Individual treatment effect estimation in the presence of unobserved confounding using proxies

a cohort study in stage III non-small cell lung cancer

Scientific Reports, 2022

Wouter A.C. van Amsterdam, MD, PhD

March 1, 2023

Affiliations & Collaborators

- University Medical Center Utrecht
    - Joost Verhoeff, MD PhD*
    - Netanja Harlianto
    - Gijs Bartholomeus
    - Pim de Jong, MD PhD*
    - Tim Leiner, MD PhD*
    - Anne van Lindert, MD
    - Rene Eijkemenas, PhD
- New York University
    - Aahlad Puli
    - Rajesh Ranganath, PhD*
- Me
    - Then: PhD student at UMCU
    - Now: Sr. research scientist at Babylon Health
    - 2023: Assistent Prof. at UMCU

*PhD advisors

Introduction

Motivation

- important decision in cancer care: how to treat?

Motivation

- important decision in cancer care: how to treat?
- weigh probability of benefit versus probability of harm for available treatments

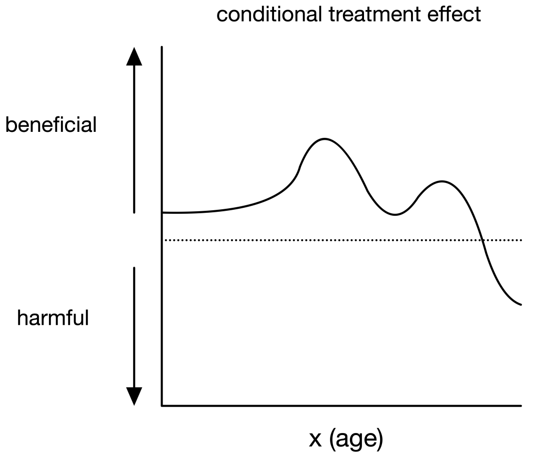
Motivation

- important decision in cancer care: how to treat?
- weigh probability of benefit versus probability of harm for available treatments
- (relative) efficacy of treatments determined in RCTs for a certain population (average treatment effect: ATE)

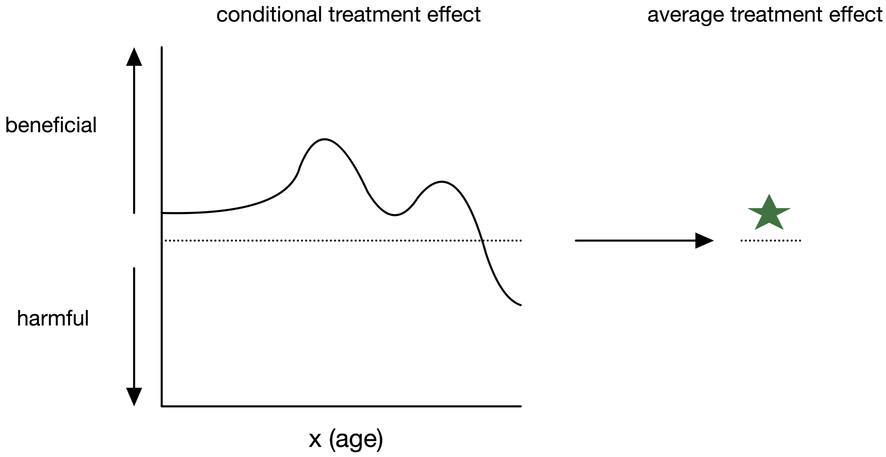
Motivation

- important decision in cancer care: how to treat?
- weigh probability of benefit versus probability of harm for available treatments
- (relative) efficacy of treatments determined in RCTs for a certain population (average treatment effect: ATE)
- for personalized treatment decision making, want to known treatment effect conditional on patient characteristics X (conditional average treatment effect: CATE)

Ideal scenario: CATE known



RCT estimates average treatment effect, which is average CATE



Why estimate CATE in observational data

CATE estimation requires too many randomized participants to have sufficient power

Potential benefits of observational data:

- Bigger sample size

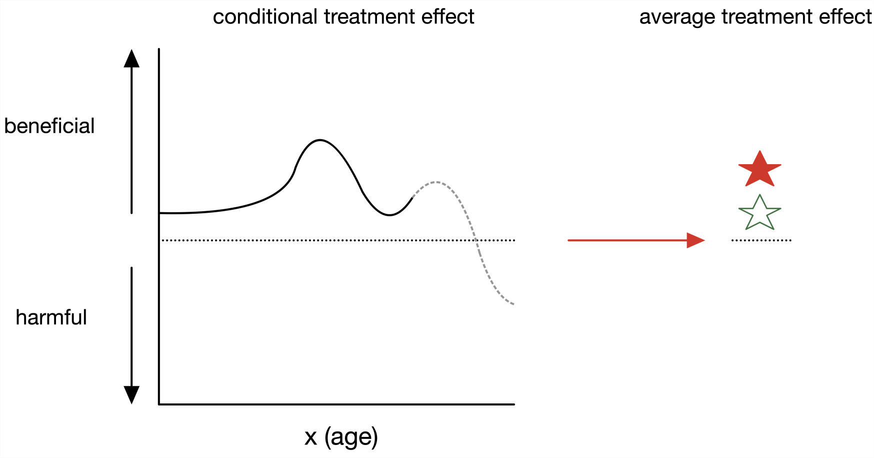
Why estimate CATE in observational data

CATE estimation requires too many randomized participants to have sufficient power

