Time-varying treatments and exposures

Consider time-varying treatment or exposure

e.g., Multicenter AIDS Cohort Study

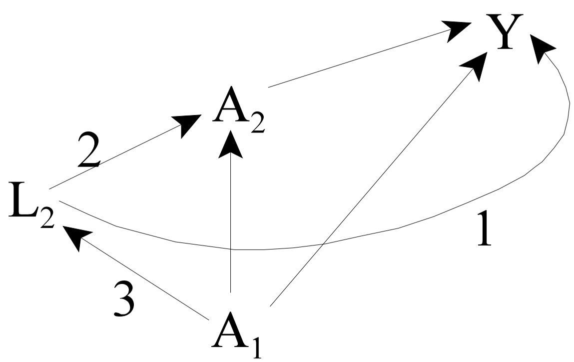
effect of AZT on Kaposi’s Sarcoma (KS)

subjects whose immune system declines (low CD4 count) likely to receive AZT

CD4 count affected by earlier AZT

CD4 count independently associated with KS

sketch DAG



important feature: consider same treatment received at different times as separate treatment; separate nodes in DAG

important for understanding setting

other examples:

studies of industrial exposures

workers in factory exposed to harmful chemicals (e.g., asbestos)

those who develop respiratory difficulties, other severe health problems leave work

receive no additional exposure

sketch DAG

same as above

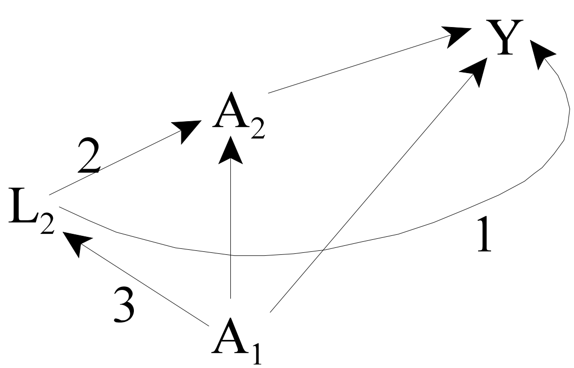
how can one analyze data?

Compare mortality of subjects who receive high doses, low doses; don't control for other variables

high doses look very good, even if no exposure effect; sicker people receive lower doses

Control for employment status, do same comparison

what is wrong with this (consider DAG)?



Controlling for variable on causal pathway; can't get overall effect of treatment

can one estimate effect of A_1? of A_2? Of both?

How can we determine what is correct causal effect, correct methods?

1. Define causal effects

2. Consider whether methods estimate properly under stated assumptions

How can we define causal effects of time-varying treatment?

If consider *joint* effects of multiple treatments $A_1,...,A_m$

consider what would happen if received treatment levels $a_1,...,a_m$

compare $Y^{a_1,...,a_m}$ to $Y^{a_1',...,a_m'}$ for possibly different levels

for general comparisons, what are appropriate choices of comparison levels of treatment?

2 axes:

axis 1: local versus global differences

local: $Y^{a_1,a_2,\ldots,a_m}$ versus $Y^{a_1',a_2,\ldots,a_m}$ or $Y^{a_1,a_2',\ldots,a_m}$

treatments received differ only in 1 element; at 1 point in time

e.g., effect of 1 day of AZT, of 1 day of working in cotton factory

global: $Y^{a_1,a_2,\ldots,a_m}$ versus $Y^{a_1',a_2',\ldots,a_m'}$

treatments differ at many points in time, in many elements

e.g., effect of several years of AZT, of several years of drug therapy, etc.

consider industrial exposure example again

what is global comparison?

$Y^{1,1,\ldots,1}$ versus $Y^{0,0,\ldots,0}$?

Exposure to industrial chemicals may not be feasible once person sick, develops respiratory symptoms

May be of no practical interest to estimate

Formulate alternative:

Compare what would happen if subject exposed to chemicals until develops symptoms to what would happen if not exposed at all

Formulate as comparison of treatment/exposure regimes

Regime: a plan, analogous to protocol in clinical trial, which specifies what treatment a subject is to receive at any point in time

denote by G

G consists of functions $g(\cdot)$ of previous covariate history

let L_k denote covariate levels at time k

use overbars to denote history:

$\overline{L}_k \equiv \{X, L_1, \ldots, L_k\}$ is covariate history through k

$a_k = g_k(\bar{l}_k)$; a_k is treatment assigned by regime at k for subject with covariate history $\bar{l}_k$ through k

example:

for industrial exposure, if l_k denotes respiratory symptoms or severe illness

1 regime: $g_k(\bar{l}_k) = 0$

another regime: $g_k(\bar{l}_k) = 1 \; \text{if} \; l_k = 0$
$\qquad\qquad 0 \; \text{if} \; l_k = 1$

1st regime: *fixed*: not dependent on covariate history

2nd regime: *dynamic*: depends on covariate history compare potential outcomes under different regimes

$Y^G, \; Y^{G'}$

or compare expectations or distributions

$E(Y^G), \; E(Y^{G'})$, etc.

Discuss estimation of causal effects

Assumptions

Estimation methods

- G-computation algorithm
- Marginal structural models
- G-estimation/structural nested models

generalize methods discussed earlier

discuss for continuous outcome Y measured at end of fixed follow-up period

methods generalize to other types of outcomes

Discuss when simpler methods applicable

Will illustrate with numerical example

Generalize condition for estimation, applicable to all methods

Which one?

How?

Ignorable treatment assignment

$\overline{A}_k$ treatment/exposure history

$\overline{L}_k$ covariate history

$\underline{Y}=\{Y^G\}$ set of all potential outcomes

no unmeasured confounders/sequential ignorability:

$pr(A_k|\overline{L}_k,\overline{A}_{k-1},\underline{Y})=pr(A_k|\overline{L}_k,\overline{A}_{k-1})$ (probabilities should also be strictly positive)

describe meaning

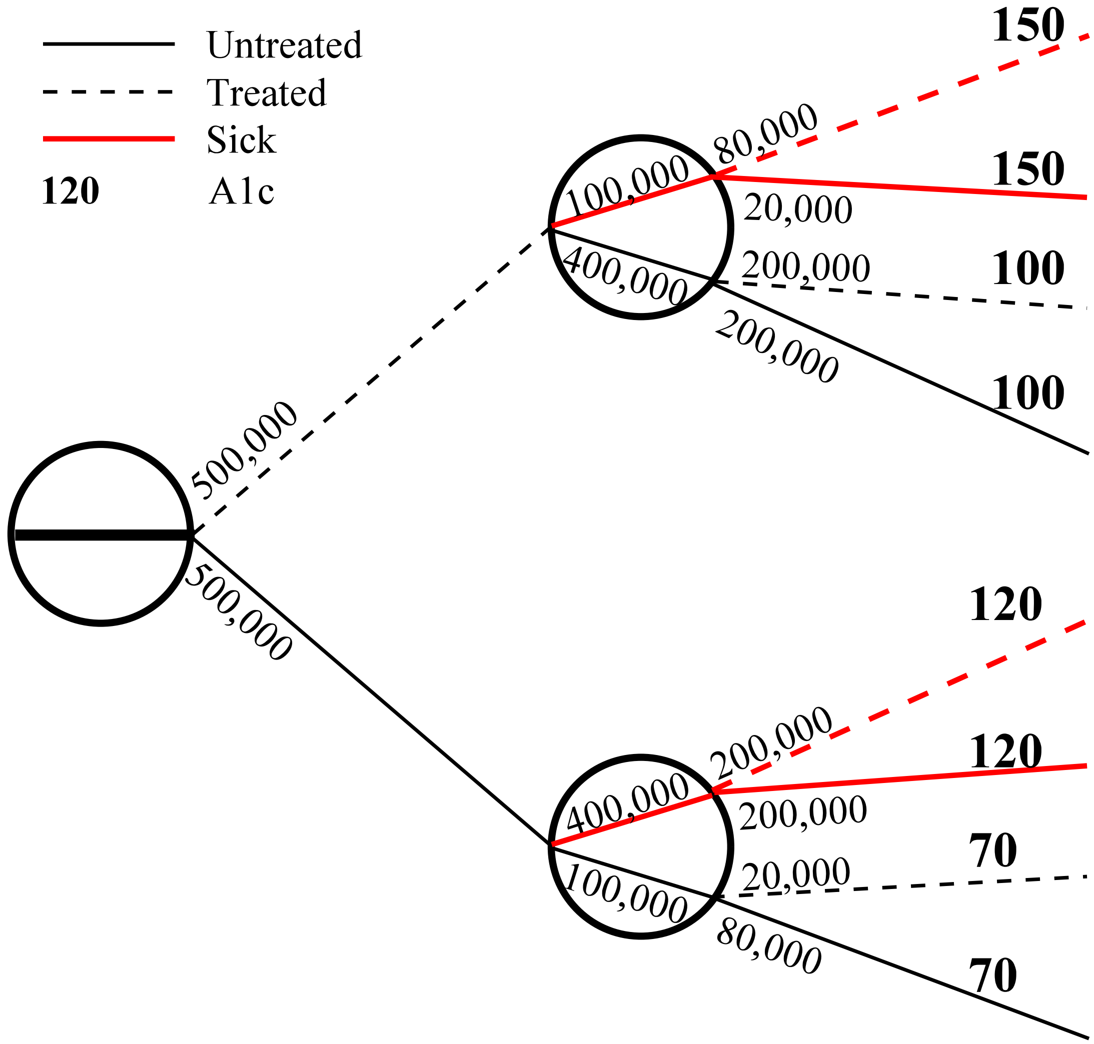
various variants, as with scalar treatments hypothetical example:

