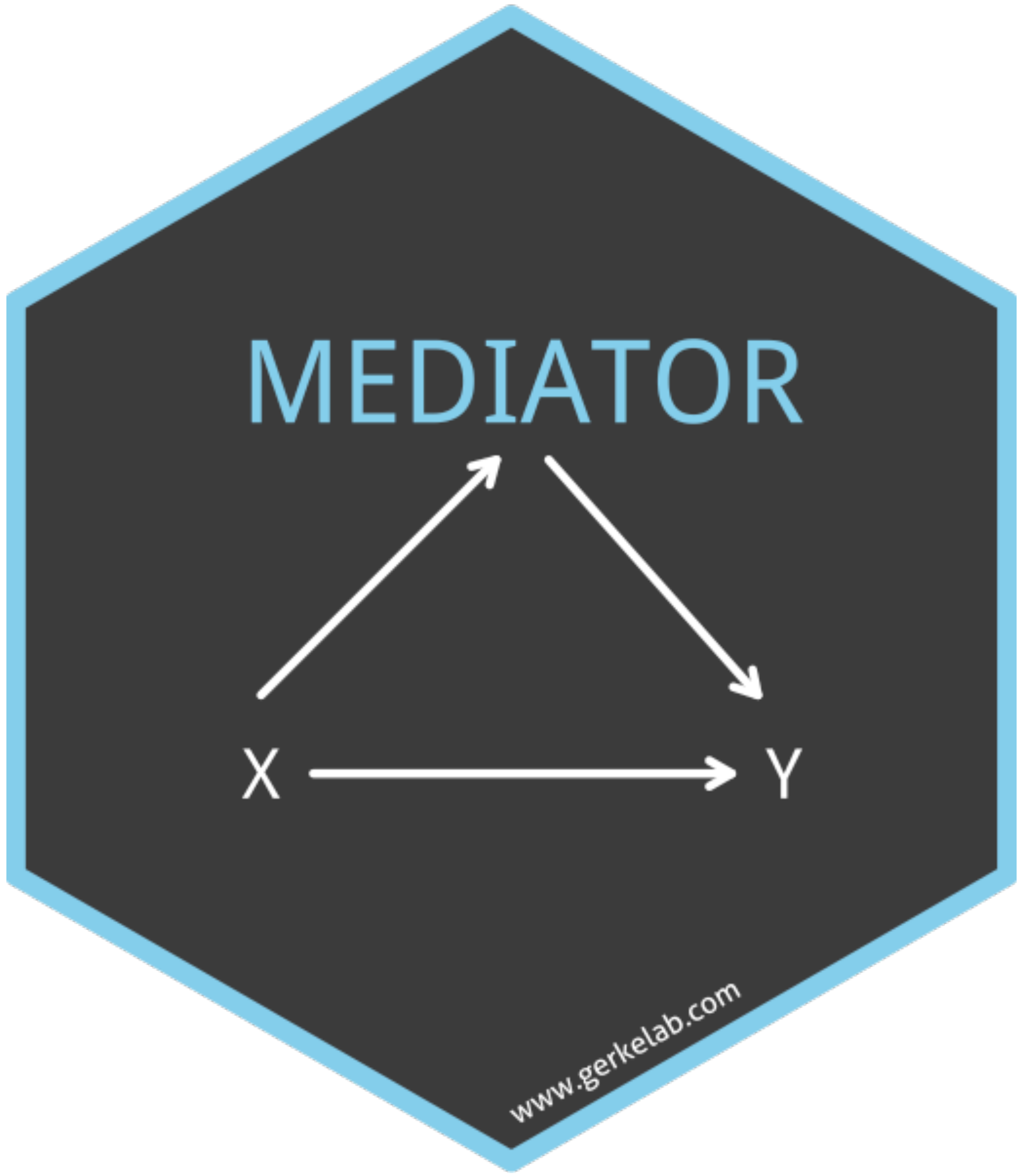


mediator

an R package for conducting causal mediation analyses

Jordan Creed

@jhcreed

MOFFITT CANCER CENTER

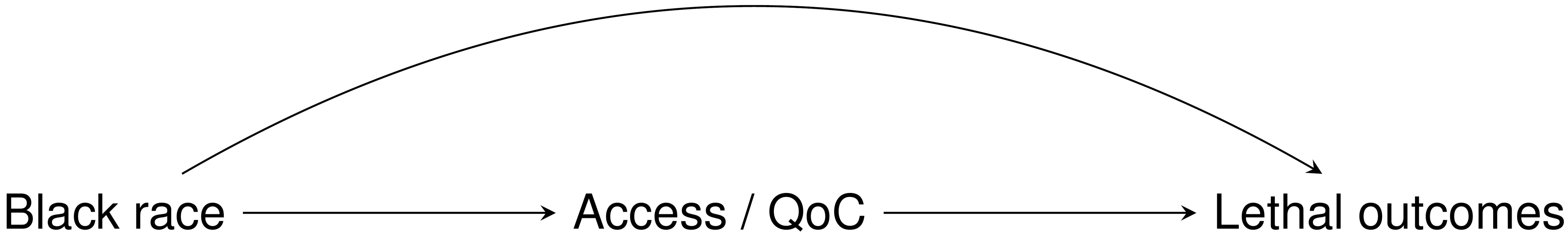
Mediation analysis

Examines an exposure and outcome through an intermediate variable (mediator)