Potential benefits of observational data:

- Bigger sample size
- New 'biomarkers'

RCT gives us ATE in selected population



Why estimate CATE in observational data

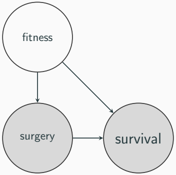
CATE estimation requires too many randomized participants to have sufficient power

Potential benefits of observational data:

- Bigger sample size
- New ‘biomarkers’

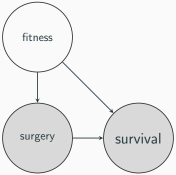
Wider population than RCT

Confounding: surgery vs chemoradiation



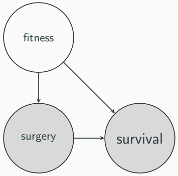
- Surgical treatment is preferred for patients who are fit for surgery

Confounding: surgery vs chemoradiation



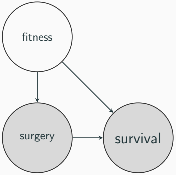
- Surgical treatment is preferred for patients who are fit for surgery
- Patients who are fit for surgery have better cancer survival, regardless of treatment

Confounding: surgery vs chemoradiation



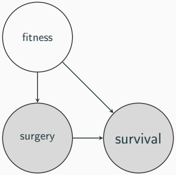
- Surgical treatment is preferred for patients who are fit for surgery
- Patients who are fit for surgery have better cancer survival, regardless of treatment
- survival(surgery=1)

Confounding: surgery vs chemoradiation



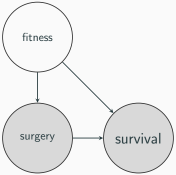
- Surgical treatment is preferred for patients who are fit for surgery
- Patients who are fit for surgery have better cancer survival, regardless of treatment
- survival(surgery=1)
- $\neq$ survival(do(surgery=1))

Confounding: surgery vs chemoradiation



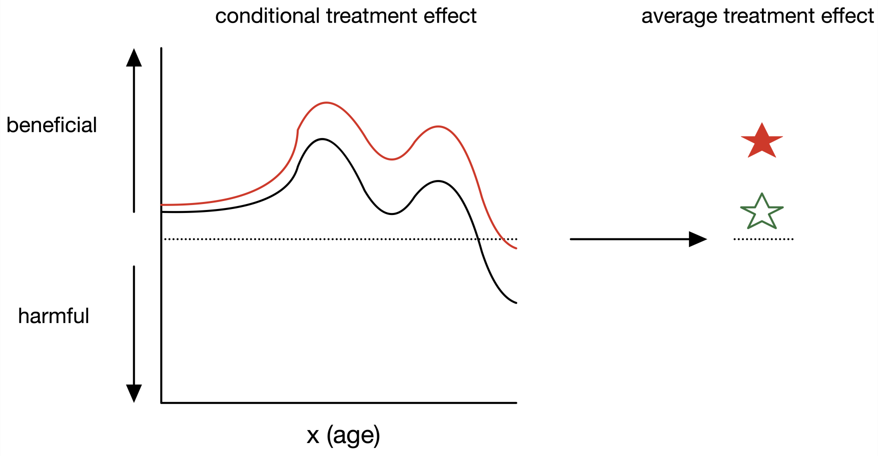
- Surgical treatment is preferred for patients who are fit for surgery
- Patients who are fit for surgery have better cancer survival, regardless of treatment
- survival(surgery=1)
- $\neq$ survival(do(surgery=1))
- Per-fitness group survival estimation resolves confounding bias (stratified / standardized / conditional / adjusted / back-door / point-treatment g-formula)

Confounding: surgery vs chemoradiation



- Per-fitness group survival estimation resolves confounding bias (stratified / standardized / conditional / adjusted / back-door / point-treatment g-formula)
- **Requires measurement of overall fitness**

CATE estimation from observational data when some confounders are unknown



Causal inference from observational data requires (additional) assumptions

- Confounders known and measured (conditional ignorability)
- Positivity: $0 < p(t|x) < 1$

Randomization ensures this

The cost of causal inference

RCTs:

- time

Observational:

Always:

The cost of causal inference

RCTs:

- time
- money

Observational:

Always:

The cost of causal inference

RCTs:

- time
- money
- non-random patient selection

Observational:

Always:

The cost of causal inference

RCTs:

- time
- money
- non-random patient selection

Observational:

- identifying assumptions (no unobserved confoundin

Always:

The cost of causal inference

RCTs:

- time
- money
- non-random patient selection

Observational:

- identifying assumptions (no unobserved confoundin

Always:

- causal inference in practice is costly,

The cost of causal inference

RCTs:

- time
- money
- non-random patient selection

Observational:

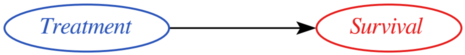
- identifying assumptions (no unobserved confoundin

Always:

