MEDIATION ANALYSIS

Application of multivariate failure time approach in cancer research

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

Causal Mediation Analysis

- **Objectives**
  - Evaluate the direct effect of race on patients survival disparities
  - Evaluate the indirect effect of race - effect of time to treatment, on patients survival disparities
- **Variables**
  - Mediator(M): time to surgery
  - Exposure(T): race
  - Outcome(Y) - survival time

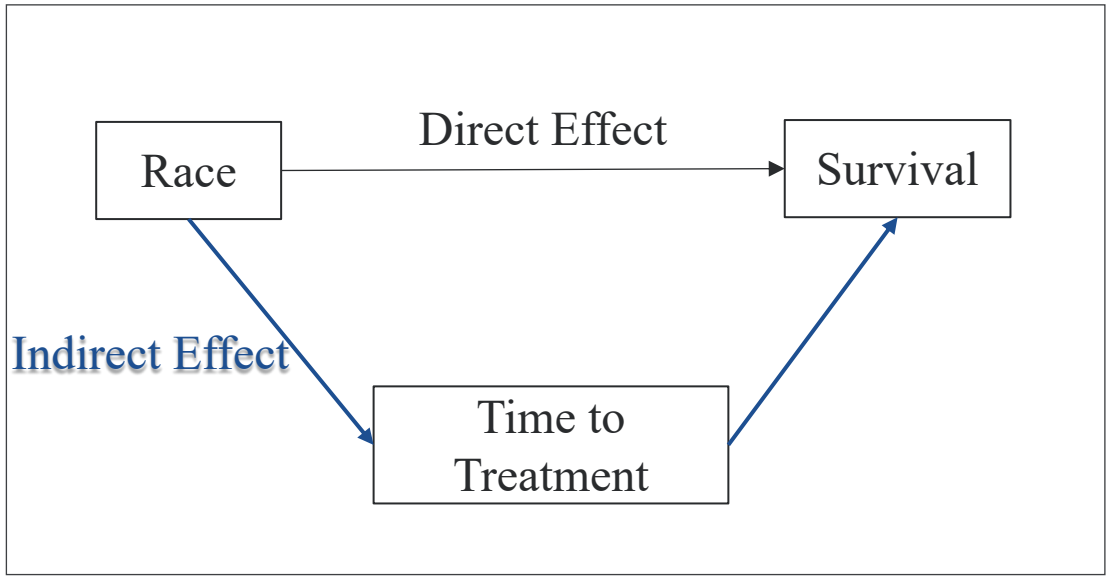

Figure1 Diagram of Causal Relationship

Data and Methods

- **Study Population**
  - non-Hispanic White and non-Hispanic Black cancer patients
  - histologically confirmed diagnosis of colorectal cancer
  - diagnosed above the age of 18
  - followed up for seven years since the date of diagnosis
- **Data Resources**
  - Cancer Care Outcomes Research and Surveillance(CanCORS) cohort study
  - Information on survival time from diagnosis, time from diagnosis to first-line treatment, stage at diagnosis, age, gender, race/ethnicity, etc. was obtained

Nonparametric Estimation

- **Kaplan-Meier Curve**
  - Patients diagnosed at stage IV suffer from the lowest survival rate
  - Lost of follow-up may cause higher censoring rate since 2000 days
- **Log-rank Test**
  - Two-sided p-value < 0.0001
  - Significant survival disparities