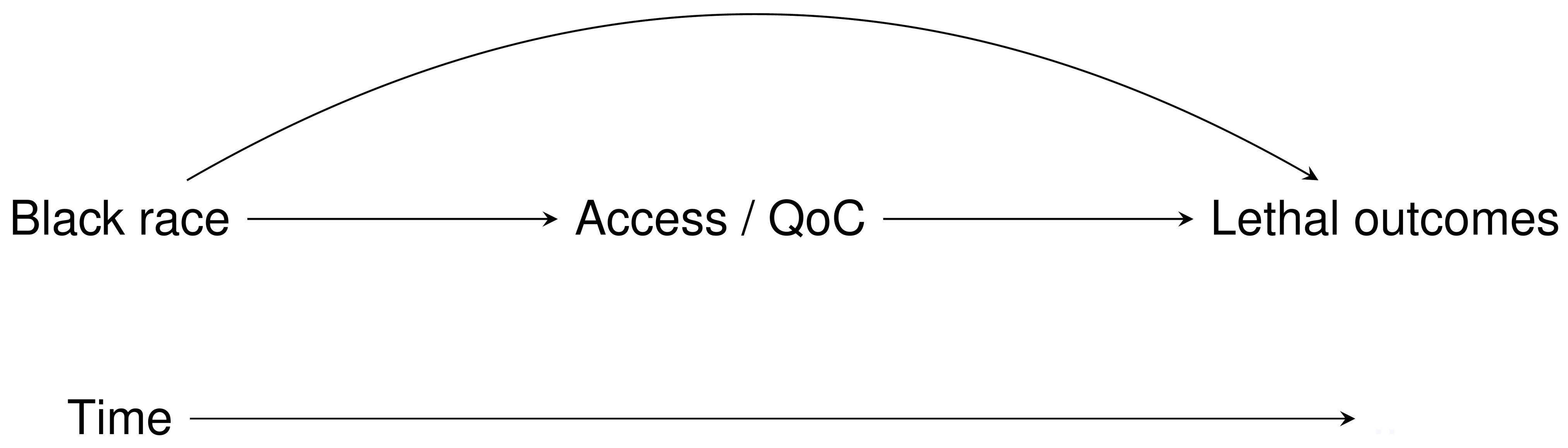
Causal relationship

- Understanding the mechanics behind an association
- NOT prediction

Goal of mediation analysis: Estimate the direct and indirect effects



Diected Acyclic Graphs (DAGs)



DAGs contain variables of interest and common causes

Quickly assess assocations between variables

Rules for reading DAGs

- Modern Epidemiology - Chapter 12
- Causal Inference - https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/

Tool for drawing DAGs: https://apps.gerkelab.com/shinyDAG/

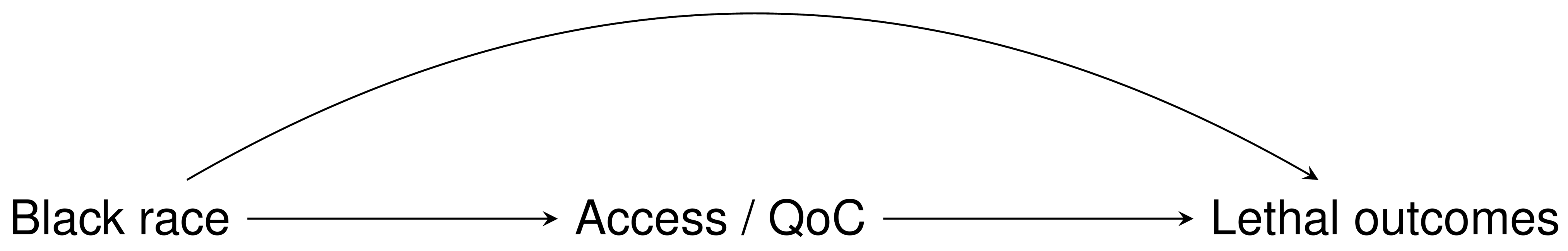
Classic mediation analysis

Baron and Kenny aka the product method

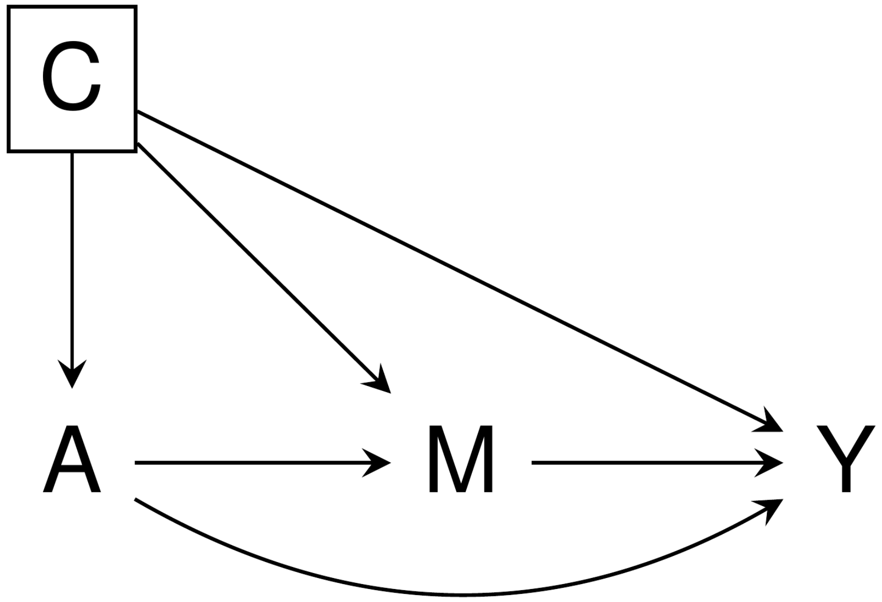
- https://www.sesp.org/files/The%20Moderator-Baron.pdf
- Over 90,000 citations

Criteria to be a mediator:

- Changing the exposure change the mediator (Race -> Access/ QoC)
- The mediator affect the outcome (Access / Qoc -> Lethal outcomes)
- Changing the exposure change the outcome (😬 unless the indirect and direct effects cancel out)
- After controlling for the mediator, the previously significant relationship between the exposure and the outcome is no longer significant (😬 unless it partial mediaton)