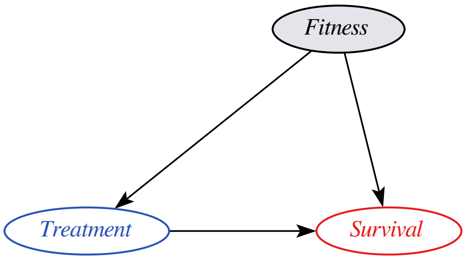
- causal inference in practice is costly,
- **but it's because we're acquiring something of value**

PROTECT

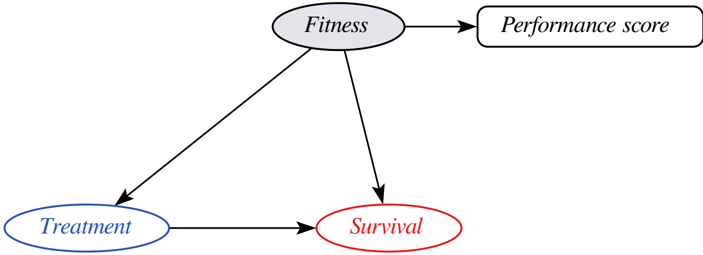
DAG: Directed Acyclic Graph



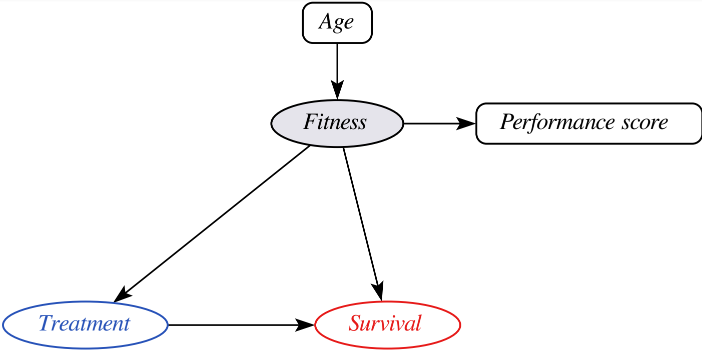
DAG: Directed Acyclic Graph



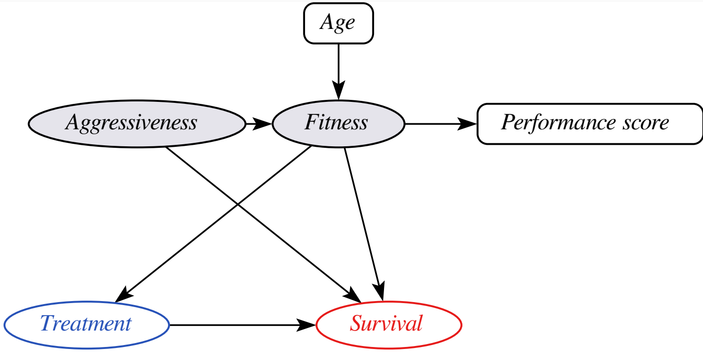
DAG: Directed Acyclic Graph



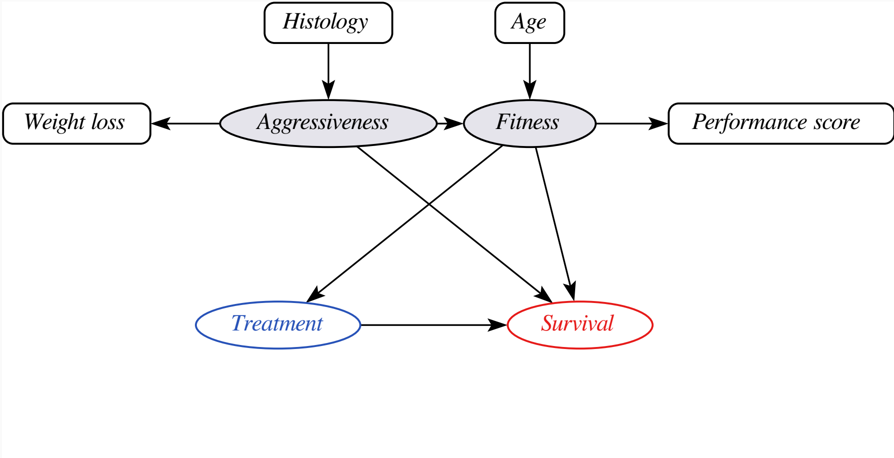
DAG: Directed Acyclic Graph



DAG: Directed Acyclic Graph



DAG: Directed Acyclic Graph



DAG: unobserved confounder Fitness

Can't do adjustment because of unobserved confounder fitness, only have proxy variables (performance score)

- PROTECT: PROxy based individual Treatment EffeCT modeling in cancer

DAG: unobserved confounder Fitness

Can't do adjustment because of unobserved confounder fitness, only have proxy variables (performance score)

- PROTECT: PROxy based individual Treatment EffeCT modeling in cancer
- Estimate latent variable model $\Pr(Y|\mathrm{do}(T), X) = E_{F|X}\Pr(Y|\mathrm{do}(T), X, F) = E_{F|X}\Pr(Y|T, X, F)$

DAG: unobserved confounder Fitness

Can't do adjustment because of unobserved confounder fitness, only have proxy variables (performance score)

- PROTECT: PROxy based individual Treatment EffeCT modeling in cancer
- Estimate latent variable model $\Pr(Y|\text{do}(T), X) = E_{F|X} \Pr(Y|\text{do}(T), X, F) = E_{F|X} \Pr(Y|T, X, F)$
- Question 1: is treatment effect identified?

