Examining Racial Disparities in Healthcare Expenditures via Causal Path-Specific Analysis

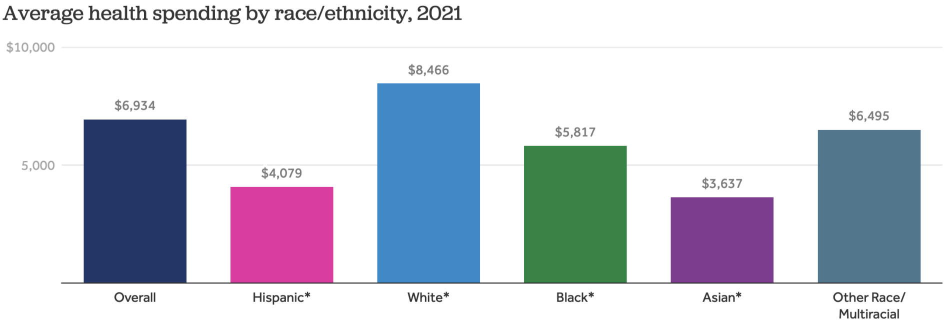
Xiaxian Ou

In collaborations with: Xinwei He , Razieh Nabi

Department of Biostatistics and Bioinformatics
Rollins School of Public Health
Emory University

Backgrounds

- Evidence from the **Medical Expenditures Panel Survey 2021 data** has highlighted the existence of racial disparities in the United States [1].
- Differences in healthcare expenditures reflect **inequitable utilization** of healthcare services



[1] Source: KFF analysis of 2021 Medical Expenditure Panel Survey

Backgrounds

Limitations of previous studies:

- Rely on **regression coefficients** in models with controlling confounders.
- Mediation analysis techniques, such as the Baron-Kenny approach, are often limited by **assumptions of linearity** (Baron and Kenny, 1986).
- Large **first-order bias** of the plug-in estimator.

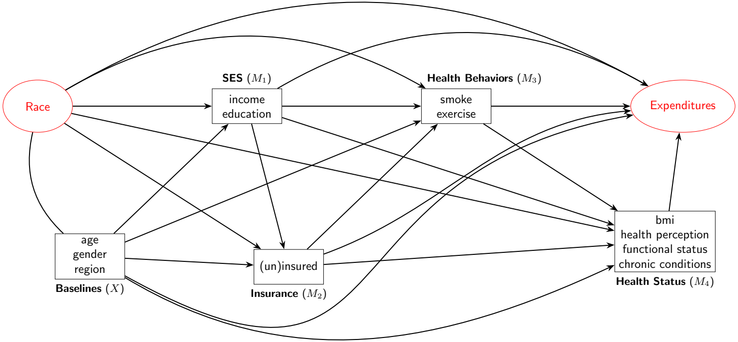
Motivations:

- A multitude of **interrelated factors** complicates analysis.
- A flexible, nonparametric framework based on **counterfactual formalization** in path-specific analysis (Pearl, 2009).
- Estimator derived from **efficient influence function** (EIF) and modeling technique involving **SuperLearner**.

Causal Path-Specific Effect Analysis

1. Estimand
2. Identification
3. Multiply Robust Estimators

Estimand

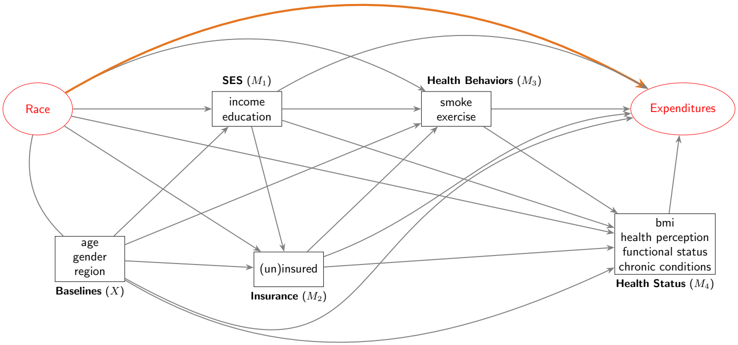


A nested potential outcome:

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

where $r_0 \in \{0, 1\}$ and $\mathbf{r} = (r_1, r_2, r_3, r_4) \in \{\mathbf{0} \cup \{\mathbf{1}_k : k = 1, 2, 3, 4\}\}$, where $\mathbf{1}_k$ denotes an indicator vector of size 4 with the k-th element being 1 (activating the pathways through M_k), and all other elements set to 0

Estimand

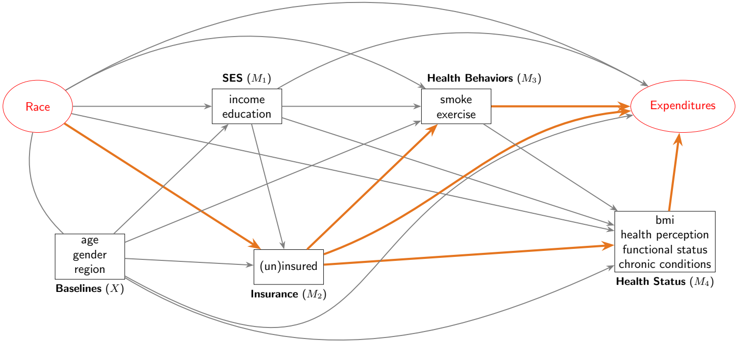


A nested potential outcome (example):

$$\phi(1, \mathbf{0}) := Y\Big(1, \underbrace{M_1(0)}_{:=M_1^c}, \underbrace{M_2(0, M_1^c)}_{:=M_2^c}, \underbrace{M_3(0, M_1^c, M_2^c)}_{:=M_3^c}, M_4(0, M_1^c, M_2^c, M_3^c)\Big) \ ,$$

The healthcare expenditures would be if individuals were White, while all mediating factors (SES, insurance, health behaviors, health status) remain at levels observed for Black individuals.

Estimand



A nested potential outcome (example):

$$\phi(0, \mathbf{1}_2) := Y\Big(0, \underbrace{M_1(0)}_{:=M_1^c}, \underbrace{M_2(1, M_1^c)}_{:=M_2^c}, \underbrace{M_3(0, M_1^c, M_2^c)}_{:=M_3^c}, M_4(0, M_1^c, M_2^c, M_3^c)\Big) \ ,$$

The healthcare expenditures of a hypothetical population would be where everyone is Black and insurance is set to the levels observed for White individuals.

Estimand

Average potential outcome

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

Direct effect and effect through each mediator:

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

- $\rho_{R\to Y}$: The average change in healthcare expenditures if individuals were White vs. Black, while all mediating factors (SES, insurance, health behaviors, health status) remain at levels observed for Black individuals (discrimination).
- $\rho_{R\to M_k \rightsquigarrow Y}$: It compares the expenditures of a hypothetical population where everyone is Black to one where M_k is set to the levels observed for White individuals.

Identification

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

$$
\begin{aligned}
Y(r_0,\overline{m}_4), M_4(r_4,\overline{m}_3), M_3(r_3,\overline{m}_2), M_2(r_2,m_1), M_1(r_1) &\perp R \mid X \ , \\
Y(r_0,\overline{m}_4), M_4(r_4,\overline{m}_3), M_3(r_3,\overline{m}_2), M_2(r_2,m_1) &\perp M_1(r) \mid R, X \ , \\
Y(r_0,\overline{m}_4), M_4(r_4,\overline{m}_3), M_3(r_3,\overline{m}_2) &\perp M_2(r,m_1) \mid M_1, R, X \ , \\
Y(r_0,\overline{m}_4), M_4(r_4,\overline{m}_3) &\perp M_3(r,\overline{m}_2) \mid \overline{M}_2, R, X \ , \\
Y(r_0,\overline{m}_4) &\perp M_4(r,\overline{m}_3) \mid \overline{M}_3, R, X \ .
\end{aligned}
$$