Baron and Kenny approach



If A is the exposure, Y the outcome, M the mediator and C the covariates

- Y and M are continuous

Step 1: Fit E[M] = $\beta_0 + \beta_1 a + \beta_2 c$

Step 2: Fit E[Y] = $\theta_0 + \theta_1 a + \theta_2 m + \theta_4 c$

Direct effect = $\theta_1 a$

- Direct effect is the exposure effect on the outcome at a fixed level of the mediator

Indirect effect = $\theta_2 \beta_1$

- Indirect effect is the effect on the outcome of changes of the exposure which operate through the mediator

👉 Does not accommodate exposure-mediator interactions

Causal Inference Methods

Allowing exposure-mediator interactions

Step 1: Fit E[M] = $\beta_0 + \beta_1 a + \beta_2 c$

- same as previously

Step 2: Fit E[Y] = $\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4 c$

Assuming a binary exposure changing level a* = 0 to a = 1

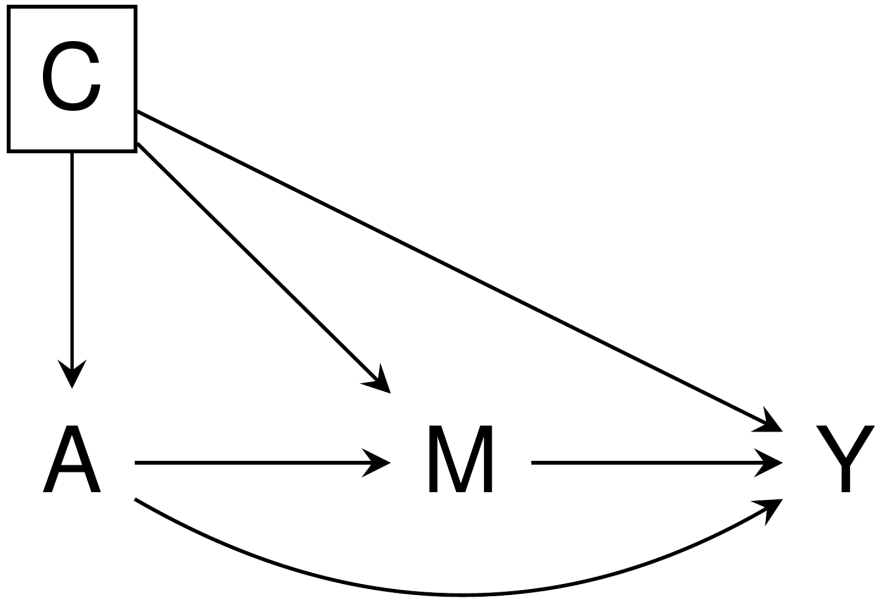
Controlled direct effect = $(\theta_1 + \theta_3 m)(a - a^*)$

Natural direct effect = $\{\theta_1 + \theta_3(\beta_0 + \beta_1 a^* + \beta_2 c)\}(a - a^*)$

Natural indirect effect = $(\theta_2 \beta_1 + \theta_3 \beta_1 a)(a - a^*)$

👉 When no exposure-mediator interaction is present, θ_3 = 0

- CDE = NDE = θ_1 and NIE = $\theta_2 \beta_1$
- Same as the direct and indirect effects in Baron and Kenny



Interpreting the estimates

Controlled direct effect = $(\theta_1 + \theta_3 m)(a - a^*)$

- How much the outcome would change if exposure changed from a$*$ to a and the mediator were controlled at level m in the population

Natural direct effect = $\{\theta_1 + \theta_3(\beta_0 + \beta_1 a^* + \beta_2 c)\}(a - a^*)$

- How much the outcome would change if exposure were set at level a versus a$*$ but for each individual the mediator were kept at the level it would have taken in the absence of exposure

Natural indirect effect = $(\theta_2\beta_1 + \theta_3\beta_1 a)(a - a^*)$

- How much the outcome would change if exposure were controlled at level a but the mediator were changed from the level it would take with a$*$ to the level it would take with a

Total effect of A = NDE + NIE

Mediation or confounding ...

Original Investigation

May 23, 2019

Association of Black Race With Prostate Cancer–Specific and Other-Cause Mortality