observational study of treatment of high blood sugar, measured by hemoglobin A1C (measures long-term burden of blood sugar)

illness at time 1 is associated with higher blood sugar

treatment in each time period decreases blood sugar in A1C by 10 units

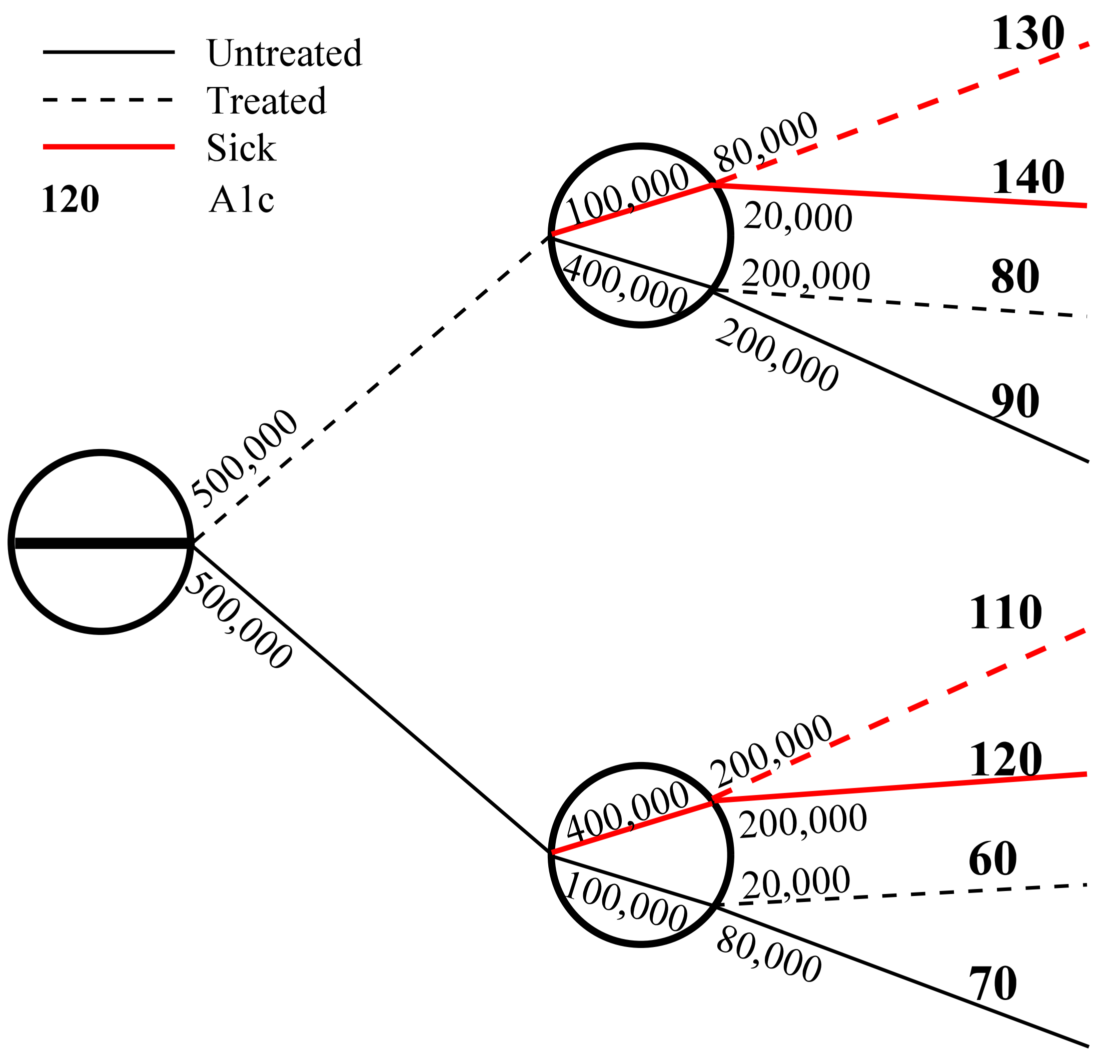
Hypothetical data: potential outcomes



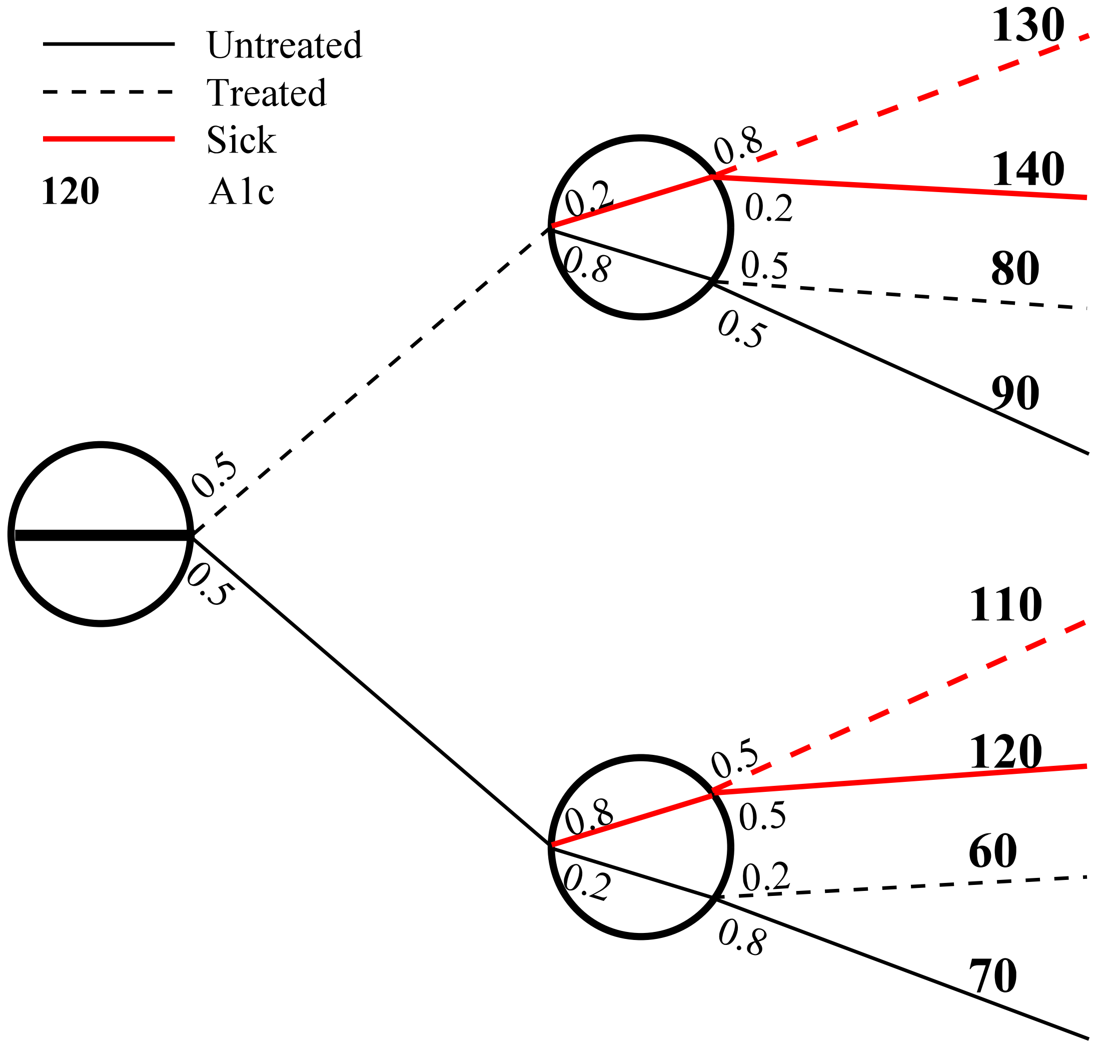
right circumference points: comparable groups

effect of each increment of treatment: lower A1c by 10 (see next panel)