Identification

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

Identification formula:

$$\gamma_{\text{inact}} = \int y dP(y \mid R = 0, x) dP(x) \ ,$$

$$\gamma_{R \to Y} = \int y dP(y \mid \overline{m}_4, R = 1, x) \prod_{k=1}^{4} dP(m_k \mid \overline{m}_{k-1}, R = 0, x) dP(x) \ ,$$

$$\gamma_{R \to M_k \rightsquigarrow Y} = \int y dP(y \mid \overline{m}_4, R = 0, x) dP(m_k \mid \overline{m}_{k-1}, R = 1, x) \prod_{\substack{j=1 \\ j \neq k}}^{4} dP(m_j \mid \overline{m}_{j-1}, R = 0, x) dP(x) \ ,$$

★ Large **first-order bias** $P\Phi(\hat{Q})$ of the plug-in estimator: $\gamma^{\text{plug-in}}(\hat{Q}) = \gamma(Q) - P\Phi(\hat{Q}) + R_2(\hat{Q}, Q)$

Multiply robust estimators

$$
\begin{aligned}
\gamma^{+}_{R\to Y}(\hat{Q}) = \frac{1}{n}\sum_{i=1}^{n}\Bigg\{ & \frac{R_i}{1-\hat{\pi}(X_i)}\,\frac{1-\hat{g}_4(\overline{M}_{4,i},X_i)}{\hat{g}_4(\overline{M}_{4,i},X_i)}\,\{Y_i-\hat{\mu}_4(\overline{M}_{4,i},R=1,X_i)\} \\
& + \frac{1-R_i}{1-\hat{\pi}(X_i)}\left\{\hat{\mu}_4(\overline{M}_{4,i},R=1,X_i)-\hat{\mathcal{C}}_{\mu_4}(R=0,X_i)\right\} + \hat{\mathcal{C}}_{\mu_4}(R=0,X_i)\Bigg\}\ , \\
\gamma^{+}_{R\to M_k\rightsquigarrow Y}(\hat{Q}) = \frac{1}{n}\sum_{i=1}^{n}\Bigg\{ & \frac{1-R_i}{1-\hat{\pi}(X_i)}\,\frac{\hat{g}_k(\overline{M}_{k,i},X_i)}{1-\hat{g}_k(\overline{M}_{k,i},X_i)}\,\frac{1-\hat{g}_{k-1}(\overline{M}_{k-1,i},X_i)}{\hat{g}_{k-1}(\overline{M}_{k-1,i},X_i)}\,\{Y_i-\hat{\mu}_k(\overline{M}_{k,i},R=0,X_i)\} \\
& + \frac{R_i}{1-\hat{\pi}(X_i)}\,\frac{1-\hat{g}_{k-1}(\overline{M}_{k-1,i},X_i)}{\hat{g}_{k-1}(\overline{M}_{k-1,i},X_i)}\left\{\hat{\mu}_k(\overline{M}_{k,i},R=0,X_i)-\hat{\mathcal{B}}_k(\overline{M}_{k-1,i},R=1,X_i)\right\} \\
& + \frac{1-R_i}{1-\hat{\pi}(X_i)}\left\{\hat{\mathcal{B}}_k(\overline{M}_{k-1,i},R=1,X_i)-\hat{\mathcal{C}}_{\mathcal{B}_k}(r_1,X_i)\right\} + \hat{\mathcal{C}}_{\mathcal{B}_k}(r_1,X_i)\Bigg\}\ ,\ \ k=1,2,3,4\ .
\end{aligned}
$$

- $\pi(x) = P(R=1|x)$;
- $g_k(\overline{m}_k, x) = P(R=1|\overline{m}_k, x)$
- $\mu_k(\overline{m}_k, r_0, x) = \mathbb{E}(Y|\overline{m}_k, r_0, x)$, $\mathcal{B}_k(\overline{m}_{k-1}, r_k, x) = \mathbb{E}(\mu_k(\overline{m}_k, r_0, x)|\overline{m}_{k-1}, r_k, x)$,
- $\mathcal{C}_{\mathcal{B}_k}(r_1, x) = \mathbb{E}[\mathcal{B}_k(\overline{m}_{k-1}, r_k, x)|r_1, x]$ for $k=2,3,4$, and $\mathcal{C}_{\mu_4}(r_1, x) = \mathbb{E}[\mu_4(\overline{m}_4, r_y, x)|r_1, x]$;