Identification, definition

Definition 3.2.3 (Identifiability)
Let $Q(M)$ be any computable quantity of a model M. We say that Q is identifiable *in a class $\mathbf{M}$ of models if, for any pairs of models M_1 and M_2 from $\mathbf{M}$, $Q(M_1) = Q(M_2)$ whenever $P_{M_1}(v) = P_{M_2}(v)$. If our observations are limited and permit only a partial set F_M of features (of $P_M(v)$) to be estimated, we define Q to be* identifiable from F_M *if $Q(M_1) = Q(M_2)$ whenever $F_{M1} = F_{M2}$.*

Loosely:

- Defined with respect to model family M

Identification, definition

Definition 3.2.3 (Identifiability)
Let $Q(M)$ be any computable quantity of a model M. We say that Q is identifiable *in a class* $\mathbf{M}$ *of models if, for any pairs of models M_1 and M_2 from $\mathbf{M}$, $Q(M_1) = Q(M_2)$ whenever $P_{M_1}(v) = P_{M_2}(v)$. If our observations are limited and permit only a partial set F_M of features (of $P_M(v)$) to be estimated, we define Q to be* identifiable from F_M *if $Q(M_1) = Q(M_2)$ whenever $F_{M1} = F_{M2}$.*

Loosely:

- Defined with respect to model family M
- If models $m_1, m_2 \in M$ fit the data equally well

Identification, definition

Definition 3.2.3 (Identifiability)
Let $Q(M)$ be any computable quantity of a model M. We say that Q is identifiable *in a class $\mathbf{M}$ of models if, for any pairs of models M_1 and M_2 from $\mathbf{M}$, $Q(M_1) = Q(M_2)$ whenever $P_{M_1}(v) = P_{M_2}(v)$. If our observations are limited and permit only a partial set F_M of features (of $P_M(v)$) to be estimated, we define Q to be* identifiable from F_M *if $Q(M_1) = Q(M_2)$ whenever $F_{M1} = F_{M2}$.*

Loosely:

- Defined with respect to model family M
- If models $m_1, m_2 \in M$ fit the data equally well
- m_1, m_2 agree on the treatment effect

Identification, our case

Here:

- Not identified given data and DAG

Identification, our case

Here:

- Not identified given data and DAG
- Need additional assumptions

Identification, our case

Here:

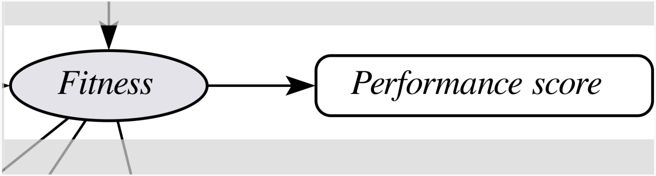
- Not identified given data and DAG
- Need additional assumptions
- Should come from domain knowledge

Identification, our case

Here:

- Not identified given data and DAG
- Need additional assumptions
- Should come from domain knowledge
- Naturally expressed as (parts-of) the data generating process

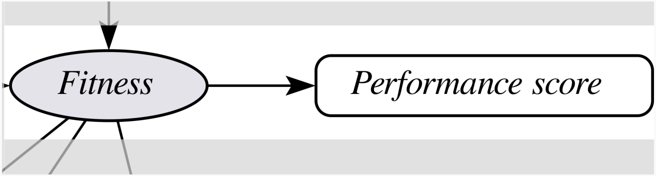
PROTECT step 2: add parametric assumptions



Assumption

- expected WHO performance score always higher with higher fitness

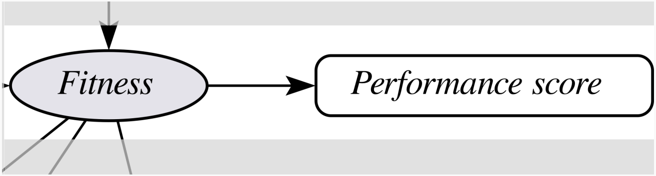
PROTECT step 2: add parametric assumptions



Assumption

- expected WHO performance score always higher with higher fitness
- e.g. logistic regression, fix sign of Fitness

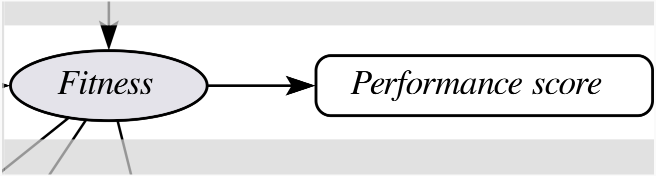
PROTECT step 2: add parametric assumptions



