | 0.321 |
| $R \to M_3 \rightsquigarrow Y$ | 0.981(0.959~1.003) | 0.089 |
| $R \to M_4 \to Y$ | 1.023(0.954~1.092) | 0.513 |
| $R \to Y$ | 1.772(1.616~1.929) | **0.000** |
| Total effect | 2.138(1.894~2.382) | **0.000** |
| Whites vs Asians* | | |
| $R \to M_1 \rightsquigarrow Y$ | 0.975(0.884~1.067) | 0.598 |
| $R \to M_2 \rightsquigarrow Y$ | 1.091(1.024~1.157) | **0.007** |
| $R \to M_3 \rightsquigarrow Y$ | 0.970(0.903~1.036) | 0.373 |
| $R \to M_4 \to Y$ | 1.418(1.242~1.594) | **0.000** |
| $R \to Y$ | 2.399(2.073~2.724) | **0.000** |
| Total effect | 2.863(2.377~3.350) | **0.000** |
| Whites vs Hispanics* | | |
| $R \to M_1 \rightsquigarrow Y$ | 1.450(1.344~1.557) | **0.000** |
| $R \to M_2 \rightsquigarrow Y$ | 1.372(1.306~1.439) | **0.000** |
| $R \to M_3 \rightsquigarrow Y$ | 1.076(1.014~1.137) | **0.016** |
| $R \to M_4 \to Y$ | 1.426(1.322~1.531) | **0.000** |
| $R \to Y$ | 2.097(1.916~2.279) | **0.000** |
| Total effect | 4.634(4.141~5.128) | **0.000** |

| Path | Effect(95%CI) | p value |
|---|---|---|
| Blacks vs Asians* | | |
| $R \to M_1 \rightsquigarrow Y$ | 0.835(0.721~0.949) | **0.004** |
| $R \to M_2 \rightsquigarrow Y$ | 1.079(1.009~1.149) | **0.027** |
| $R \to M_3 \rightsquigarrow Y$ | 0.974(0.931~1.017) | 0.233 |
| $R \to M_4 \to Y$ | 1.440(1.242~1.637) | **0.000** |
| $R \to Y$ | 1.044(0.876~1.212) | 0.610 |
| Total effect | 1.307(1.032~1.583) | **0.029** |
| Blacks vs Hispanics* | | |
| $R \to M_1 \rightsquigarrow Y$ | 1.192(1.130~1.254) | **0.000** |
| $R \to M_2 \rightsquigarrow Y$ | 1.478(1.393~1.562) | **0.000** |
| $R \to M_3 \rightsquigarrow Y$ | 1.023(0.986~1.060) | 0.225 |
| $R \to M_4 \to Y$ | 1.302(1.202~1.402) | **0.000** |
| $R \to Y$ | 1.024(0.943~1.104) | 0.568 |
| Total effect | 2.085(1.774~2.396) | **0.000** |
| Asians vs Hispanics* | | |
| $R \to M_1 \rightsquigarrow Y$ | 1.768(1.569~1.967) | **0.000** |
| $R \to M_2 \rightsquigarrow Y$ | 1.265(1.176~1.355) | **0.000** |
| $R \to M_3 \rightsquigarrow Y$ | 0.999(0.980~1.017) | 0.891 |
| $R \to M_4 \to Y$ | 0.788(0.719~0.857) | **0.000** |
| $R \to Y$ | 1.015(0.939~1.091) | 0.697 |
| Total effect | 1.855(1.521~2.189) | **0.000** |

arXiv > stat > arXiv:2504.21688

Search...

Help | Advar

Statistics > Applications

*[Submitted on 30 Apr 2025]*

# Assessing Racial Disparities in Healthcare Expenditures Using Causal Path-Specific Effects

Xiaxian Ou, Xinwei He, David Benkeser, Razieh Nabi

Racial disparities in healthcare expenditures are well-documented, yet the underlying drivers remain complex and require further investigation. This study employs causal and counterfactual path-specific effects to quantify how various factors, including socioeconomic status, insurance access, health behaviors, and health status, mediate these disparities. Using data from the Medical Expenditures Panel Survey, we estimate how expenditures would differ under counterfactual scenarios in which the values of specific mediators were aligned across racial groups along selected causal pathways. A key challenge in this analysis is ensuring robustness against model misspecification while addressing the zero-inflation and right-skewness of healthcare expenditures. For reliable inference, we derive asymptotically linear estimators by integrating influence function-based techniques with flexible machine learning methods, including super learners and a two-part model tailored to the zero-inflated, right-skewed nature of healthcare expenditures.

## arXiv

flexPaths Public

Unpin | Unwatch 1 | Fork 0 | Starred 3

main | 1 Branch | 0 Tags | Go to file t | Add file | Code

xxou message about pvalue and se for G and IPW 1543b43 · 2 weeks ago 26 Commits

| | | |
|---|---|---|
| R | message about pvalue and se for G and IPW | 2 weeks ago |
| data | 0824 call version | 9 months ago |
| examples | change model.propensity to model.treatment | 6 months ago |
| man | change model.propensity to model.treatment | 6 months ago |
| pic | onlySL in super learner | 2 months ago |
| vignettes | pathsInfo | 7 months ago |

**About**

Causal Path-Specific Analysis: flexible number of treatments and mediators, flexible pathways, flexible decomposition, flexible fitted models, flexible estimators

Readme

Activity

3 stars

1 watching

0 forks

## package

Xiaxian Ou, MSPH | Emory University | xiaxian.ou@emory.edu | https://xxou.github.io/