Multiply robust estimators

Advantages

1. These estimators are robust against model misspecification.
2. These estimators are particularly well-suited for incorporating data-adaptive methods
3. Desirable statistical properties such as asymptotic normality and $\sqrt{n}$-consistency

Empirical Data Analysis

1. Zero inflation and right skewed data
2. SuperLearner

MEPS Data

- The **Medical Expenditures Panel Survey (MEPS)**, is 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.
- Here, we used the household components of the **2009 MEPS** data. The initial sample size was 20,889, which was trimmed to 20,316 after focusing on **non-Hispanic Whites** (9,830), **non-Hispanic Blacks** (3,905), **Asians** (1,431) and **Hispanics** (5,150).

Variables

1. Self-reported race;
2. Mediators:
   - SES (M_1): income, education
   - Insurance access (M_2): uninsured
   - Health behaviors (M_3): smoking, exercise
   - Health status (M_4): BMI, health perception, functional status (daily living activities, functional, or sensory abilities, social limitations) and chronic conditions (diabetes, asthma, high blood pressure, coronary heart disease, angina, myocardial infarction, stroke, emphysema, cholesterol, arthritis, and cancer).
3. Covariates: age, gender, region
4. Outcome: Annual total healthcare expenditures — the sum of out-of-pocket payments and payments by private insurance, Medicaid, Medicare, and other sources.

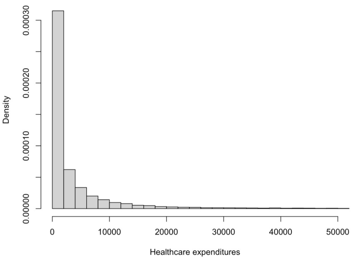
Challenge

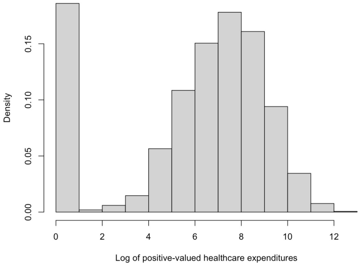
Zero-inflated right-skewed data

Two-part model

- Part 1: The probability of a non-zero response, $P(Y > 0 \mid X)$,
- Part 2: The probability distribution of the positive responses $\mathbb{E}[\log(Y) \mid Y > 0, X]$

The conditional mean: $\mathbb{E}[\mathbb{I}(Y > 0)\log(Y) \mid X] = P(Y > 0 \mid X) \times \mathbb{E}[\log(Y) \mid Y > 0, X]$.





Model estimation

SuperLearner

SuperLearner: an ensemble learning method that combines flexible statistical and machine learning models via cross-validation to reduce model misspecification and improve predictive accuracy (Van der Laan et al., 2007).

- Binomial family: `glm`, `glm.interaction`, `randomForest`, `xgboost`, `dbarts`
- Gaussian family: `glmnet`, `polymars`, `lm`, `dbarts`

Results and Discussion

Results

Interpretation :

- The **scaled geometric mean** of healthcare expenditures would be 1.787 times if individuals were White vs. Black, while all mediating factors remain at levels observed for Black individuals.
- If SES for Blacks were hypothetically aligned with that of Whites, the **scaled geometric mean** of their healthcare expenditures would increase to 1.098 (95% CI: 1.035 – 1.161) times.