What happens

- model family M now smaller: M'; $m_2 \in M, m_2 \notin M'$

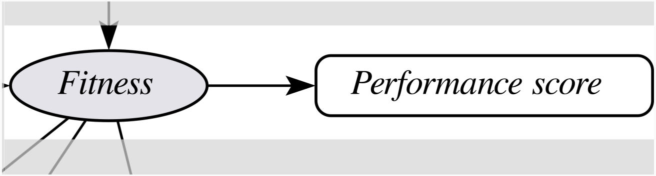
PROTECT step 2: add parametric assumptions



What happens

- model family M now smaller: M'; $m_2 \in M, m_2 \notin M'$
- treatment effect now potentially identified* in model family M' (* checks but no gaurantees)

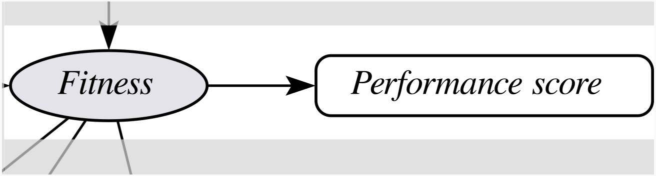
PROTECT step 2: add parametric assumptions



What happens

- model family M now smaller: M'; $m_2 \in M, m_2 \notin M'$
- treatment effect now potentially identified* in model family M' (* checks but no gaurantees)
- can exclude m_2 that says higher fitness $\implies$ worse performance score

PROTECT step 2: add parametric assumptions



What happens

- model family M now smaller: M'; $m_2 \in M, m_2 \notin M'$
- treatment effect now potentially identified* in model family M' (* checks but no gaurantees)
- can exclude m_2 that says higher fitness $\implies$ worse performance score
- 'buying' identification with assumptions we're comfortable with making

Estimation

- Parametric models for all observed variables and latent factors

Estimation

- Parametric models for all observed variables and latent factors
- Latent factor model estimation

Estimation

- Parametric models for all observed variables and latent factors
- Latent factor model estimation
- (Bayesian) (non-linear) structural equation modeling

Estimation

- Parametric models for all observed variables and latent factors
- Latent factor model estimation
- (Bayesian) (non-linear) structural equation modeling
- Identification criterion: unimodal posterior over treatment effect (necessary, not sufficient)

PROTECT step 3: model selection

- may have multiple parameterizations of model

PROTECT step 3: model selection

- may have multiple parameterizations of model
- lead to different treatment effect estimates, which one is right?

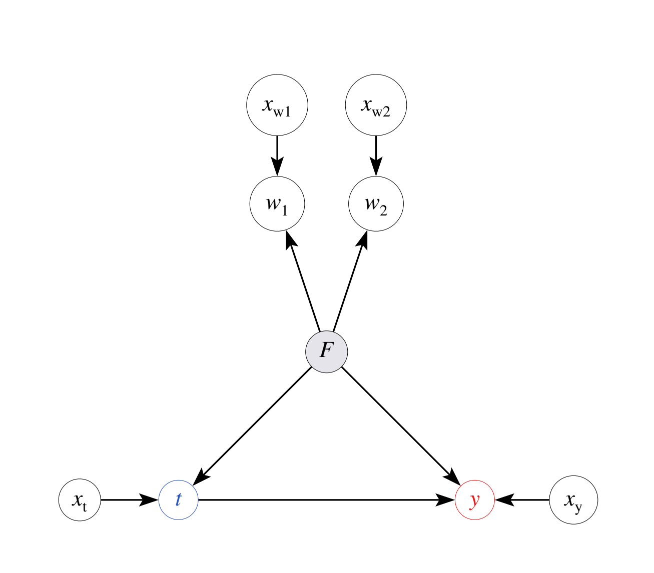
PROTECT step 3: model selection

- may have multiple parameterizations of model
- lead to different treatment effect estimates, which one is right?
- want data-driven model selection procedure

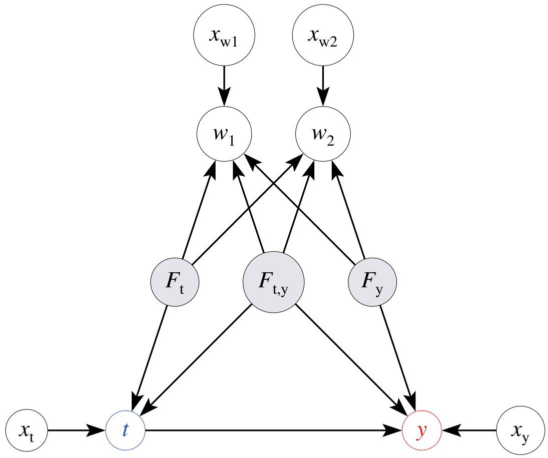
PROTECT step 3: model selection

- may have multiple parameterizations of model
- lead to different treatment effect estimates, which one is right?
- want data-driven model selection procedure
- derive criteria from existing assumptions

PROTECT DAG model selection assumed



PROTECT DAG model selection actual



Potential issues

- Much easier to predict treatment than survival

Potential issues

- Much easier to predict treatment than survival
- Latent factor F might copy treatment (i.e. be F_t)

Potential issues

- Much easier to predict treatment than survival
- Latent factor F might copy treatment (i.e. be F_t)
- Invalid treatment effect estimation as not estimating confounder $F_{t,y}$.