Robert T. Dess, MD[1]; Holly E. Hartman, MS[2]; Brandon A. Mahal, MD[3]; et al

Author Affiliations | Article Information

JAMA Oncol. 2019;5(7):975-983. doi:10.1001/jamaoncol.2019.0826

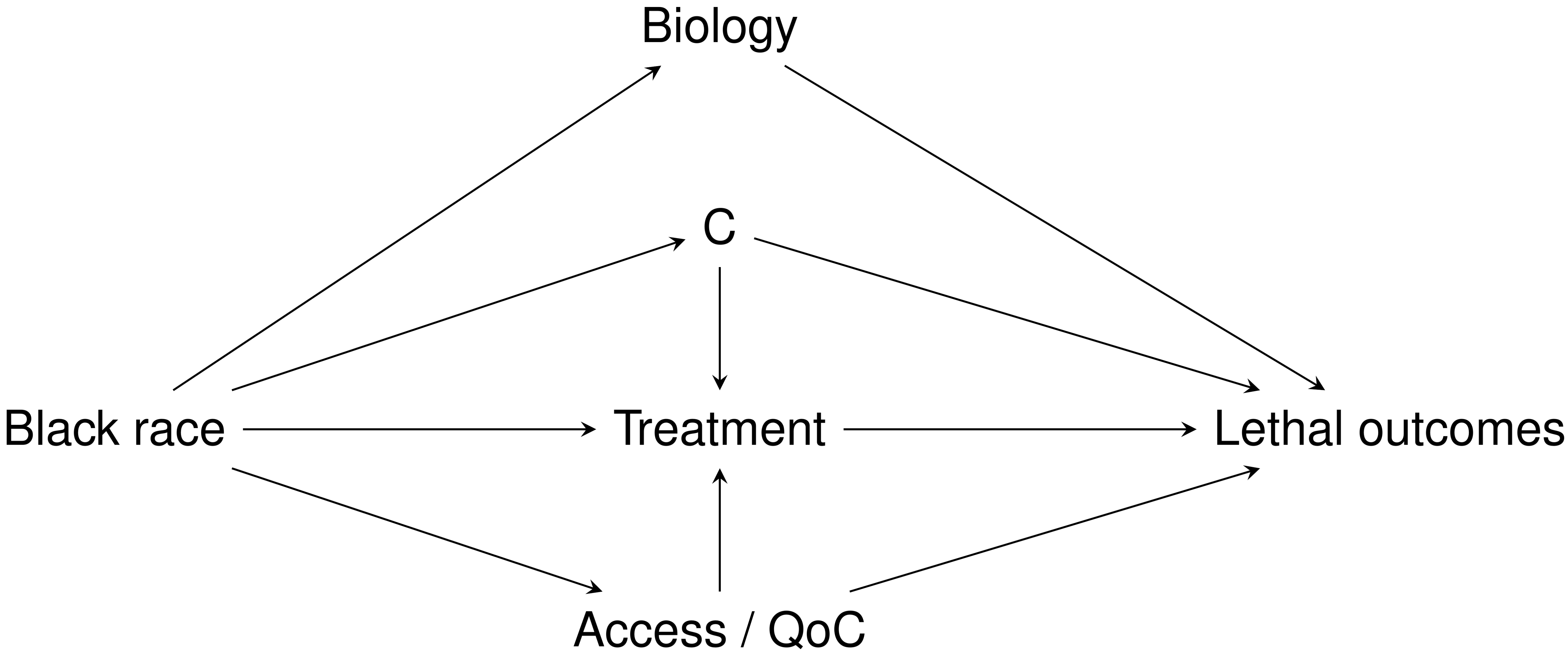
Editorial Comment

Related Articles

Key Points

Question Is black race associated with worse prostate cancer outcomes after controlling for known prognostic variables and access to care?

Mediation or confounding ...



Mediation or confounding ...

Second, our approach highlights the challenges of interpreting population-based data.[24] We adjusted for age, insurance, and a newly released validated socioeconomic status variable. Moreover, we adjusted for cancer- and treatment-related confounders, including the newly released quality-assured PSA values, which were a significant limitation in prior SEER analyses.[25] Inclusion of these crucial prognostic factors substan-

mediator - https://github.com/GerkeLab/mediator

README.md

mediator

build passing codecov 86%

The goal of `mediator` is to conduct causal mediation analysis under the counterfactual framework, allowing interation between the exposure and mediator [Valeri 2013]. Currently, `mediator` estimates the controlled direct effect (CDE), natural direct effect (NDE), natural indirect effect (NIE), total effect (TE) and proportion mediated (PM) and their 95% confidence intervals.

Installation

You can install `mediator` from github with:

```
# install.packages("devtools")
devtools::install_github("gerkelab/mediator")
```

???

Explain the goals for the package, its outputs and installation

mediator - quick start

Required arguments

- `data =` the data for performing the analysis
- `out.model =` fitted model object for the outcome
  - `glm`, `lm` or `coxph`
- `med.model =` fitted model object for the mediator
  - `glm` or `lm`
- `treat =` character string indicating the name of the treatment/exposure variable