Whites vs Blacks*		
Path	Effect(95%CI)	p value
$R \rightarrow M_1 \rightsquigarrow Y$	1.098(1.035˜1.161)	**0.001**
$R \rightarrow M_2 \rightsquigarrow Y$	1.009(0.974˜1.044)	0.606
$R \rightarrow M_3 \rightsquigarrow Y$	0.974(0.951˜0.998)	**0.035**
$R \rightarrow M_4 \rightarrow Y$	1.035(0.963˜1.107)	0.337
$R \rightarrow Y$	1.787(1.629˜1.945)	**0.000**
Total effect	2.106(1.865˜2.347)	**0.000**

Results

★ (1) **Total effect**: inequality of healthcare access and utilization.

Path	Effect(95%CI)	p value
Whites vs Blacks*		
$R \to M_1 \rightsquigarrow Y$	1.098(1.035~1.161)	**0.001**
$R \to M_2 \rightsquigarrow Y$	1.009(0.974~1.044)	0.606
$R \to M_3 \rightsquigarrow Y$	0.974(0.951~0.998)	**0.035**
$R \to M_4 \to Y$	1.035(0.963~1.107)	0.337
$R \to Y$	1.787(1.629~1.945)	**0.000**
Total effect	2.106(1.865~2.347)	**0.000**
Whites vs Asians*		
$R \to M_1 \rightsquigarrow Y$	0.945(0.834~1.055)	0.339
$R \to M_2 \rightsquigarrow Y$	1.054(0.992~1.116)	0.081
$R \to M_3 \rightsquigarrow Y$	0.982(0.927~1.036)	0.509
$R \to M_4 \to Y$	1.358(1.171~1.545)	**0.000**
$R \to Y$	2.521(2.159~2.883)	**0.000**
Total effect	2.805(2.304~3.306)	**0.000**
Whites vs Hispanics*		
$R \to M_1 \rightsquigarrow Y$	1.572(1.438~1.705)	**0.000**
$R \to M_2 \rightsquigarrow Y$	1.377(1.305~1.450)	**0.000**
$R \to M_3 \rightsquigarrow Y$	1.089(1.020~1.159)	**0.009**
$R \to M_4 \to Y$	1.436(1.307~1.564)	**0.000**
$R \to Y$	2.044(1.865~2.223)	**0.000**
Total effect	4.647(4.143~5.150)	**0.000**

Path	Effect(95%CI)	p value
Blacks vs Asians*		
$R \to M_1 \rightsquigarrow Y$	0.837(0.692~0.981)	**0.043**
$R \to M_2 \rightsquigarrow Y$	1.024(0.947~1.101)	0.531
$R \to M_3 \rightsquigarrow Y$	0.970(0.917~1.023)	0.271
$R \to M_4 \to Y$	1.475(1.243~1.708)	**0.000**
$R \to Y$	1.111(0.917~1.305)	0.237
Total effect	1.297(1.008~1.585)	**0.022**
Blacks vs Hispanics*		
$R \to M_1 \rightsquigarrow Y$	1.268(1.194~1.343)	**0.000**
$R \to M_2 \rightsquigarrow Y$	1.486(1.384~1.588)	**0.000**
$R \to M_3 \rightsquigarrow Y$	1.053(1.006~1.099)	**0.022**
$R \to M_4 \to Y$	1.367(1.183~1.552)	**0.000**
$R \to Y$	0.988(0.910~1.066)	0.770
Total effect	2.111(1.791~2.431)	**0.000**
Asians vs Hispanics*		
$R \to M_1 \rightsquigarrow Y$	1.960(1.706~2.213)	**0.000**
$R \to M_2 \rightsquigarrow Y$	1.342(1.221~1.463)	**0.000**
$R \to M_3 \rightsquigarrow Y$	0.998(0.978~1.018)	0.846
$R \to M_4 \to Y$	0.811(0.716~0.906)	**0.000**
$R \to Y$	0.988(0.912~1.064)	0.760
Total effect	1.884(1.541~2.227)	**0.000**

Results

★ (2) **SES** and **insurance** were important mediators.