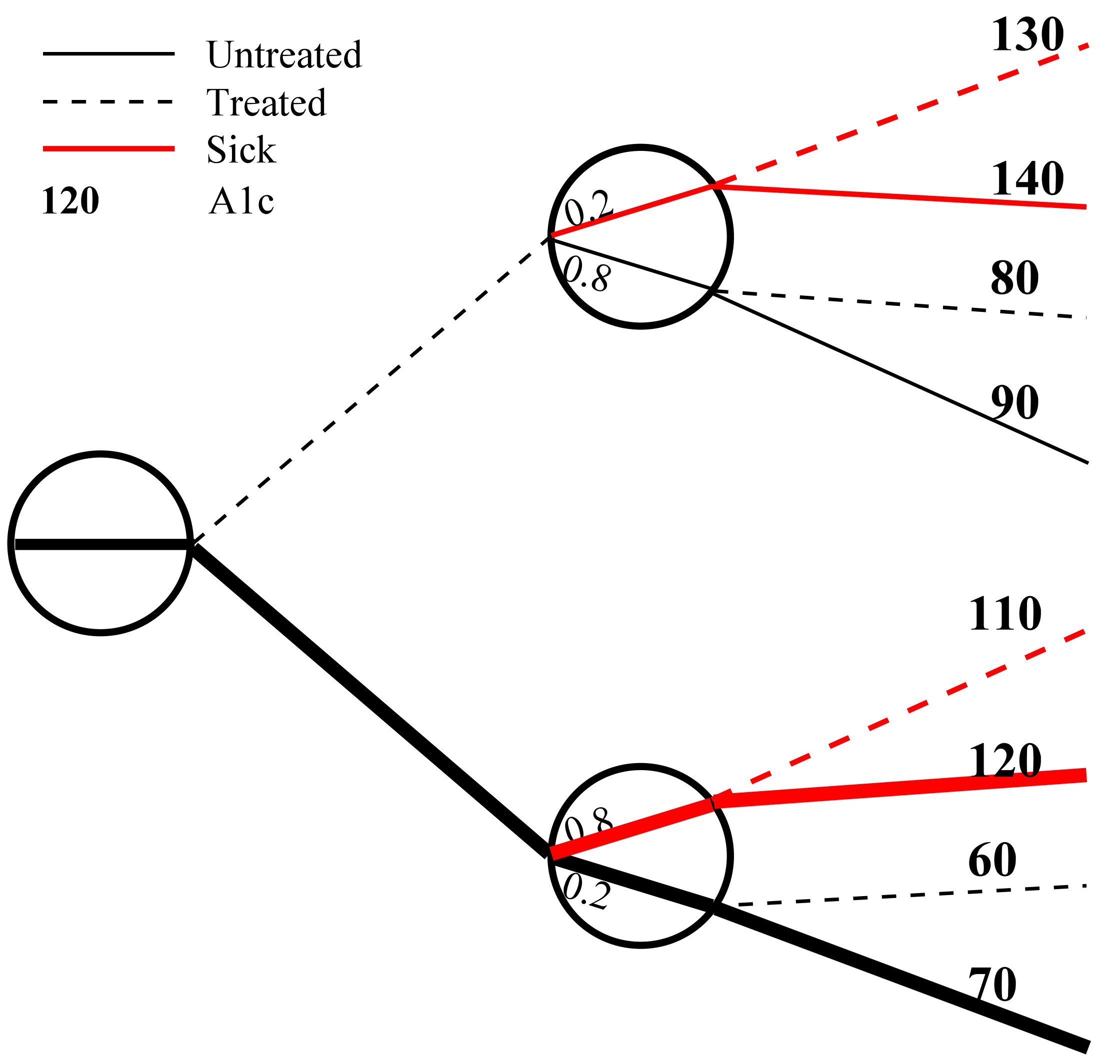
Observed data



compute probabilities from graph



compute what would happen if everyone not treated:



highlight paths consistent with regime "never treat"

remove probabilities of treatment from graph

multiply probabilities, expectations, together

sum over paths

$$E(Y^{00}) = \sum_{l_1} pr(L_1^{00} = l_1) E(Y^{00} | L_1^{00} = l_1)$$

$$= \sum_{l_1} pr(L_1^{00} = l_1 | A_0 = 0) E(Y^{00} | L_1^{00} = l_1 | A_0 = 0, A_1 = 0)$$

$$= \sum_{l_1} pr(L_1 = l_1 | A_0 = 0) E(Y | L_1 = l_1 | A_0 = 0, A_1 = 0)$$

so here, $E(Y^{00}) = 0.8 * 120 + 0.2 * 70 = 110$

similarly,

$$E(Y^{11}) = 0.2 * 130 + 0.8 * 80 = \ 90$$
$$E(Y^{10}) = 0.2 * 140 + 0.8 * 90 = 100$$
$$E(Y^{01}) = 0.8 * 110 + 0.2 * 60 = 100$$

simple example of G-computation algorithm (Robins, 1986) general form:

$$E(Y^G) = \sum_{\bar{l}} E(Y^G | \bar{L}_M^G = \bar{l}) \prod_{m=1}^{M} pr(L_m^G = l_m | \bar{L}_{m-1}^G)$$

$$= \sum_{\bar{l}} E(Y | \bar{L}_M = \bar{l}_m, \bar{A}_M = \bar{a}_M^G) \prod_{m=1}^{M} pr(L_m = l_m | \bar{L}_{m-1}, \bar{A}_{m-1} = \bar{a}_{m-1}^G)$$

where a_m^G indicates that treatment history through m is consistent with regime G

generalizes standardization for scalar treatment: there, no product over times

what if $\bar{L}_M$, $\bar{A}_M$ high dimension? (Inevitable if m moderate, etc.)

need to model $E(Y|\cdot)$

no need to model $pr(L_m = l_m|\cdot)$; can use empirical distribution function

how can one compute variance?

Delta method (very tedious for high dimensional problems)

bootstrap

Bayesian

model specification:

$$E(Y^G) = \sum_{\bar{l}} E(Y|\bar{L}_M = \bar{l}_m, \bar{A}_M = \bar{a}_M^G) \prod_{m=1}^{M} pr(L_m = l_m | \bar{L}_{m-1}, \bar{A}_{m-1} = \bar{a}_{m-1}^G)$$

can one see from formula the effect of a regime on outcome?

how can one choose parameters in these models to represent hypothesis of no effect?

Need to synthesize information from both parts of likelihood ($E(Y|\cdot)$, $pr(L_m|\cdot)$)

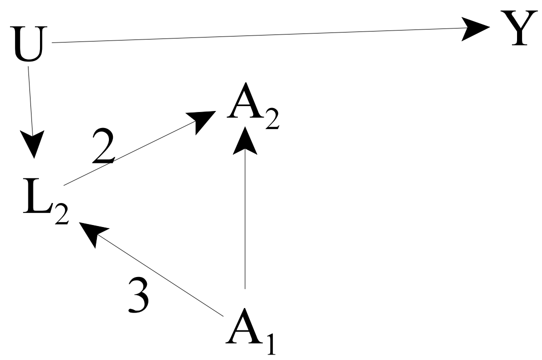
Difficult to see directly from parameters in each model what overall treatment effect is

In general, no compact way of representing the null hypothesis

In particular, if have confounding by variable affected by treatment (see DAG), will have $E(Y|\overline{A}_M, \overline{L}_m) \neq E(Y|\overline{L}_m)$ even if no effect of treatment on outcome

How can this happen?

Consider ineffective method for screening for cancer



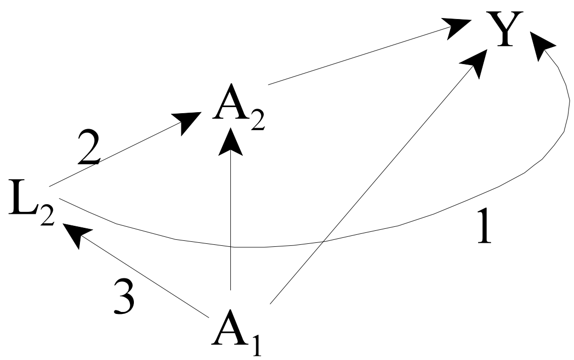
A	screening
L	diagnosis
U	underlying state of tumor after time 1

screening advances diagnosis, which is related to but does not affect outcome

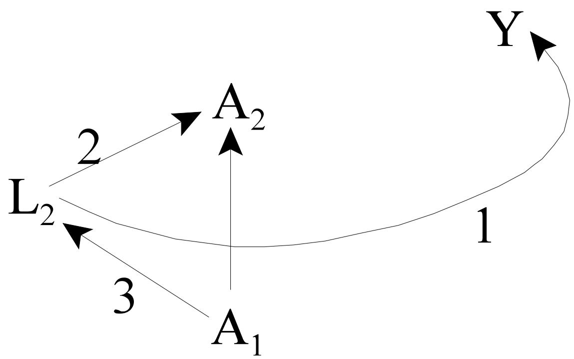
Result: when synthesizing information, model for *Y* inevitably misspecified

Can lead to invalid tests of null

How can one have A have effect on Y but regression of Y on A, L, Y show no dependence on A? (Consider DAG)