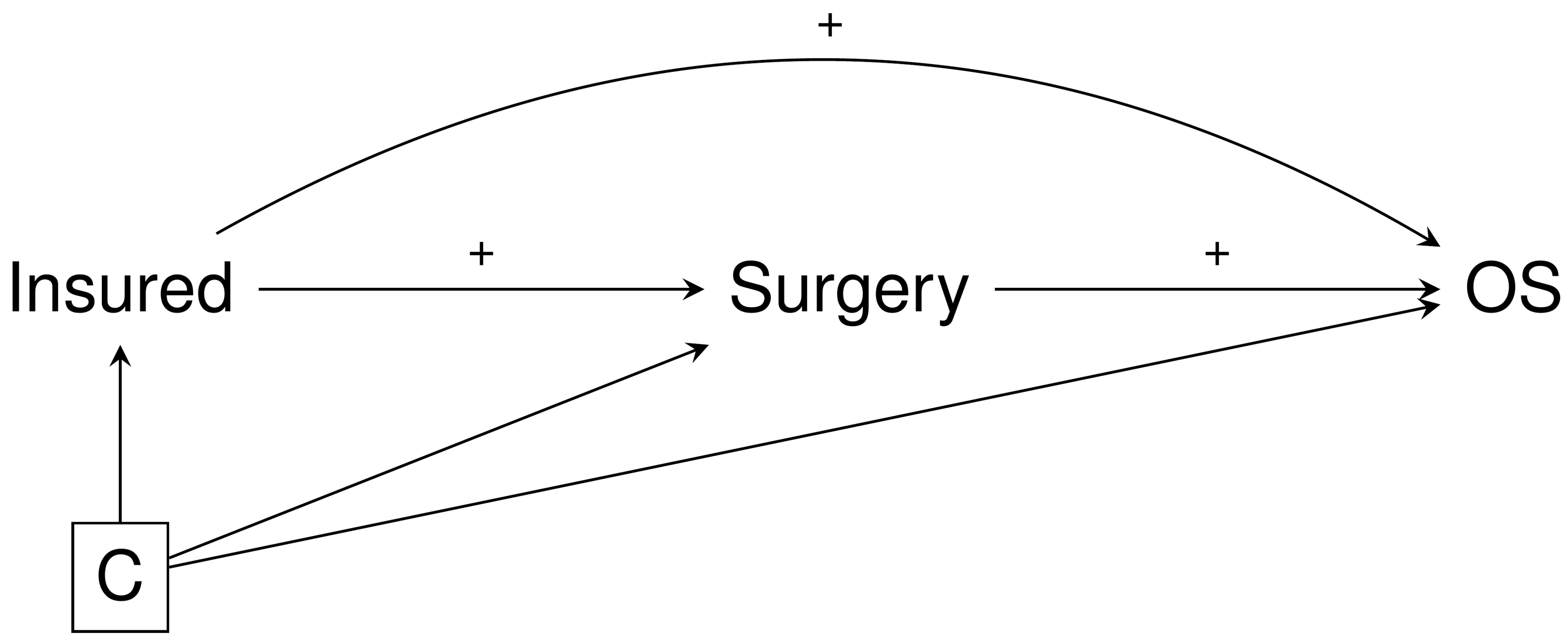
Default arguments

- `a =` numeric value indicating the exposure level
  - default = 1
- `a_star =` numeric value indicating the compared exposure level
  - default = 0
- `m =` numeric value indicating the level of the mediator
  - default = 1
- `boot_rep =` numeric value indicating the number of repetitions to use when utalizing bootstrap to calculate confidence intervals
  - deault = 0 (Delta method)

Is the effect of being insured on overall survival mediated through surgery in prostate cancer …

Based on > 500,000 prostate cancer cases in the National Cancer Data Base

- Those who are insured are more likely to receive surgery than those who are uninsured
- Those who are insured have better overall survival than those who are uninsured
- Those who receive surgery have better overall survival than those who do not receive surgery



Is the effect of being insured on overall survival mediated through surgery in prostate cancer ...

```r
library(mediator)
library(tidyverse)
library(survival)

dat %>%
  # do some data cleaning here ...
  mediator(out.model = coxph(Surv(DX_LASTCONTACT_DEATH_MONTHS, death) ~
                             insured + surgery +
                             PSA + cGleasonSum + AGE + racev2 +
                             (insured*surgery),
                           data = .),
           med.model = glm(surgery ~ insured +
                           PSA + cGleasonSum + AGE + racev2,
                           data = ., family = "binomial"),
                   treat = "insured")
```

Full code available at: https://github.com/jhcreed/bsp2020-mediator

Is the effect of being insured on overall survival mediated through surgery in prostate cancer ...

Effect	Estimate	Lower 95% CI	Upper 95% CI
CDE	**0.61238**	**0.56128**	**0.66814**
NDE	**0.53448**	**0.49786**	**0.57378**
NIE	**0.92258**	**0.91319**	**0.93207**
Total Effect	**0.49310**	**0.45907**	**0.52965**
Proportion Mediated	**0.08163**		

👉 The number of potential estimates for the CDE is equal to the number of levels that the mediator can take

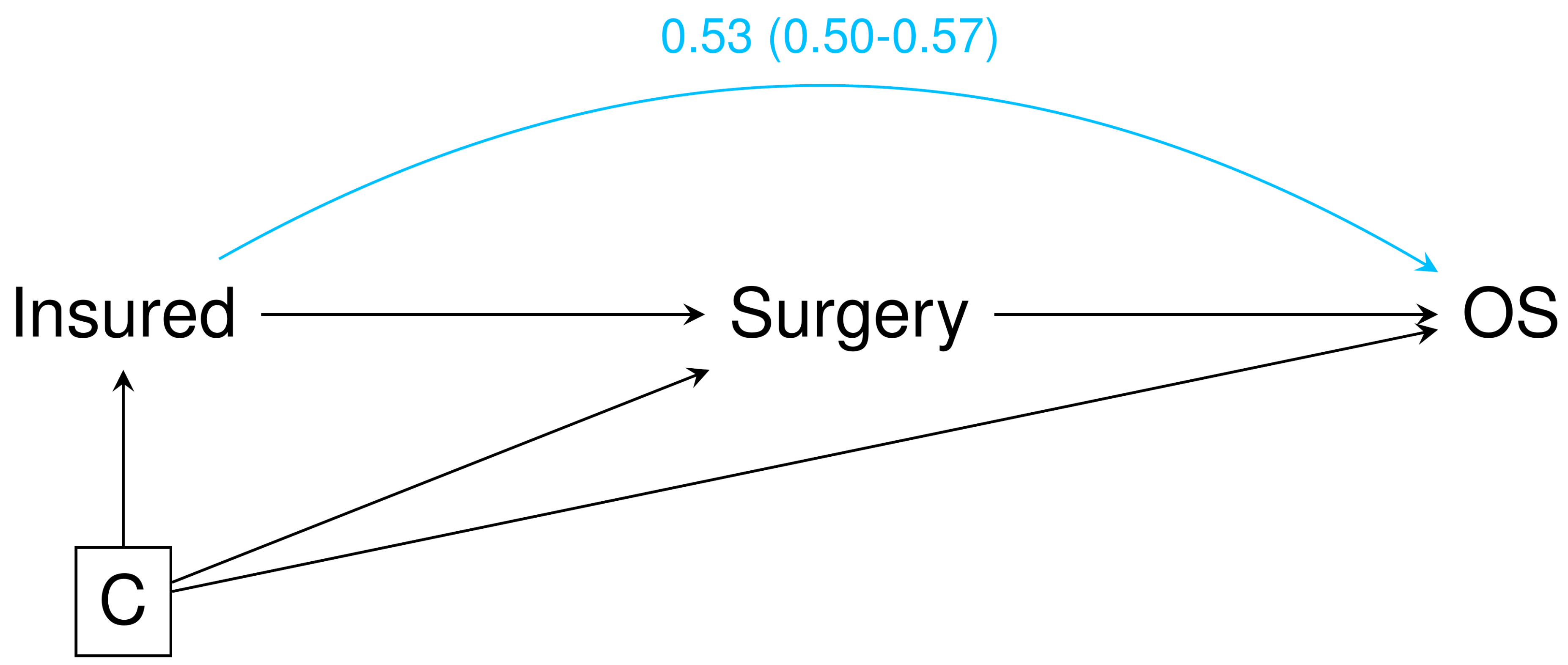
- CDE of being insured compared to uninsured when forcing surgery not to occur is 0.47 (0.42-0.52)
- CDE of being insured compared to uninsured when forcing surgery to occur is 0.61 (0.56-0.67)

Full code available at: https://github.com/jhcreed/bsp2020-mediator

Is the effect of being insured on overall survival mediated through surgery in prostate cancer ...