Path	Effect(95%CI)	p value
Whites vs Blacks*		
$R \to M_1 \rightsquigarrow Y$	1.098(1.035~1.161)	**0.001**
$R \to M_2 \rightsquigarrow Y$	1.009(0.974~1.044)	0.606
$R \to M_3 \rightsquigarrow Y$	0.974(0.951~0.998)	**0.035**
$R \to M_4 \to Y$	1.035(0.963~1.107)	0.337
$R \to Y$	1.787(1.629~1.945)	**0.000**
Total effect	2.106(1.865~2.347)	**0.000**
Whites vs Asians*		
$R \to M_1 \rightsquigarrow Y$	0.945(0.834~1.055)	0.339
$R \to M_2 \rightsquigarrow Y$	1.054(0.992~1.116)	0.081
$R \to M_3 \rightsquigarrow Y$	0.982(0.927~1.036)	0.509
$R \to M_4 \to Y$	1.358(1.171~1.545)	**0.000**
$R \to Y$	2.521(2.159~2.883)	**0.000**
Total effect	2.805(2.304~3.306)	**0.000**
Whites vs Hispanics*		
$R \to M_1 \rightsquigarrow Y$	1.572(1.438~1.705)	**0.000**
$R \to M_2 \rightsquigarrow Y$	1.377(1.305~1.450)	**0.000**
$R \to M_3 \rightsquigarrow Y$	1.089(1.020~1.159)	**0.009**
$R \to M_4 \to Y$	1.436(1.307~1.564)	**0.000**
$R \to Y$	2.044(1.865~2.223)	**0.000**
Total effect	4.647(4.143~5.150)	**0.000**

Path	Effect(95%CI)	p value
Blacks vs Asians*		
$R \to M_1 \rightsquigarrow Y$	0.837(0.692~0.981)	**0.043**
$R \to M_2 \rightsquigarrow Y$	1.024(0.947~1.101)	0.531
$R \to M_3 \rightsquigarrow Y$	0.970(0.917~1.023)	0.271
$R \to M_4 \to Y$	1.475(1.243~1.708)	**0.000**
$R \to Y$	1.111(0.917~1.305)	0.237
Total effect	1.297(1.008~1.585)	**0.022**
Blacks vs Hispanics*		
$R \to M_1 \rightsquigarrow Y$	1.268(1.194~1.343)	**0.000**
$R \to M_2 \rightsquigarrow Y$	1.486(1.384~1.588)	**0.000**
$R \to M_3 \rightsquigarrow Y$	1.053(1.006~1.099)	**0.022**
$R \to M_4 \to Y$	1.367(1.183~1.552)	**0.000**
$R \to Y$	0.988(0.910~1.066)	0.770
Total effect	2.111(1.791~2.431)	**0.000**
Asians vs Hispanics*		
$R \to M_1 \rightsquigarrow Y$	1.960(1.706~2.213)	**0.000**
$R \to M_2 \rightsquigarrow Y$	1.342(1.221~1.463)	**0.000**
$R \to M_3 \rightsquigarrow Y$	0.998(0.978~1.018)	0.846
$R \to M_4 \to Y$	0.811(0.716~0.906)	**0.000**
$R \to Y$	0.988(0.912~1.064)	0.760
Total effect	1.884(1.541~2.227)	**0.000**

Results

★ (3) **Direct effect**: The direct effects were significant in the comparisons between Whites and any minority (unmeasered variables or discrimination).