Potential issues

- Much easier to predict treatment than survival
- Latent factor F might copy treatment (i.e. be F_t)
- Invalid treatment effect estimation as not estimating confounder $F_{t,y}$.
- How to test?

Potential issues

- Much easier to predict treatment than survival
- Latent factor F might copy treatment (i.e. be F_t)
- Invalid treatment effect estimation as not estimating confounder $F_{t,y}$.
- How to test?
- On held out data:

Potential issues

- Much easier to predict treatment than survival
- Latent factor F might copy treatment (i.e. be F_t)
- Invalid treatment effect estimation as not estimating confounder $F_{t,y}$.
- How to test?
- On held out data:
  - Model with $F|t, y$ should predict proxy W better than model with $F|t$

Potential issues

- Much easier to predict treatment than survival
- Latent factor F might copy treatment (i.e. be F_t)
- Invalid treatment effect estimation as not estimating confounder $F_{t,y}$.
- How to test?
- On held out data:
  - Model with $F|t, y$ should predict proxy W better than model with $F|t$
  - Model with $F|t, y$ should predict proxy W better than model with $F|y$

Potential issues

- Much easier to predict treatment than survival
- Latent factor F might copy treatment (i.e. be F_t)
- Invalid treatment effect estimation as not estimating confounder $F_{t,y}$.
- How to test?
- On held out data:
    - Model with $F|t, y$ should predict proxy W better than model with $F|t$
    - Model with $F|t, y$ should predict proxy W better than model with $F|y$
- (this is necessary but not sufficient test)

Potential issues

- Much easier to predict treatment than survival
- Latent factor F might copy treatment (i.e. be F_t)
- Invalid treatment effect estimation as not estimating confounder $F_{t,y}$.
- How to test?
- On held out data:
  - Model with $F|t, y$ should predict proxy W better than model with $F|t$
  - Model with $F|t, y$ should predict proxy W better than model with $F|y$
- (this is necessary but not sufficient test)
- final inference: Bayesian Model Average for all models that pass the tests

Final inference

Given models m_j, Bayesian Model Average for all models that pass the tests

PROTECT: proxy based individual treatment effect modeling in cancer

PROTECT has three steps:

1. complete a scaffold DAG for application

PROTECT: proxy based individual treatment effect modeling in cancer

PROTECT has three steps:

1. complete a scaffold DAG for application
2. provide additional domain knowledge to the DAG in the form of parametric assumptions

PROTECT: proxy based individual treatment effect modeling in cancer

PROTECT has three steps:

1. complete a scaffold DAG for application
2. provide additional domain knowledge to the DAG in the form of parametric assumptions
3. follow model selection procedure

Application to state III non-small cell lung cancer

Non-small Cell Lung Cancer

- NSCLC Stage III: 'curate'

Non-small Cell Lung Cancer

- NSCLC Stage III: 'curate'
    - Concurrent chemotherapy and radiotherapy

Non-small Cell Lung Cancer

- NSCLC Stage III: 'curate'
  - Concurrent chemotherapy and radiotherapy
  - Sequential chemotherapy and radiotherapy

Non-small Cell Lung Cancer

- NSCLC Stage III: 'curate'
    - Concurrent chemotherapy and radiotherapy
    - Sequential chemotherapy and radiotherapy
- Randomized controlled trials (RCT): Hazard ratio for concurrent vs sequential: $HR = 0.84$

Non-small Cell Lung Cancer

- NSCLC Stage III: 'curate'
  - Concurrent chemotherapy and radiotherapy
  - Sequential chemotherapy and radiotherapy
- Randomized controlled trials (RCT): Hazard ratio for concurrent vs sequential: $\mathrm{HR} = 0.84$
- Is concurrent better for all patients? Can we quantify effect based on pre-treatment characteristics?

Non-small Cell Lung Cancer

- NSCLC Stage III: 'curate'
  - Concurrent chemotherapy and radiotherapy
  - Sequential chemotherapy and radiotherapy
- Randomized controlled trials (RCT): Hazard ratio for concurrent vs sequential: $\mathrm{HR} = 0.84$
- Is concurrent better for all patients? Can we quantify effect based on pre-treatment characteristics?
- Data overview

Non-small Cell Lung Cancer

- NSCLC Stage III: 'curate'
  - Concurrent chemotherapy and radiotherapy
  - Sequential chemotherapy and radiotherapy
- Randomized controlled trials (RCT): Hazard ratio for concurrent vs sequential: $\text{HR} = 0.84$
- Is concurrent better for all patients? Can we quantify effect based on pre-treatment characteristics?
- Data overview
  - 504 patients (out of 743 stage III) from 8 hospitals