NDE: the effect of the exposure (insured) on the outcome (overall survival) if the pathway from the exposure to the mediator (surgery) was removed

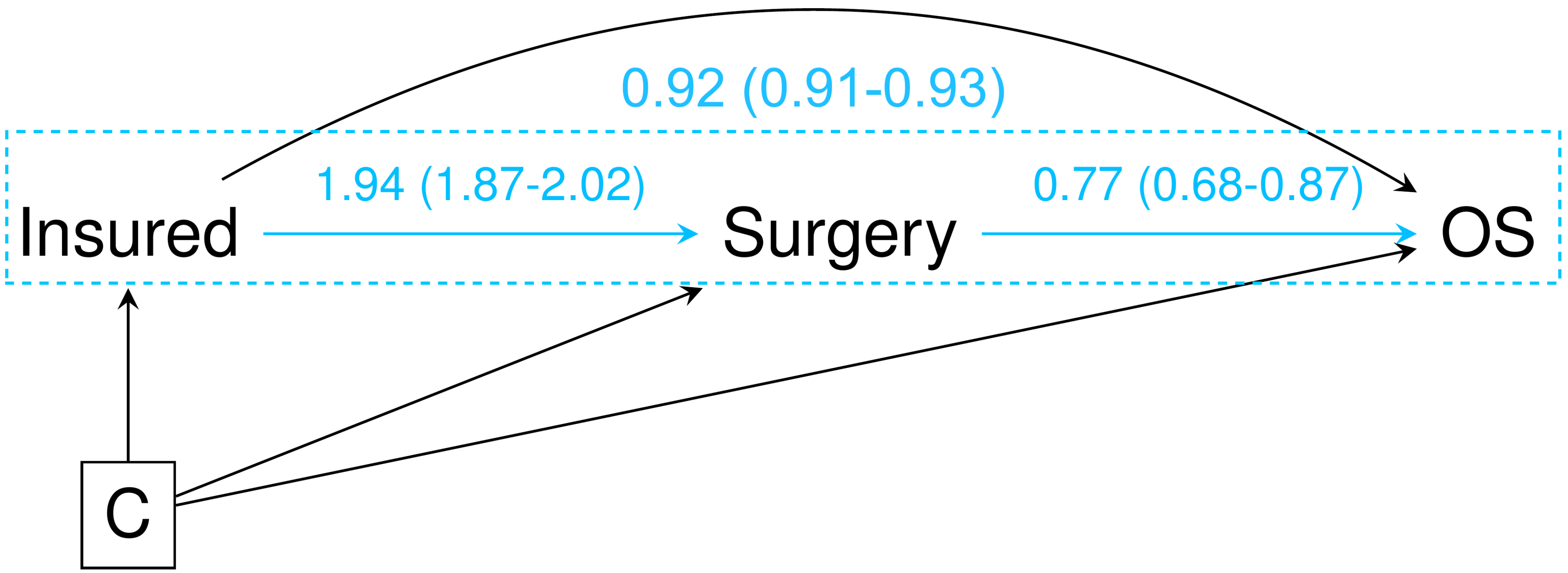
- HR for being insured compared to uninsured, when each individual's surgical status is kept at the level it would take in the absence of insurance status, is 0.53 (0.50-0.57)



Is the effect of being insured on overall survival mediated through surgery in prostate cancer ...

NIE: the effect of the exposure (insured) that operates by changing the mediator (surgery)

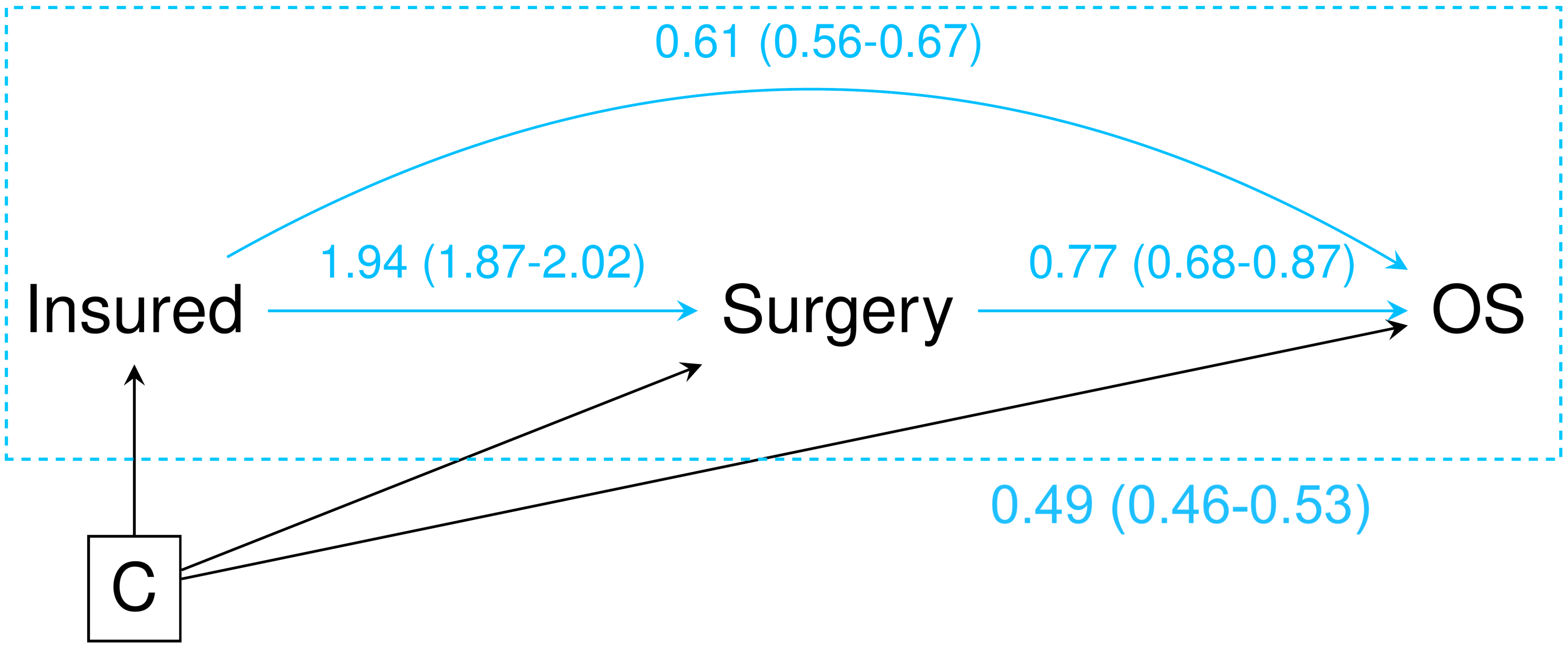
- HR for being insured, if surgical status was changed from the level it would take if insurance status was uninsured to the level if insurance status was insured, is 0.92 (0.91-0.93)