Path	Effect(95%CI)	p value
Whites vs Blacks*		
$R \to M_1 \rightsquigarrow Y$	1.098(1.035~1.161)	**0.001**
$R \to M_2 \rightsquigarrow Y$	1.009(0.974~1.044)	0.606
$R \to M_3 \rightsquigarrow Y$	0.974(0.951~0.998)	**0.035**
$R \to M_4 \to Y$	1.035(0.963~1.107)	0.337
$R \to Y$	1.787(1.629~1.945)	**0.000**
Total effect	2.106(1.865~2.347)	**0.000**
Whites vs Asians*		
$R \to M_1 \rightsquigarrow Y$	0.945(0.834~1.055)	0.339
$R \to M_2 \rightsquigarrow Y$	1.054(0.992~1.116)	0.081
$R \to M_3 \rightsquigarrow Y$	0.982(0.927~1.036)	0.509
$R \to M_4 \to Y$	1.358(1.171~1.545)	**0.000**
$R \to Y$	2.521(2.159~2.883)	**0.000**
Total effect	2.805(2.304~3.306)	**0.000**
Whites vs Hispanics*		
$R \to M_1 \rightsquigarrow Y$	1.572(1.438~1.705)	**0.000**
$R \to M_2 \rightsquigarrow Y$	1.377(1.305~1.450)	**0.000**
$R \to M_3 \rightsquigarrow Y$	1.089(1.020~1.159)	**0.009**
$R \to M_4 \to Y$	1.436(1.307~1.564)	**0.000**
$R \to Y$	2.044(1.865~2.223)	**0.000**
Total effect	4.647(4.143~5.150)	**0.000**

Path	Effect(95%CI)	p value
Blacks vs Asians*		
$R \to M_1 \rightsquigarrow Y$	0.837(0.692~0.981)	**0.043**
$R \to M_2 \rightsquigarrow Y$	1.024(0.947~1.101)	0.531
$R \to M_3 \rightsquigarrow Y$	0.970(0.917~1.023)	0.271
$R \to M_4 \to Y$	1.475(1.243~1.708)	**0.000**
$R \to Y$	1.111(0.917~1.305)	0.237
Total effect	1.297(1.008~1.585)	**0.022**
Blacks vs Hispanics*		
$R \to M_1 \rightsquigarrow Y$	1.268(1.194~1.343)	**0.000**
$R \to M_2 \rightsquigarrow Y$	1.486(1.384~1.588)	**0.000**
$R \to M_3 \rightsquigarrow Y$	1.053(1.006~1.099)	**0.022**
$R \to M_4 \to Y$	1.367(1.183~1.552)	**0.000**
$R \to Y$	0.988(0.910~1.066)	0.770
Total effect	2.111(1.791~2.431)	**0.000**
Asians vs Hispanics*		
$R \to M_1 \rightsquigarrow Y$	1.960(1.706~2.213)	**0.000**
$R \to M_2 \rightsquigarrow Y$	1.342(1.221~1.463)	**0.000**
$R \to M_3 \rightsquigarrow Y$	0.998(0.978~1.018)	0.846
$R \to M_4 \to Y$	0.811(0.716~0.906)	**0.000**
$R \to Y$	0.988(0.912~1.064)	0.760
Total effect	1.884(1.541~2.227)	**0.000**

Discussion

1. **Targeted interventions**: To improve educational opportunities for minority populations, address systemic barriers to insurance enrollment, and equip healthcare providers with training to recognize and address implicit biases.
2. **Algorithmic fairness** for healthcare decisions: Integrating fairness-aware models could help ensure that algorithms do not perpetuate structural inequalities, instead promoting equitable healthcare access for all racial groups (Nabi and Shpitser, 2018).
3. **Race as a causal variable**: “no causation without manipulation.” we adopted the “weak interpretation” (VanderWeele and Robinson, 2014) .

Concluding Remark

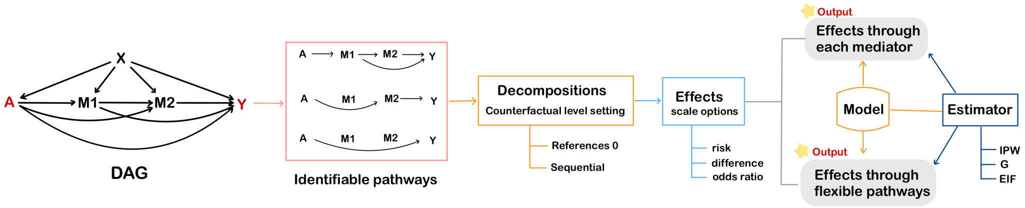
Summary: We achieved the application of causal path-specific effect analysis framework to examine racial disparities in healthcare expenditures, highlighting its potential for broader use in public health and social science research.