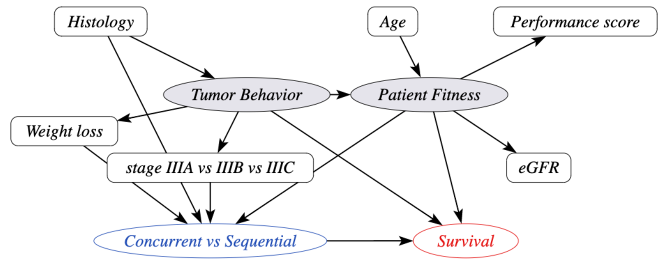
Non-small Cell Lung Cancer

- NSCLC Stage III: 'curate'
  - Concurrent chemotherapy and radiotherapy
  - Sequential chemotherapy and radiotherapy
- Randomized controlled trials (RCT): Hazard ratio for concurrent vs sequential: $\mathrm{HR} = 0.84$
- Is concurrent better for all patients? Can we quantify effect based on pre-treatment characteristics?
- Data overview
  - 504 patients (out of 743 stage III) from 8 hospitals
  - 'intention-to-treat' treatment

Non-small Cell Lung Cancer

- NSCLC Stage III: 'curate'
  - Concurrent chemotherapy and radiotherapy
  - Sequential chemotherapy and radiotherapy
- Randomized controlled trials (RCT): Hazard ratio for concurrent vs sequential: $\text{HR} = 0.84$
- Is concurrent better for all patients? Can we quantify effect based on pre-treatment characteristics?
- Data overview
  - 504 patients (out of 743 stage III) from 8 hospitals
  - 'intention-to-treat' treatment
  - age, sex, eGFR, weight loss, performance score, stage IIIA/IIIB/IIIC, histology type

PROTECT step 1: add proxies / causes of fitness and aggressiveness



PROTECT step 2: specify parametric assumptions for all arrows in the DAG

- 1-dimensional continuous fitness

PROTECT step 2: specify parametric assumptions for all arrows in the DAG

- 1-dimensional continuous fitness
- GLMs, parametric survival model

PROTECT step 2: specify parametric assumptions for all arrows in the DAG

- 1-dimensional continuous fitness
- GLMs, parametric survival model
- estimation details:

PROTECT step 2: specify parametric assumptions for all arrows in the DAG

- 1-dimensional continuous fitness
- GLMs, parametric survival model
- estimation details:
    - dealing with missing data

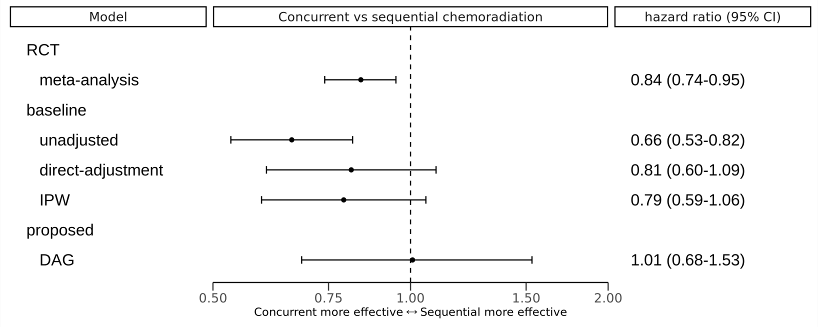
PROTECT step 2: specify parametric assumptions for all arrows in the DAG

- 1-dimensional continuous fitness
- GLMs, parametric survival model
- estimation details:
  - dealing with missing data
  - marginalized DAG (no latent factor for behavior)

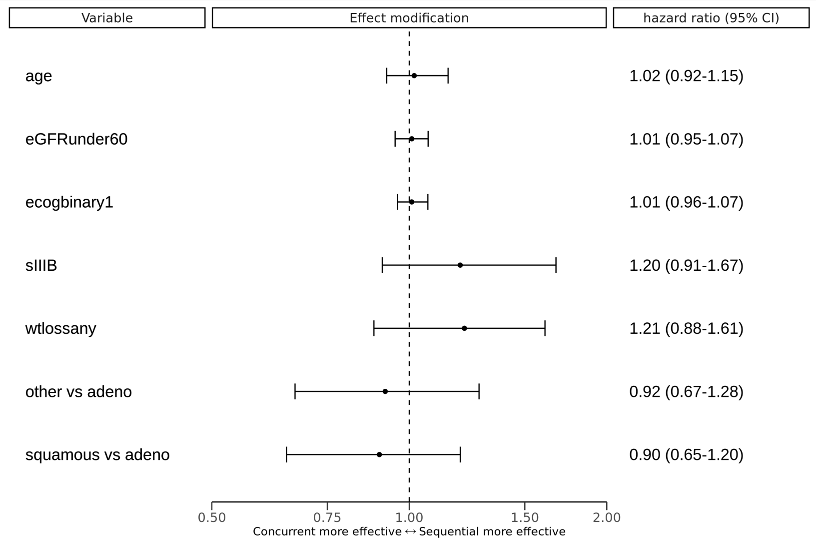
Estimation details

- HMC (NUTS)
- 16 chains
- converged chains (Rhat)
- few divergent transitions (false positives)
- unimodal posterior