Allow only delayed effects relayed through intermediate L_2



note that $\{A_1, A_2\} \perp Y | L_2$, but A_1 has effect on Y

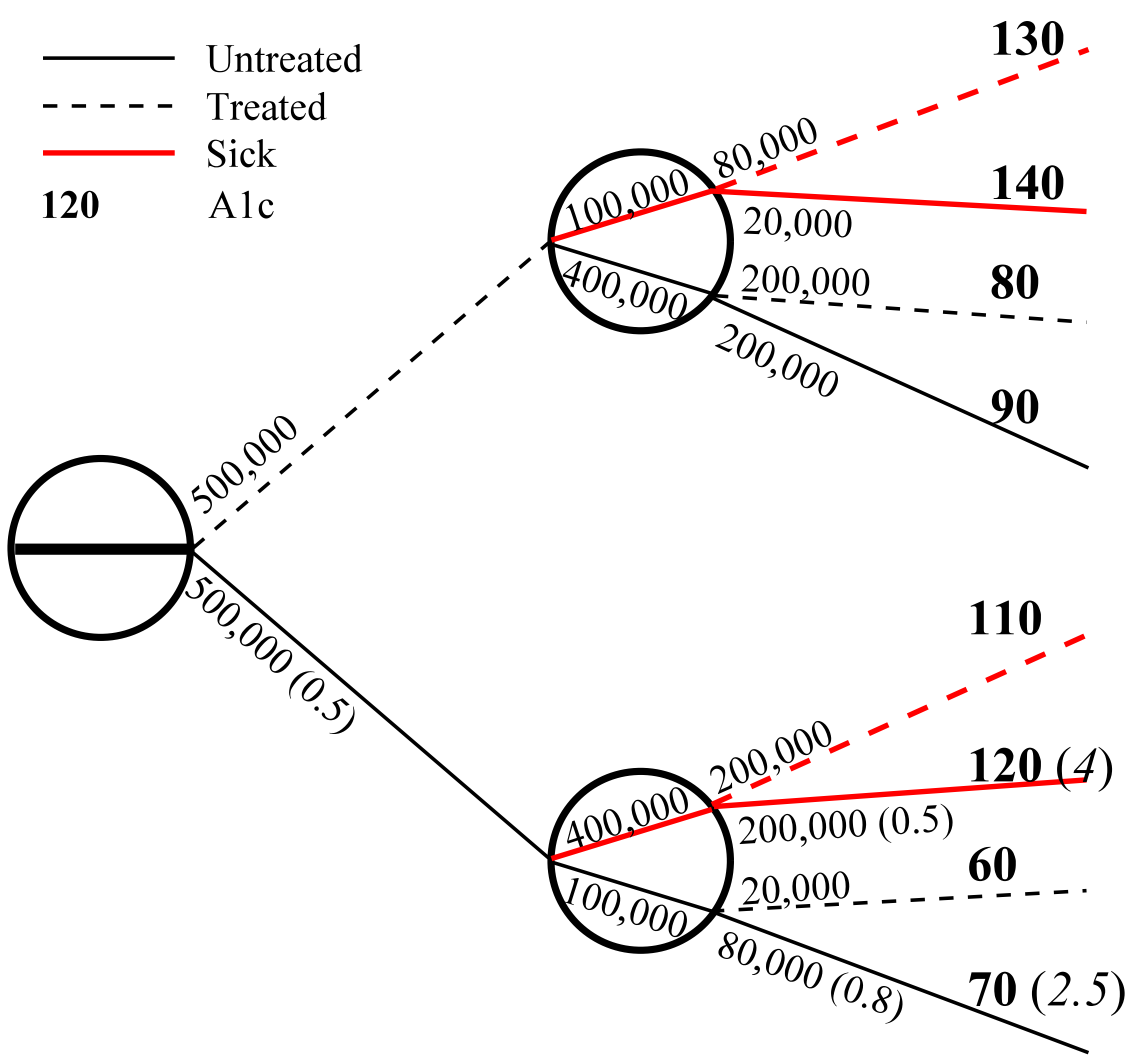
Modeling complicated, difficult to interpret

How would one specify probability model for the whole data?

How does this relate to inference for causal effects?

Consider two methods which directly model effect of treatment on outcome

Inverse probability of treatment weighting



to estimate $E(Y^{00})$, weight each subject by inverse probability of receiving treatment given previous history

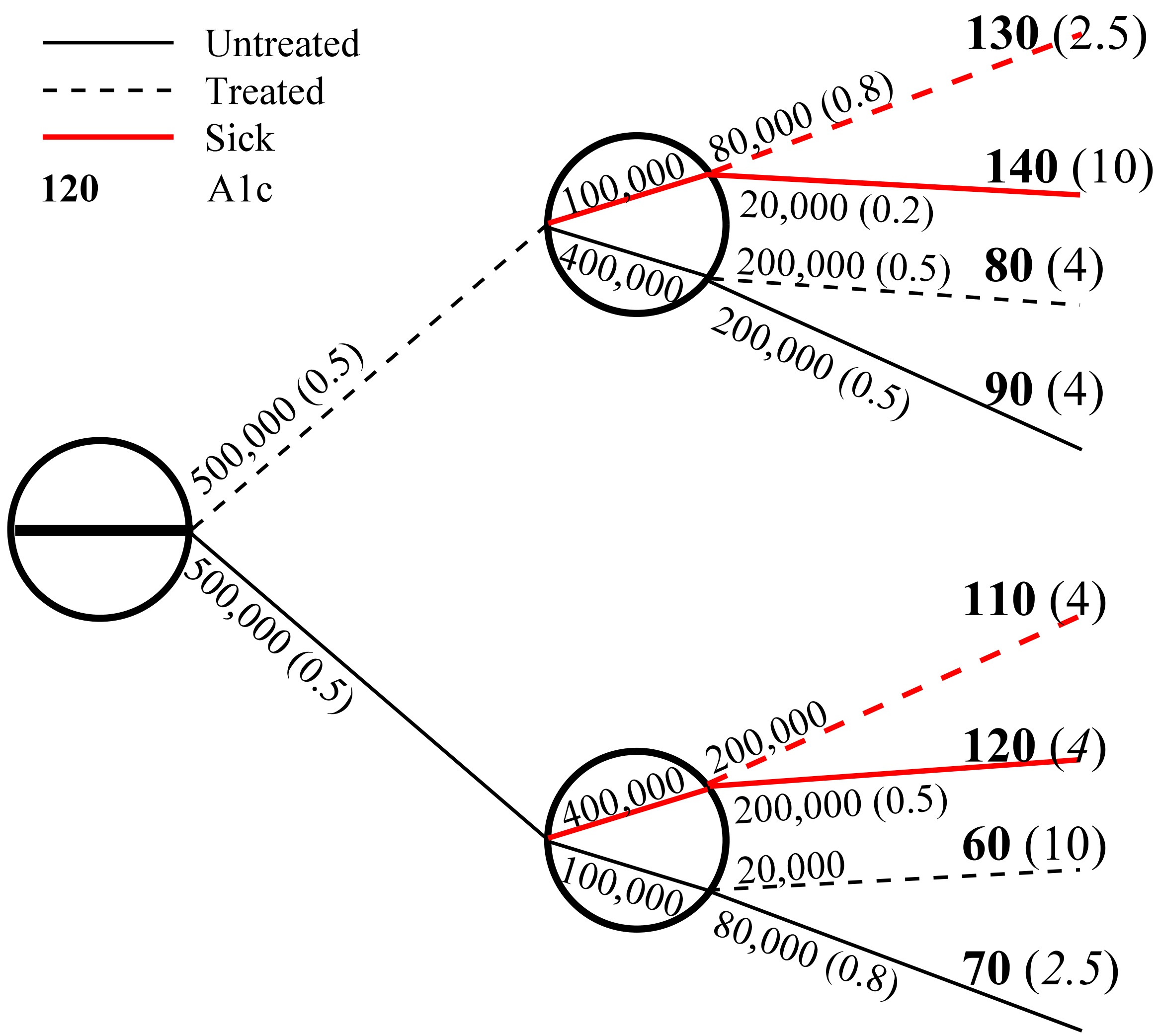
at time 1, each untreated subject stands in for 1/0.5 = 2 subjects (self and 1 comparable subject who is treated)

at time 2, untreated sick subject stands in for 1/0.5=2 subjects untreated healthy subject stands in for 1/0.8=1.25 subjects compute weighted estimate:

$$E(Y^{00}) = \frac{200{,}000*4*120+80{,}000*2.5*70}{200{,}000*4+80{,}000*2.5} = \frac{800{,}000*120+200{,}000*70}{1{,}000{,}000} = 110$$

total sample size in pseudopopulation is same as total sample size; represents what would have happened if entire population untreated

can compute weights for other regimes:



Untreated
Treated
Sick
120 A1c
500,000 (0.5)
500,000 (0.5)
100,000
400,000
80,000 (0.8)
20,000 (0.2)
200,000 (0.5)
200,000 (0.5)
130 (2.5)
140 (10)
80 (4)
90 (4)
400,000
100,000
200,000
200,000 (0.5)
20,000
80,000 (0.8)
110 (4)
120 (4)
60 (10)
70 (2.5)

describe randomized trial population created by weighting randomized trial with 1,000,000 subjects randomized to each of 4 arms/regimes: (0,0), (0,1), (1,0), (1,1)

model for $E(Y^{a_1,a_2})$:

$$E(Y^{a_1,a_2}) = 110 - 10(a_1 + a_2)$$

example of a marginal structural model

more generally; weights: $\dfrac{1}{\prod_{m=1}^{M} pr(A_m | \overline{L}_m, \overline{A}_{m-1})}$

applicable to nonparametric identification and identification of MSMs

general specification of MSMs

$$g\{E(Y^{\bar{a}_M} | V)\} = f(\bar{a}_M, V)$$

functions of treatment history, selected baseline covariates

what functions of treatment history might be used?