Is the effect of being insured on overall survival mediated through surgery in prostate cancer ...

TE: the overall effect of the exposure (insured) on the outcome (overall survival)

- HR for insured compared to uninsured, overall, is 0.49 (0.46-0.53)



SAS, STATA and R - Oh My!

`mediator` is the sister program of `%mediator` the SAS/SPSS macro developed by Valeri and VanderWeele

Confidence Intervals

- `%mediator` uses hard coded `1.96` and `-1.96` while `mediator` uses `c(-1,1)*qnorm(.975)` for the Delta method
- during bootstrapping, `%mediator` bootstraps effect estimates and CIs while `mediator` only bootstraps the CIs
- minor differences due to rounding

Speed differences

- `mediator` up to 1000x faster than `%mediator` when using bootstrapping

Covariates

- `%mediator` uses dummy variables for multi-level factors while `mediator` allows multi-level and character variables in models

`mediation` R package

- different set of terminology
- different estimation approach

Package == Reproducibility

Easy to share and implement new methods

R packages are more than just a bundle of code : tests, data, documentation, …

hen the average controlled direct effect and the average natural direct and ind…
ffects are given by

exp PER P. 23 OF VALERI/VANDERWEELE 2013

$$OR^{CDE}(m) = (\theta_1 + \theta_3 m)(a - a^*)$$

exp

$$OR^{NDE} = \frac{\exp(\theta_1 a)\{1 + \exp(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a^* + \beta_2' c)\}}{\exp(\theta_1 a^*)\{1 + \exp(\theta_2 + \theta_3 a^* + \beta_0 + \beta_1 a^* + \beta_2' c)\}}$$

$$OR^{NIE} = \frac{\{1 + \exp(\beta_0 + \beta_1 a^* + \beta_2' c)\}\{1 + \exp(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a + \beta_2' c)\}}{\{1 + \exp(\beta_0 + \beta_1 a + \beta_2' c)\}\{1 + \exp(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a^* + \beta_2' c)\}}$$

GitHub

README.md

mediator

build passing | codecov 86%

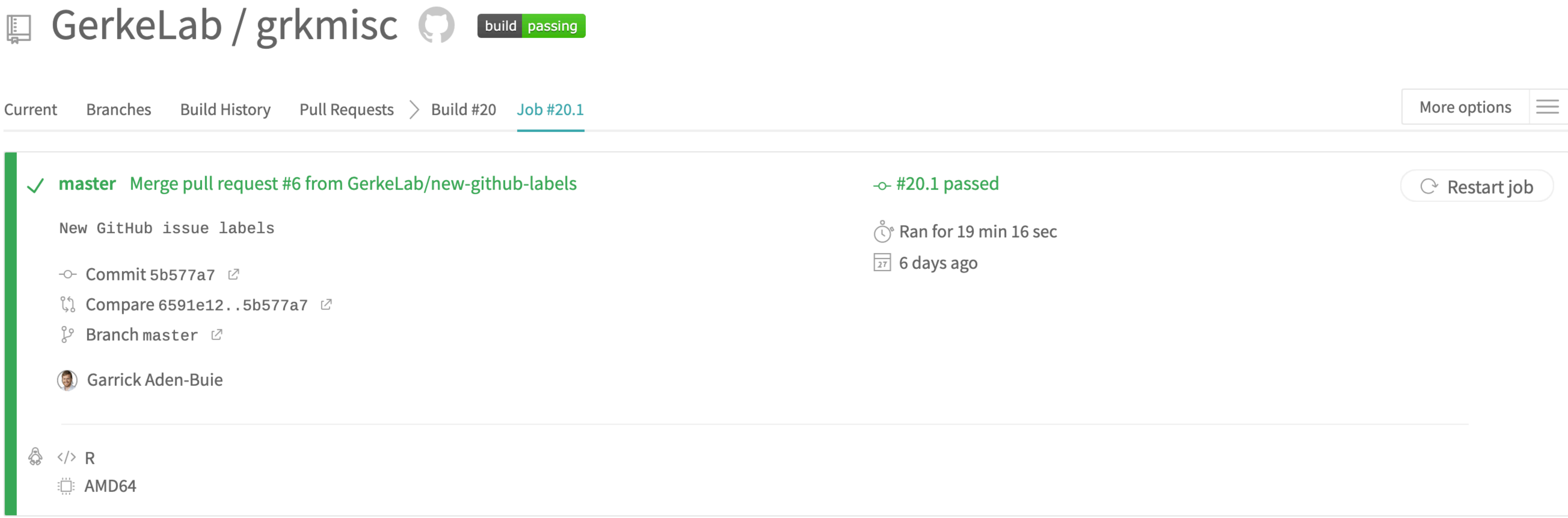
The goal of `mediator` is to conduct causal mediation analysis under the counterfactual framework, allowing interation between the exposure and mediator [Valeri 2013]. Currently, `mediator` estimates the controlled direct effect (CDE), natural direct effect (NDE), natural indirect effect (NIE), total effect (TE) and proportion mediated (PM) and their 95% confidence intervals.