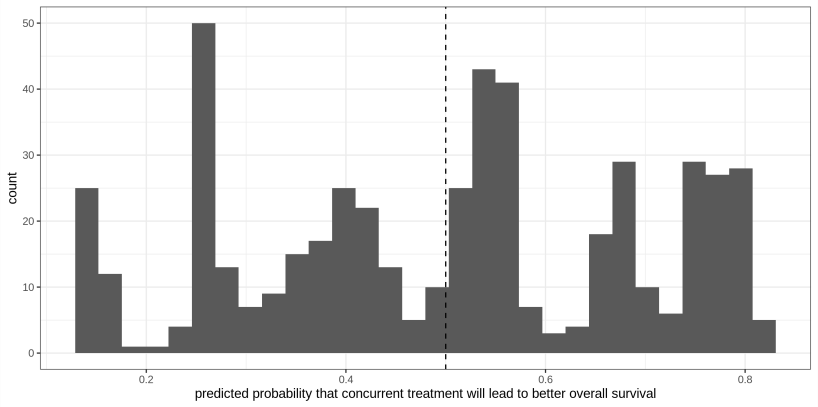
Results: Average Treatment Effects



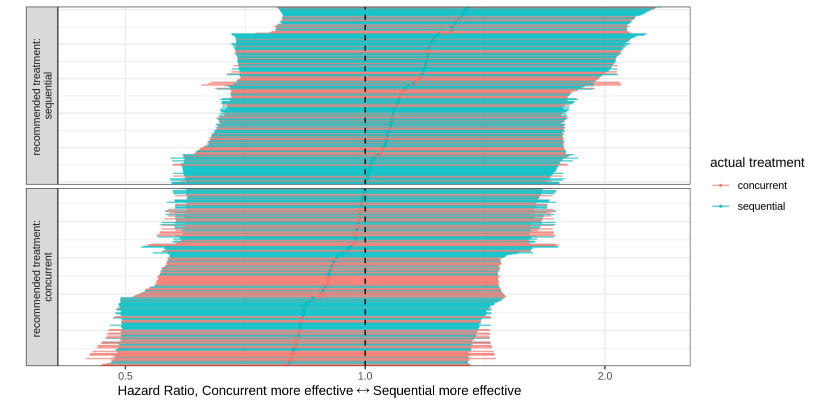
Results: (Friedman) Treatment Effect Modification



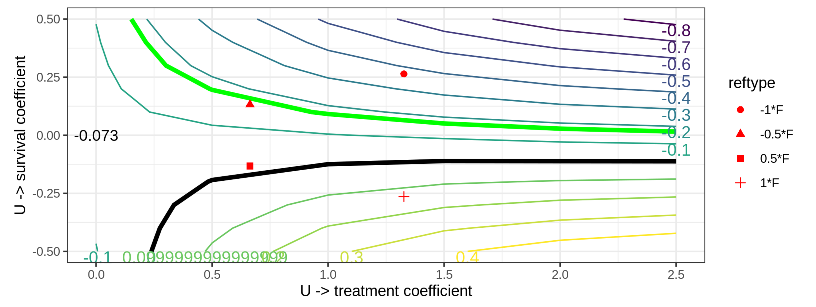
Results: Should treat probabilities



Results: Individual treatment effect estimates



sensitivity to unobserved confounder



Conclusion

- PROTECT: treatment effect estimation with proxies

Conclusion

- PROTECT: treatment effect estimation with proxies
- for NSCLC: treatment effect estimates more credible than standard adjustment, but large confidence intervals

Conclusion

- PROTECT: treatment effect estimation with proxies
- for NSCLC: treatment effect estimates more credible than standard adjustment, but large confidence intervals
- point estimate smaller than RCT estimate, robust to unobserved confounders of reasonable strength

Reflections

- cost of causal inference (CATEs, most valueable for treatment decision making):

Reflections

- cost of causal inference (CATEs, most valueable for treatment decision making):
    - RCTs: time/money, selection

Reflections

- cost of causal inference (CATEs, most valueable for treatment decision making):
  - RCTs: time/money, selection
  - observational: assumptions

Reflections

- cost of causal inference (CATEs, most valueable for treatment decision making):
  - RCTs: time/money, selection
  - observational: assumptions
- PROTECT:

Reflections

- cost of causal inference (CATEs, most valueable for treatment decision making):
    - RCTs: time/money, selection
    - observational: assumptions
- PROTECT:
    - can deal with unobserved confounding (if assumptions hold)

Reflections

- cost of causal inference (CATEs, most valueable for treatment decision making):
  - RCTs: time/money, selection
  - observational: assumptions
- PROTECT:
  - can deal with unobserved confounding (if assumptions hold)
  - always need positivity (does it hold in this application?)

Reflections

- cost of causal inference (CATEs, most valueable for treatment decision making):
  - RCTs: time/money, selection
  - observational: assumptions
- PROTECT:
  - can deal with unobserved confounding (if assumptions hold)
  - always need positivity (does it hold in this application?)
  - requires sufficient sample size

Reflections

- cost of causal inference (CATEs, most valueable for treatment decision making):
  - RCTs: time/money, selection
  - observational: assumptions
- PROTECT:
  - can deal with unobserved confounding (if assumptions hold)
  - always need positivity (does it hold in this application?)
  - requires sufficient sample size
  - no: 'from protect import estimatecate; cate = estimatecate(mydata)'