Cumulative treatment/ exposure $\sum_m a_m$

Current treatment a_M

ever treatment: $I(\sum_m a_m > 0)$

lagged versions of above

multivariable versions of above

interactions of treatment with baseline covariates V

stabilized weights: $\prod_{m=1}^{M} \frac{pr(A_m | V, \bar{A}_{m-1})}{pr(A_m | \bar{L}_m, \bar{A}_{m-1})}$

can only include in stabilizing numerator those covariates included in structural model

MSMs (as formulated here) do not model effects of dynamic regimes or interactions of treatment with post-baseline covariates

Structural nested models/G-estimation:

will consider rank-preserving models (deterministic causal effects)

propose model relating potential outcomes to each other:

$$Y^{a_1,a_2} = Y^{00} + a_1\Psi_1 + a_2\Psi_2$$

or

$$Y^{A_1,A_2} = Y^{00} + A_1\Psi_1 + A_2\Psi_2$$ (More circumspect model)

can assume
common effect of treatments received at different times (i.e., $\Psi_1 = \Psi_2$)

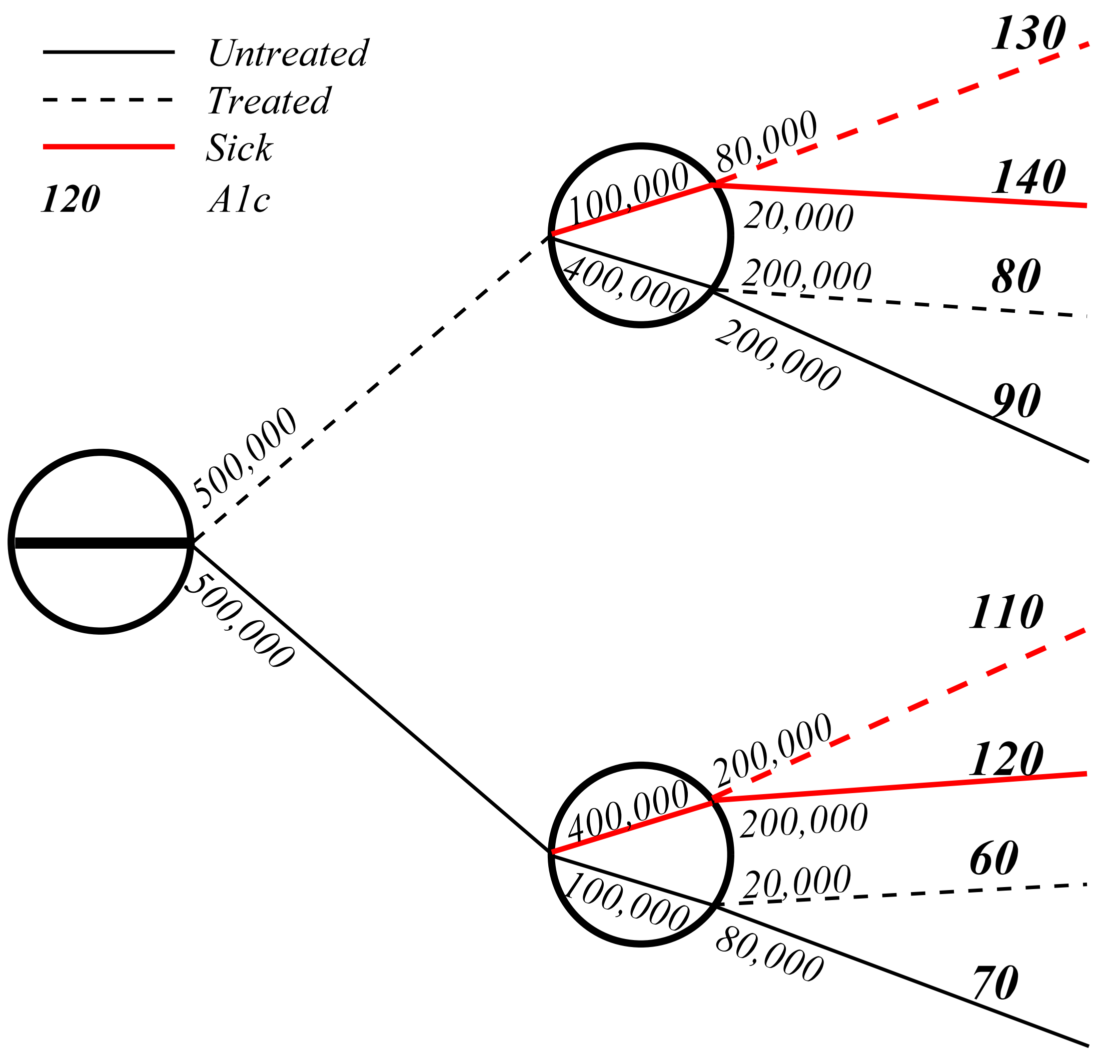
interaction of treatment with baseline on time-varying covariates (e.g.,
$Y^{A_1,A_2} = Y^{00} + A_1\Psi_1 + A_2 L_2 \Psi_2$)

etc.

to estimate parameters, compute putative potential outcome
$Y^{0,0}(\Psi) = Y - (A_1 + A_2)\Psi$ (Constant treatment effect model)

under ignorability, null hypothesis, $Y^{00}(\Psi) \perp A_m | \overline{A}_{m-1}, \overline{L}_m$

test as before with estimating equations or logistic regression



for $\Psi = 0$, test independence of A_2, $Y^{00}(\Psi)$ (Compare 130 vs, 140, etc.)

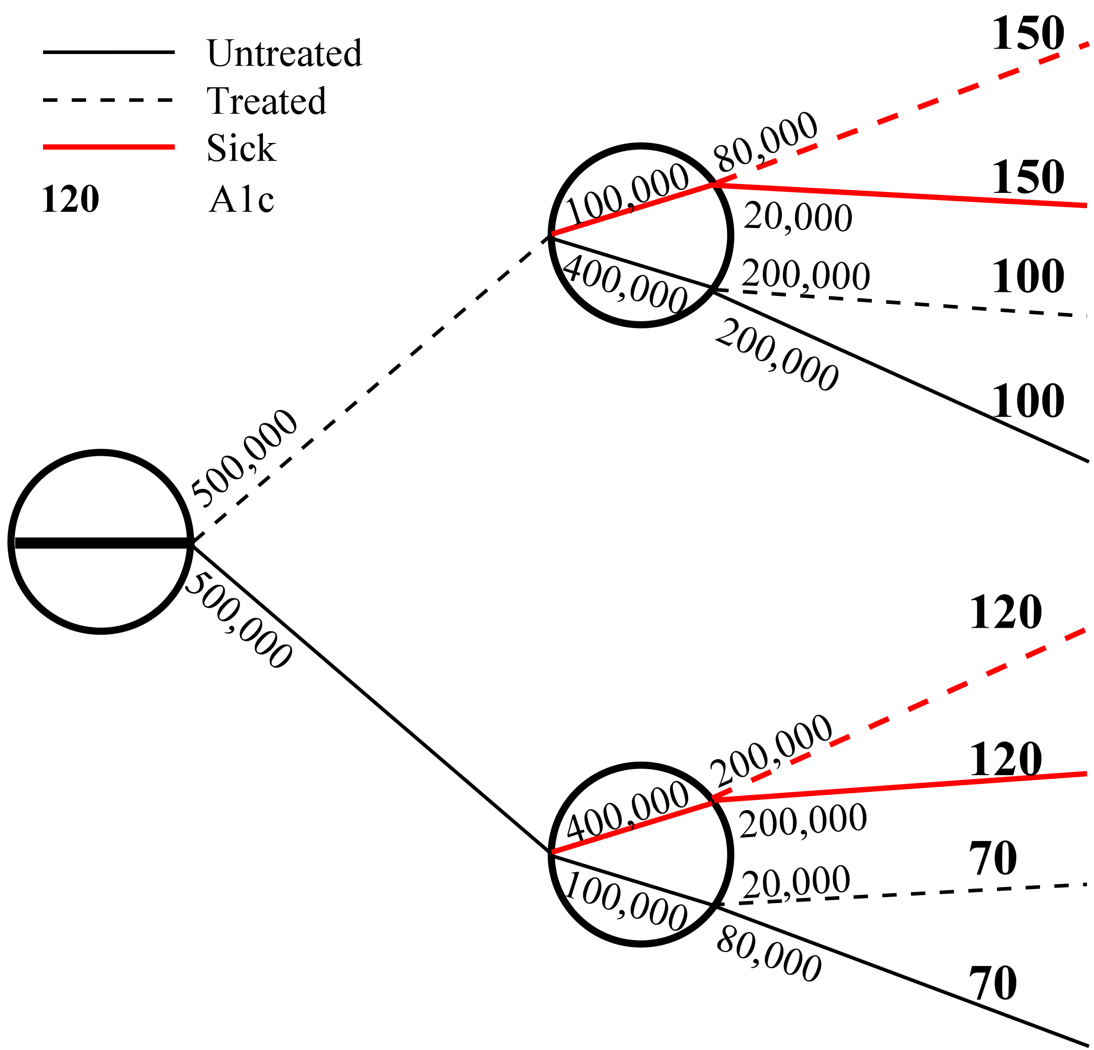
also test independence of A_1, $Y^{00}(\Psi)$:

(130*8+140*2+80*20+90*20)/50 = 94.4

vs.

(110*20+120*20+60*2+70*8) /50= 105.6

(Reject)



for $\Psi = -10$, test independence of A_1, $Y^{00}(\Psi)$

(150*8+150*2+100*20 +100*20)/50 =110

vs.

(120*20+120*20+70*2 +70*8) /50= 110

(Fail to reject)

What can one do if believe $A_1 \perp Y^{00}$ but don't believe $A_2 \perp Y^{00} | A_1, L_2$?