Future work:

- Consider methods to address selection bias and incorporate broader data sources.
- Explore additional pathways, such as those involving exposure-induced confounding.
- Conduct sensitivity analyses.

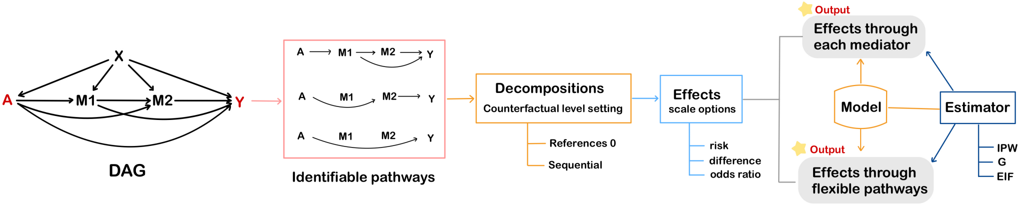
`flexPaths`: An R Package for Flexible and Robust Causal Path-Specific Effect Estimation

Workflow of single treatment




```
Info <- pathsInfo(data = singTreat, A="treat", Y="outcome1", cov_x=c("X1", "X2"),
                  M.list=list(M1="med1", M2=c('med2_1', 'med2_2')),
                  model.outcome=~bart(),  model.treatment=~bart())
```

Workflow of single treatment




```r
re0 <- pathsEffect(Info, decomposition="refer0", scale="diff"); re0
##                      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
```

Q&A

Xiaxian Ou, MSPH student (she/her/hers)
Department of Biostatistics and Bioinformatics
Emory University
xiaxian.ou@emory.edu
https://xxou.github.io/

References

Baron, R. M. and Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6):1173.

Nabi, R. and Shpitser, I. (2018). Fair inference on outcomes. In *Proceedings of the Thirty Second Conference on Association for the Advancement of Artificial Intelligence (AAAI-32nd)*. AAAI Press.

Pearl, J. (2009). *Causality*. Cambridge university press.

Van der Laan, M. J., Polley, E. C., and Hubbard, A. E. (2007). Super learner. *Statistical applications in genetics and molecular biology*, 6(1).

VanderWeele, T. J. and Robinson, W. R. (2014). On the causal interpretation of race in regressions adjusting for confounding and mediating variables. *Epidemiology*, 25(4):473–484.

Appendix - The ratio of geometric mean

Transformation

- Let $Y(r_0, \mathbf{r}) = \mathbb{I}(\phi(r_0, \mathbf{r}) > 0) \log \phi(r_0, \mathbf{r})$. The effects are defined as:

$$\exp\left(\mathbb{E}[Y(r_0, \mathbf{r}) - Y(0, \mathbf{0})]\right) \approx \frac{\left\{\prod_{i=1}^{n} \phi_{i,\text{pos}}(r_0, \mathbf{r})\right\}^{\hat{P}(\phi(r_0,\mathbf{r})>0)/n}}{\left\{\prod_{i=1}^{n} \phi_{i,\text{pos}}(0, \mathbf{0})\right\}^{\hat{P}(\phi(0,\mathbf{0})>0)/n}} = \frac{G_n\left(\phi_{\text{pos}}(r_0, \mathbf{r})\right)^{\hat{P}(\phi(r_0,\mathbf{r})>0)}}{G_n\left(\phi_{\text{pos}}(0, \mathbf{0})\right)^{\hat{P}(\phi(0,\mathbf{0})>0)}}$$

- Delta method (eg. direct effect):

$$\begin{aligned} &\sqrt{n}\Big(\exp(\rho^{+}_{R\to Y}(\hat{Q})) - \exp(\rho_{R\to Y}(Q))\Big) \\ &\qquad \to^{d} \mathcal{N}\left(0, \exp(\rho_{R\to Y}(Q))^2 \times \mathbb{E}\big[\big(\Phi_{R\to Y}(Q) - \Phi_{\text{inact}}(Q)\big)^2\big]\right) \ . \end{aligned}$$