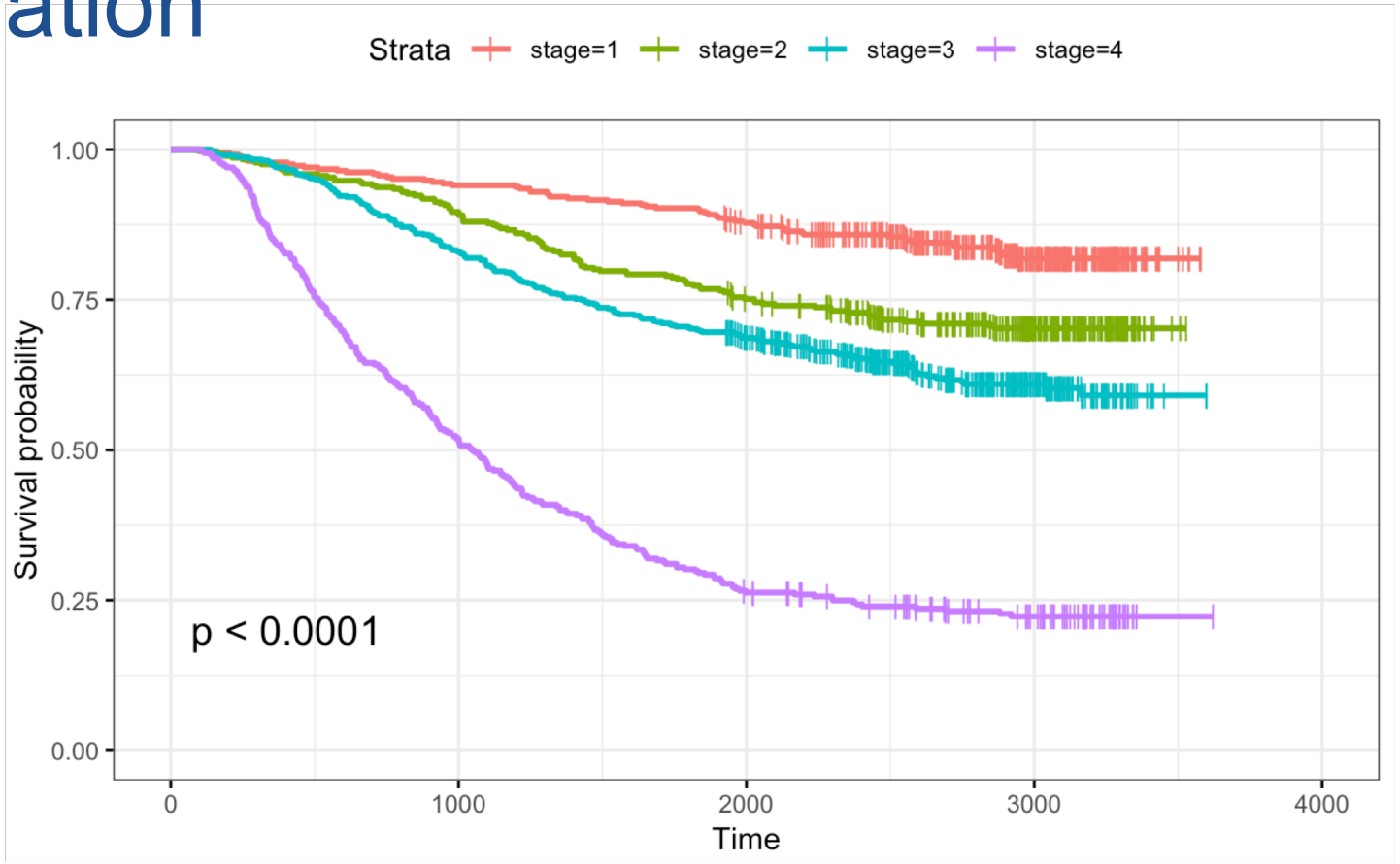

Figure2 KM curve for survival time

Parametric Estimation

- **Accelerated Failure Time**
  - Mediator $\log(\mathbf{T_{trt}}) = \beta_0 + \beta_1 I(race = 1) + \boldsymbol{\beta_2 X} + \epsilon_{\mathrm{m}}$
  - Outcome $\log(\mathrm{Surv}) = \theta_0 + \theta_1 I(race = 1) + \theta_2 \boldsymbol{T_{trt}} + \theta_3 I(race = 1) T_{trt} + \boldsymbol{\theta_4 X} + \epsilon_{\mathrm{y}}$
  - ϵ_{m} and ϵ_{y} respectively represent the random error in mediator model and outcome model
- **Effect Decomposition**
  - Direct effect $\log(Surv)^{DE} = \theta_1 + \theta_3 \, (\beta_0 + \boldsymbol{\beta_2 X})$
  - Indirect effect $\log(Surv)^{IE} = \beta_1(\theta_2 + \theta_3)$

Modeling Results

```
> fit_medi_int2=survreg(Surv(rel_days_rca_surg, surgery_status)~gender*(income+race),data=cancors)
> summary(fit_medi_int2)

Call:
survreg(formula = Surv(rel_days_rca_surg, surgery_status) ~ gender *
    (income + race), data = cancors)
                     Value Std. Error     z       p
(Intercept)         5.4917     0.1471 37.34 < 2e-16
genderMale          0.8019     0.1955  4.10 4.1e-05
income2             1.0106     0.2574  3.93 8.7e-05
income3             0.7146     0.2795  2.56 0.01058
race                0.5515     0.2360  2.34 0.01947
genderMale:income2 -1.2355     0.3311 -3.73 0.00019
genderMale:income3 -0.6548     0.3552 -1.84 0.06523
genderMale:race    -0.6394     0.3218 -1.99 0.04698
Log(scale)          0.0687     0.0439  1.56 0.11778

Scale= 1.07

Weibull distribution
Loglik(model)= -2059   Loglik(intercept only)= -2070.5
        Chisq= 23.01 on 7 degrees of freedom, p= 0.0017
Number of Newton-Raphson Iterations: 5
n= 334
```

$\beta_1 = 0.5515$

Modeling Results

```
> fit_out=survreg(Surv(surv, cens)~income+race+gender+rel_days_rca_surg,data=cancors)
> summary(fit_out)

Call:
survreg(formula = Surv(surv, cens) ~ income + race + gender +
    rel_days_rca_surg, data = cancors)
                      Value Std. Error     z      p
(Intercept)        7.505546   0.118978 63.08 <2e-16
income2           -0.034253   0.144425 -0.24  0.813
income3            0.227585   0.161746  1.41  0.159
race              -0.279738   0.141092 -1.98  0.047
genderMale        -0.160183   0.119106 -1.34  0.179
rel_days_rca_surg  0.000236   0.000113  2.09  0.037
Log(scale)        -0.076394   0.051318 -1.49  0.137

Scale= 0.926

Weibull distribution
Loglik(model)= -2174.2   Loglik(intercept only)= -2180.9
        Chisq= 13.29 on 5 degrees of freedom, p= 0.021
Number of Newton-Raphson Iterations: 5
n= 334
```

$\theta_1 = -0.27974$

$\theta_2 = 0.00024$

Discussion

- **Pros**
  - Successfully developed the framework to conduct mediation analysis for **censored data**
  - Quantified the effect of **time to treatment** in explaining racial survival disparities
- **Cons**
  - Only stage IV patients were included for analysis
  - Confounders were not fully identified
- **Improvement**
  - Perform **semi-parametric modeling** in the future analysis
  - Consider **multiple mediators** (e.g. stage of diagnosis)