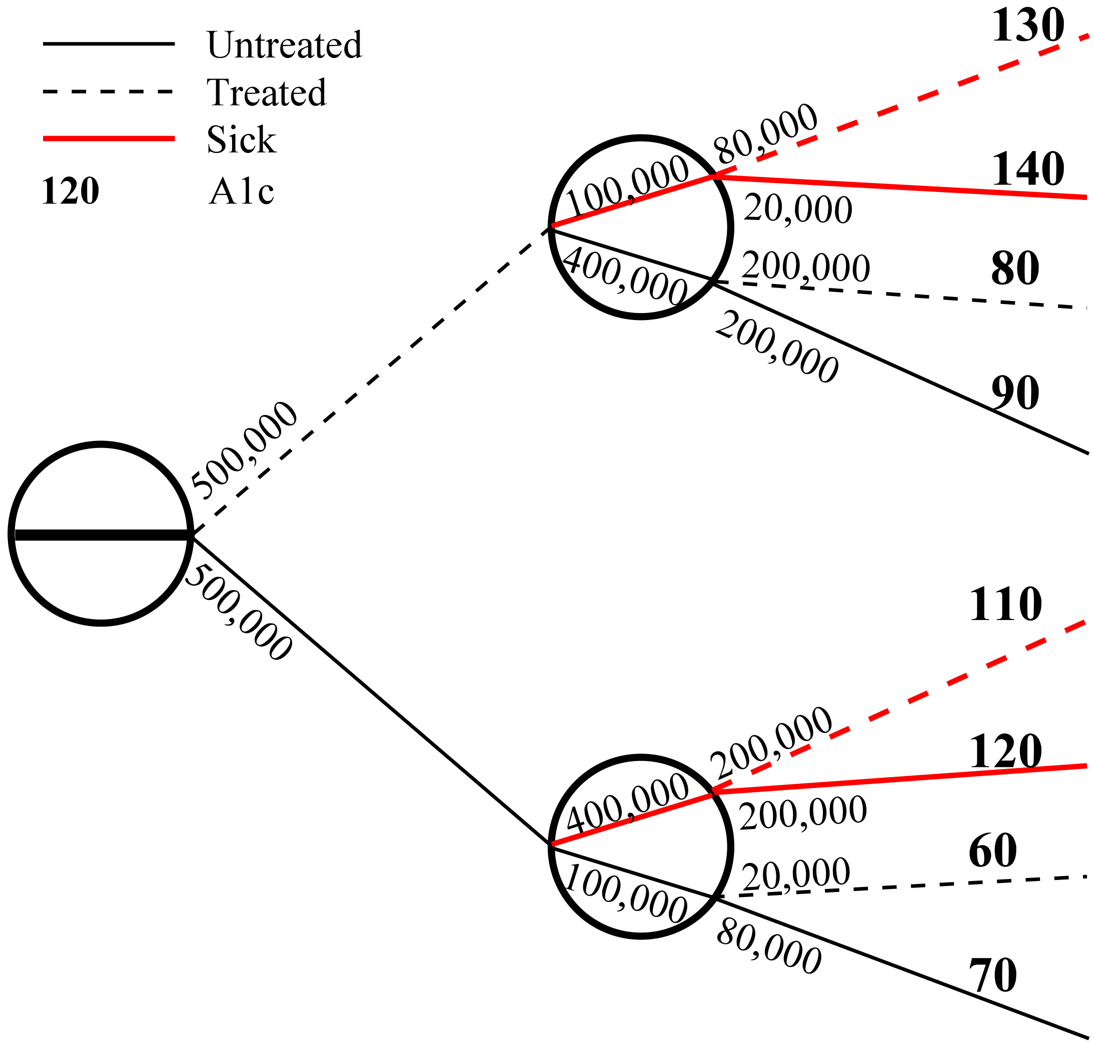
Test only first independence, not second

actually model for randomized trial with noncompliance

e.g., A_1 can represent randomization, first increment of treatment

A_2 subsequent level of treatment

standard approaches



$$E(Y|A_1) = 105.6 - 11.2A_1$$

attenuated as measure of overall effect

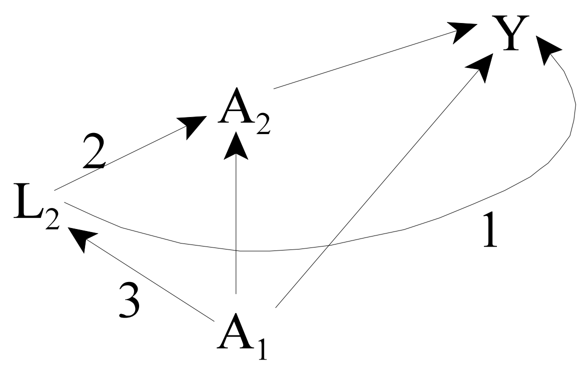
$$E(Y|A_2) = 100.8 - 1.6A_2$$
$$E(Y|A_1, A_2) = 105.71429 - 11.16883A_1 - 0.25974A_2$$
$$E(Y|A_1 + A_2) = 105.71429 - 5.71429(A_1 + A_2)$$
$$E(Y|A_1, L_2) = 68 + 17A_1 + 47L_2$$
$$E(Y|A_2, L_2) = 84.28571 - 6.10390A_2 + 37.53247L_2$$
$$E(Y|A_1, A_2, L_2) = 70 + 20A_1 - 10A_2 + 50L_2$$
$$E(Y|A_1 + A_2, L_2) = 78.59873 + 2.42038(A_1 + A_2) + 37.96178L_2$$

The only correct estimate is effect of A_2 when controlling for A_1, L_2

expected under sequentially ignorable treatment assignment
can estimate component effect of last bit of treatment

cannot estimate joint effects of treatments received at different times

DAG



conditions 1+2: confounding of effect of A_2

conditions 1+3: L_2 intermediate in path from A_1 to outcome

if interested in joint effects, in presence of conditions 1-3, no standard regression/stratification approach will be valid

direction of bias depends on analytic method, nature of associations/effects problems can be present if L_2 not intermediate but is affected by A_1, associated with A_2

SNMs can be used for modeling effects of dynamic treatment regimes

need first to specify likelihood for potential outcome, treatment, covariate processes:

$$pr(\underline{Y})\prod_{m=1}^{M}\{pr(L_m|\overline{L}_{m_1},\overline{A}_{m-1},\underline{Y})pr(A_m|\overline{L}_m,\overline{A}_{m-1})\}$$

what assumption allows writing likelihood this way?

Can also estimate parameters by using ML; why not do this?

joint distribution under regime can be derived from modified likelihood:

$$pr(Y^0)\prod_{m=1}^{M}\{pr(L_m|\overline{L}_{m-1},\overline{A}_{m-1},Y^0)\} \quad \prod g_m(\overline{L}_m)=A_m$$

algorithm:

1. Estimate causal parameters Ψ using G-estimation
2. draw an estimate of Y^0 ($Y^0(\hat{\Psi})$) from its distribution
3. Draw L_m from its conditional distribution (for m=1)
4. Assign A_m based on regime

   repeat steps 3 and 4 for m = 1 to M

5. derive Y from Y^0, $\overline{A}_m$,$\overline{L}_m$,Ψ

algorithm can be substantially simplified (eliminate recursive steps: 3 and 4) if

1. Treatment regime not dynamic

and

2. Effect of treatment does not depend on post-baseline covariates

estimating equation approach to G-estimation

$$\sum_i \sum_k (A_{ik} - p_{ik}) g\{\overline{A}_{ik-1}, \overline{L}_k, Y^0(\Psi)\} = 0$$

analogous to previous case: anything in past at time of treatment decision can be put into $g(\cdot)$

optimal function $g(\cdot) = E(S_\Psi | A_k = 1, \overline{A}_{k-1}, \overline{L}_k, Y^0) - E(S_\Psi | A_k = 0, \overline{A}_{k-1}, \overline{L}_k, Y^0)$

optimal dynamic treatment regime

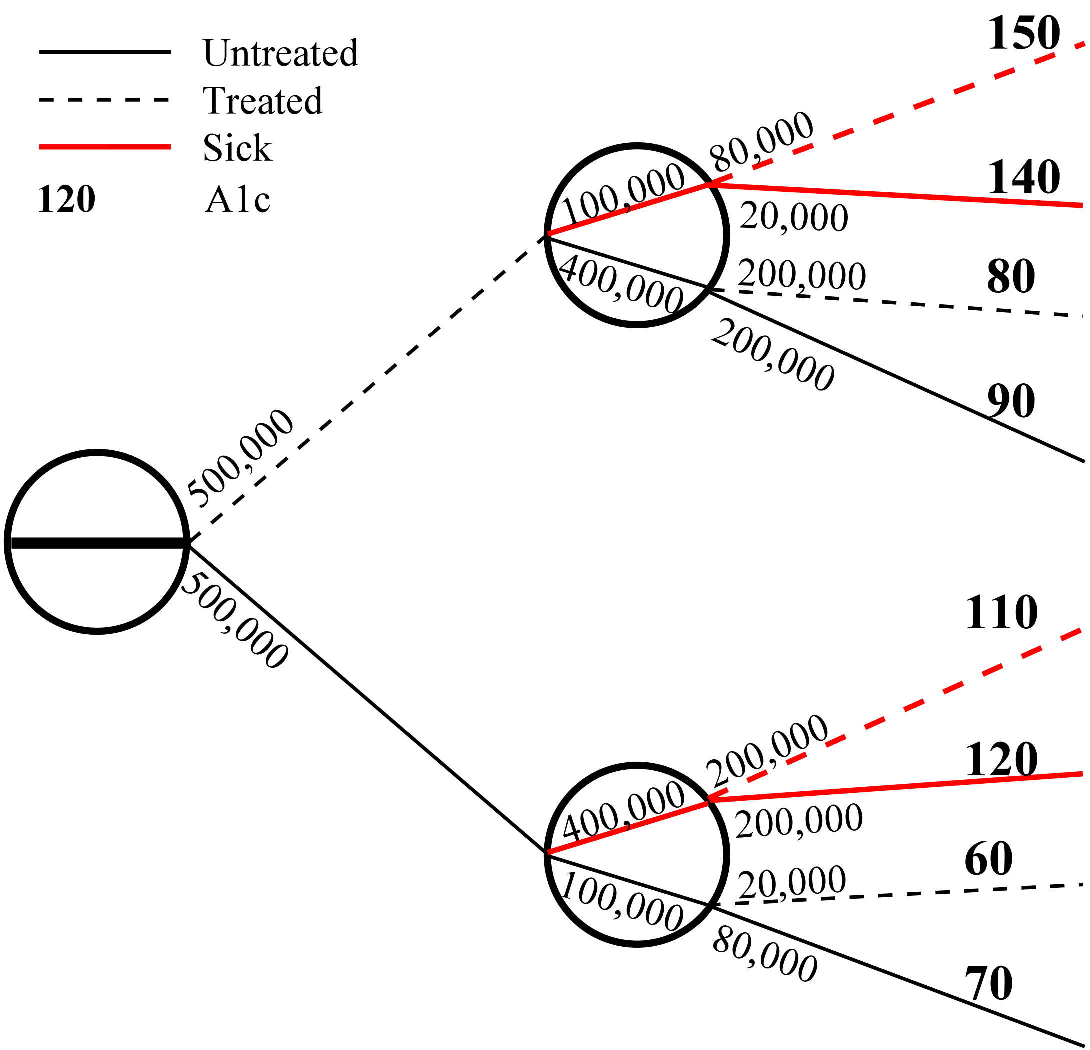
want to choose treatment sequence which will maximize benefit:

choose G: $\text{argmax}_{G'}\{E(Y^{G'})\}$

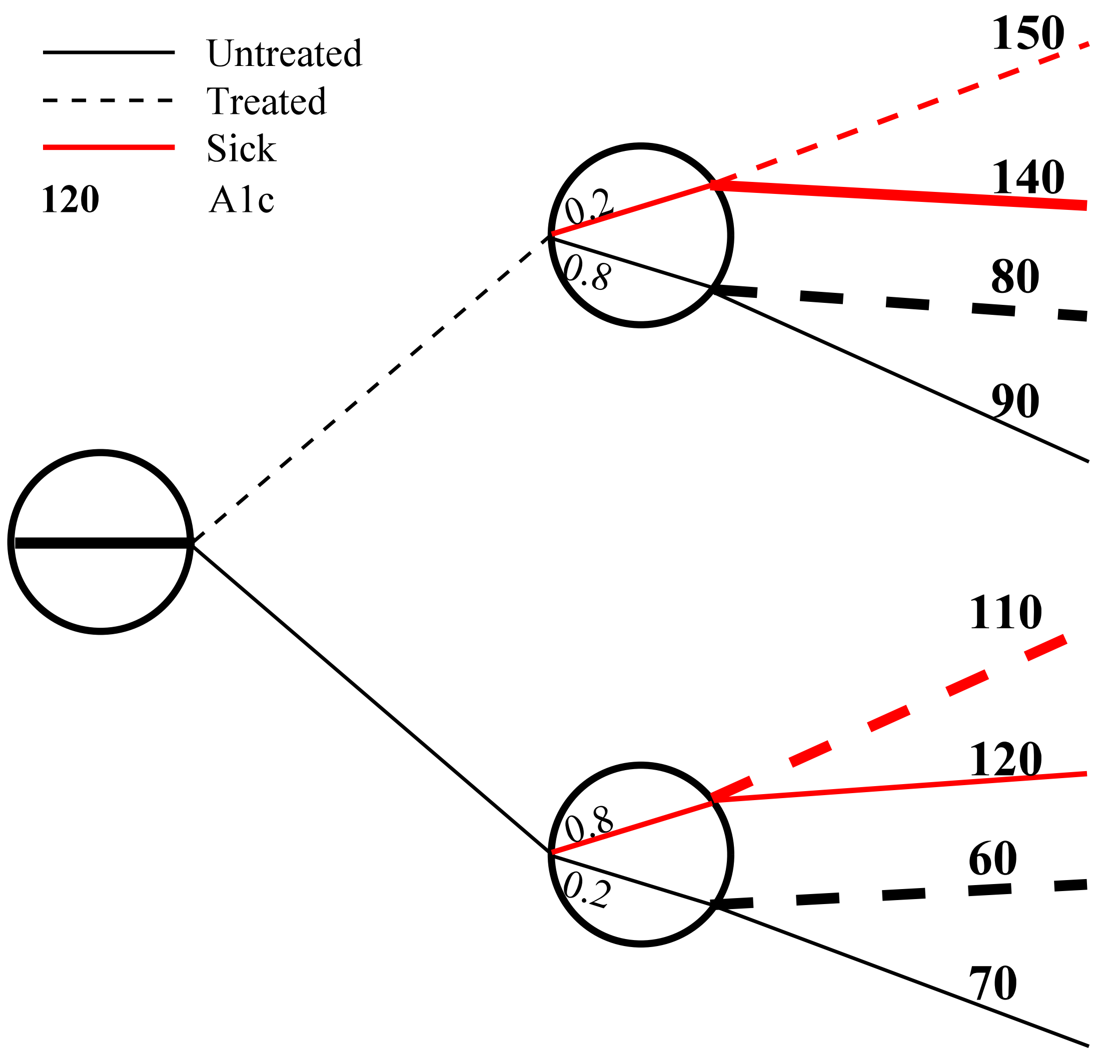
consider how to do nonparametrically in simple setting

will assume sequential ignorability:

in numerical example, will assume that sick subjects who have been treated previously develop tolerance to treatment; subsequent treatment harmful; otherwise treatment beneficial



choose optimal treatment at last stage as function of previous treatment, covariate history

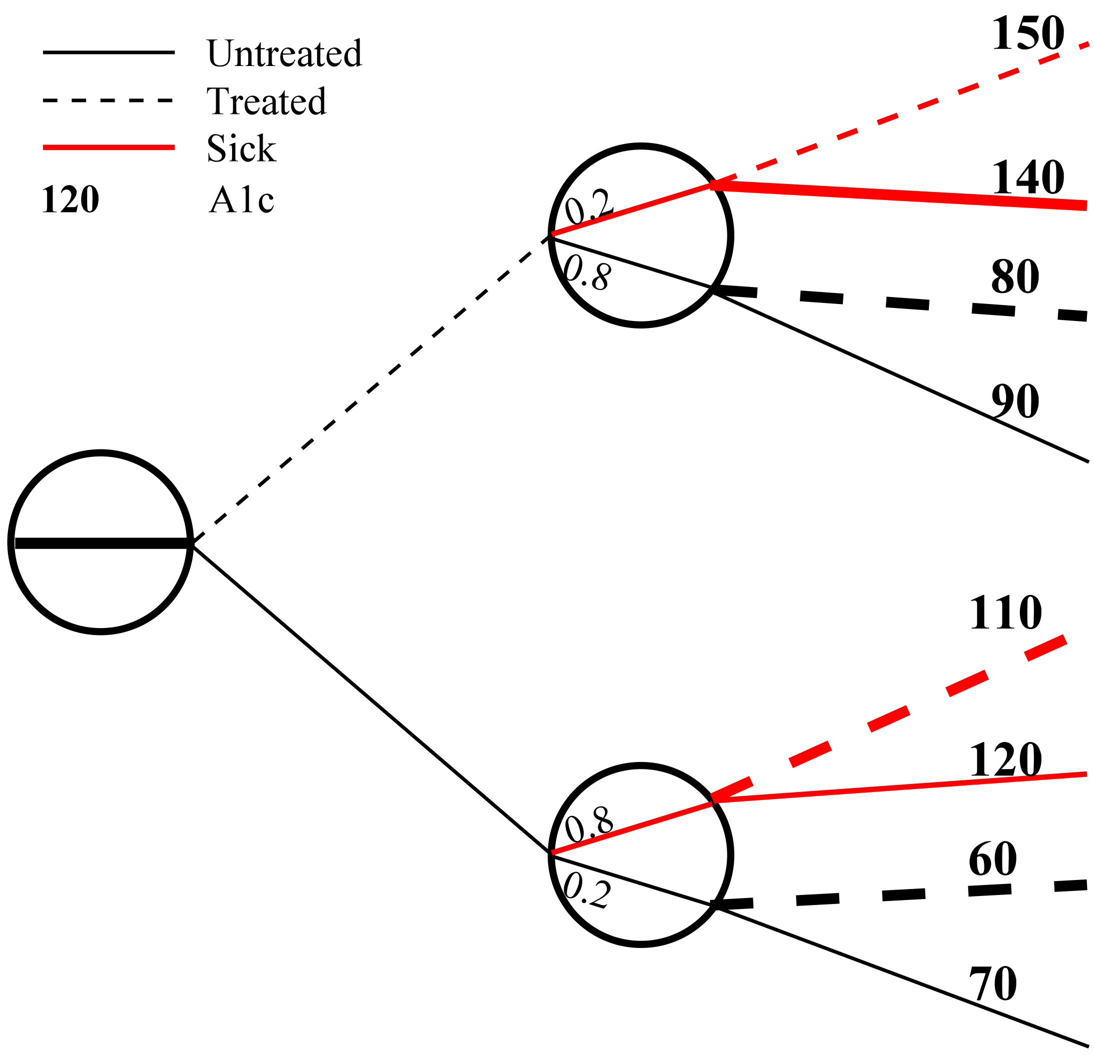


eliminate treatment probabilities from graph

highlight optimal treatment at last stage

replace #s by probabilities

how can one choose the optimal treatment at the beginning?