Installation

You can install `mediator` from github with:

```
# install.packages("devtools")
devtools::install_github("gerkelab/mediator")
```

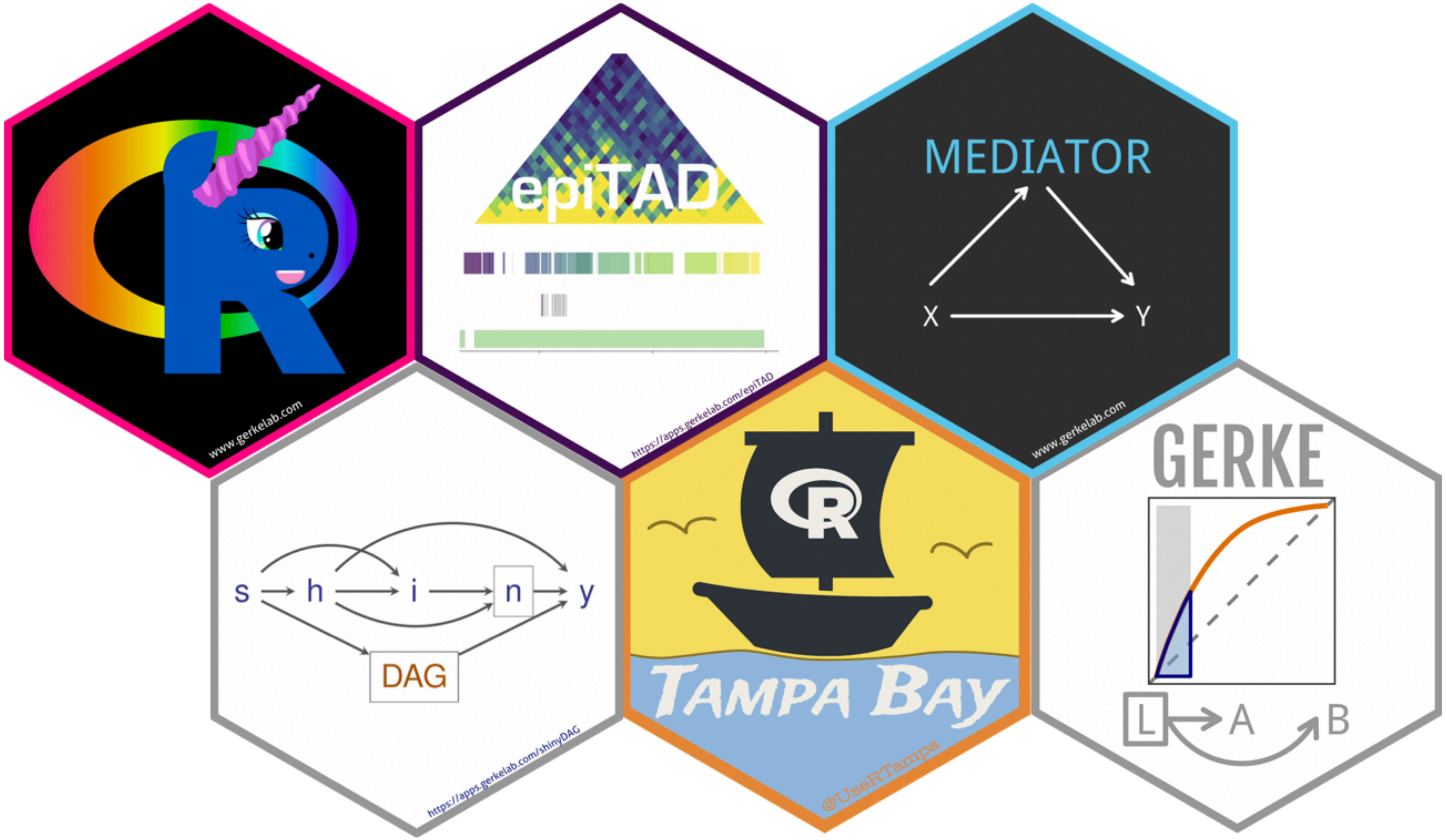
GitHub


GerkeLab / grkmisc
build passing
Current
Branches
Build History
Pull Requests
Build #20
Job #20.1
More options
master
Merge pull request #6 from GerkeLab/new-github-labels
#20.1 passed
Restart job
New GitHub issue labels
Ran for 19 min 16 sec
6 days ago
Commit 5b577a7
Compare 6591e12..5b577a7
Branch master
Garrick Aden-Buie
R
AMD64

GitHub

/ R

Files	≡	●	●	●	Coverage
delta_method.R	31	31	0	0	100.00%
effect_estimates.R	43	43	0	0	100.00%
gammas.R	156	156	0	0	100.00%
mediator.R	112	66	0	46	58.93%
sigma.R	20	20	0	0	100.00%
utils.R	30	25	0	5	83.33%
Folder Totals (6 files)	392	341	0	51	86.99%
Project Totals (6 files)	392	341	0	51	86.99%



www.gerkelab.com
epiTAD
https://apps.gerkelab.com/epiTAD
MEDIATOR
X
Y
www.gerkelab.com
s h i n y
DAG
https://apps.gerkelab.com/shinyDAG
Tampa Bay
@UseRTampa
GERKE
L A B