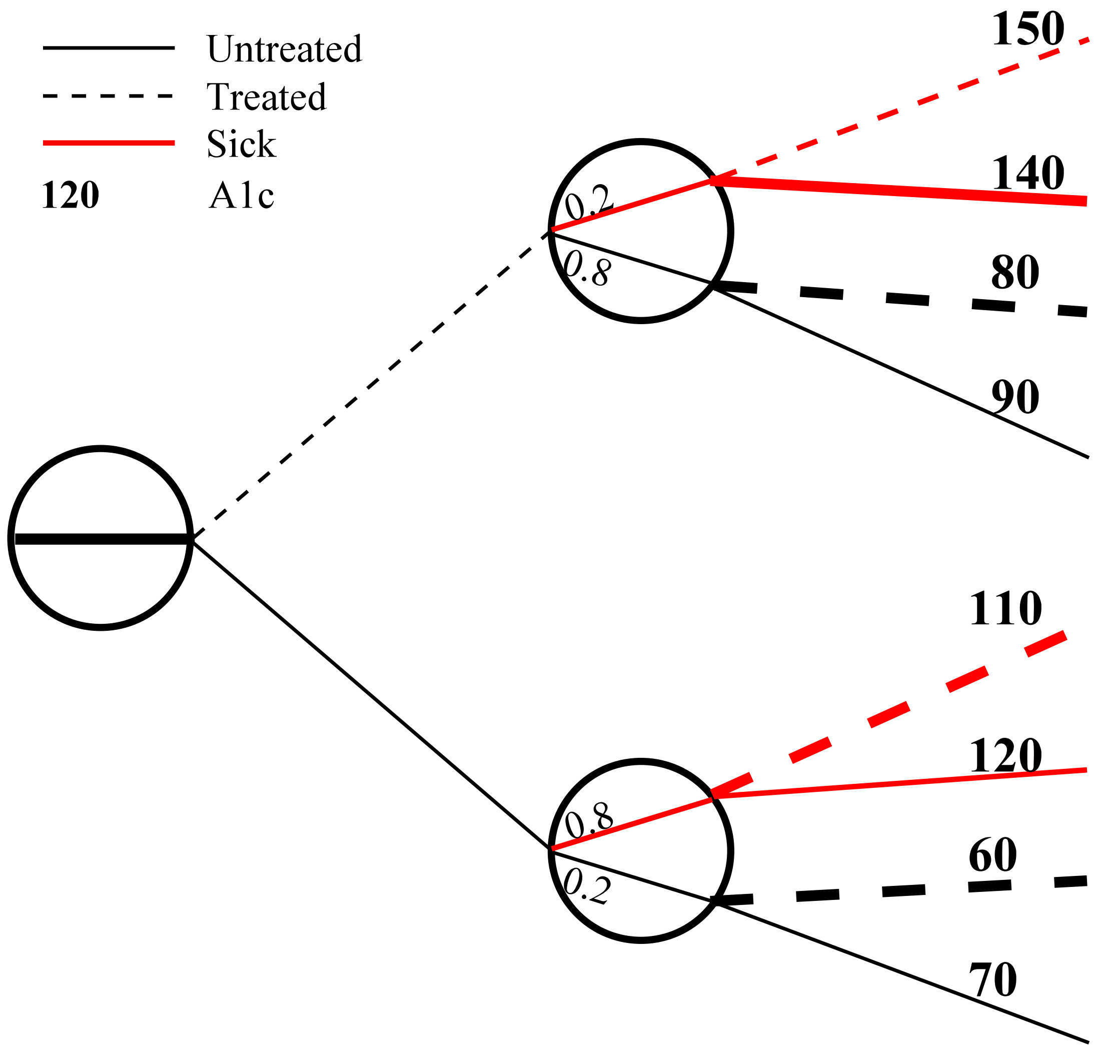
Choose the treatment at time 1 that, in conjunction with the optimal treatment at time 2, leads to the best overall outcome

use G-computation/extended standardization

don't treat at outset:
0.8(110)+0.2(60)=100

treat at outset:
0.2(140)+0.8(80)=92

state optimal regime



treat at beginning, then

treat at time 2 if still healthy

don't treat at time 2 if sick

what problems might result with estimation if many times, high dimensional covariates?

Sparse data: nonparametric estimation not possible

will need to resort to modeling

can use variant of structural nested models/G-estimation alternatively, can use variant of marginal structural models: history-adjusted MSMs

obtained by combining component MSMs from each time

let $V_k = q(X, \bar{L}_k, \bar{A}_{k-1})$ be some function of past treatment, covariate history

model distribution of potential outcomes given specified part of past V_k:

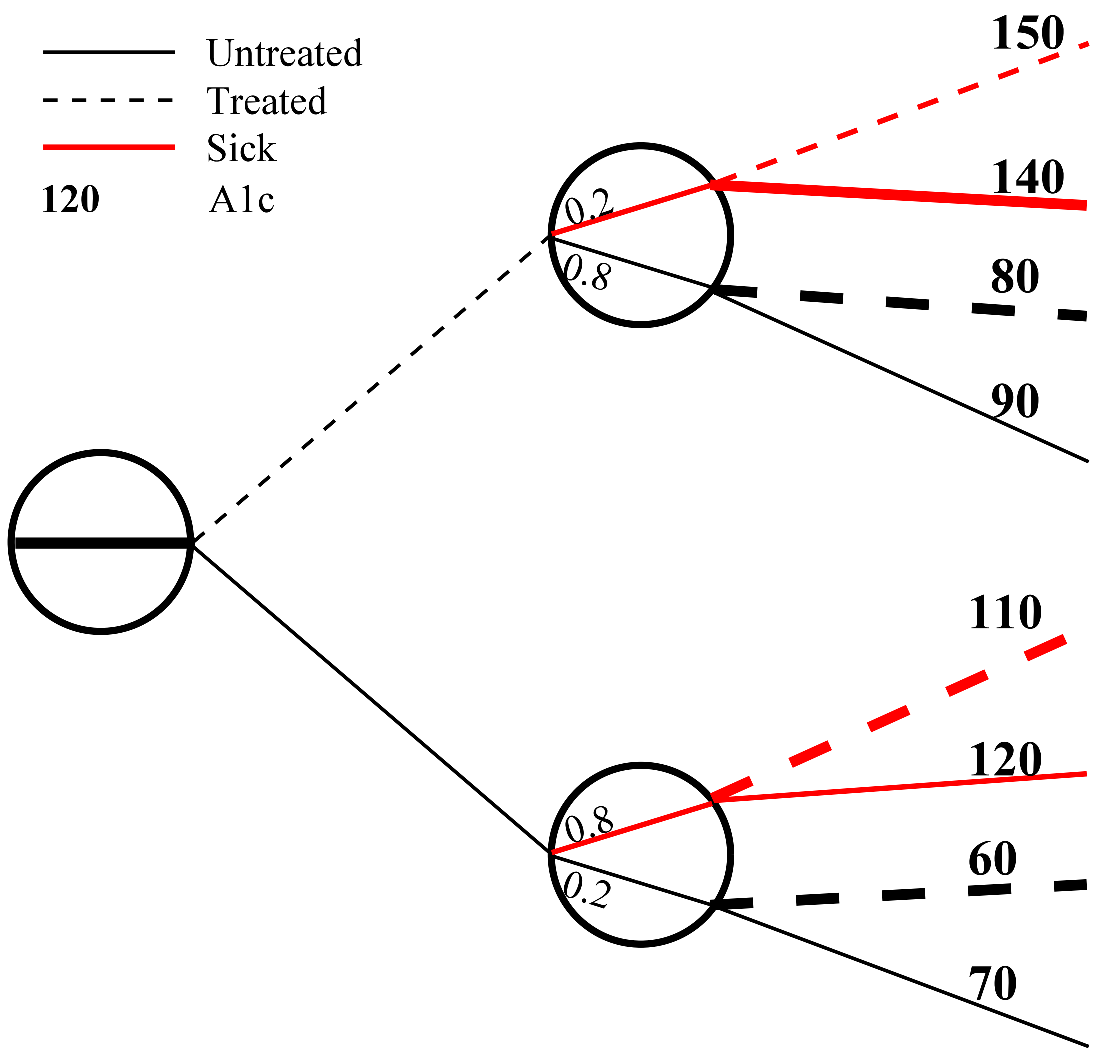
$Y^{\bar{A}_{k-1}\bar{a}_{k+}}$ outcome following observed exposure/treatment through k, $\bar{a}_{k+}$ beyond k

$$E(Y^{\bar{A}_{k-1}\bar{a}_{k+}} | V_k) = g(V_k, \bar{a}_{k+})$$

Statically optimal dynamic treatment regime:

at each point k, determine the optimal static regime from k onward

do the first part of that regime



four static regimes: expected outcome

(0,0) = 70*0.2+120*0.8=110

(0,1) = 60*0.2+110*0.8=100

(1,0) = 90*0.8+140*0.2=100

(1,1)=80*0.8+150*0.2= 94

so choose to treat at outset

optimal regime:

optimal statically optimal dynamic regime: same as optimal regime

not always case

censored data:

censoring by death

censored failure-time data:

analytic approaches

G-computation algorithm

inverse probability of (non-)censoring weighted estimation

G-estimation:

structural nested failure-time models

artificial censoring for estimation