Application of Causal Inference Methods to Improve Treatment of HIV in Resource-Limited Settings

Maya Petersen

works.bepress.com/maya_petersen

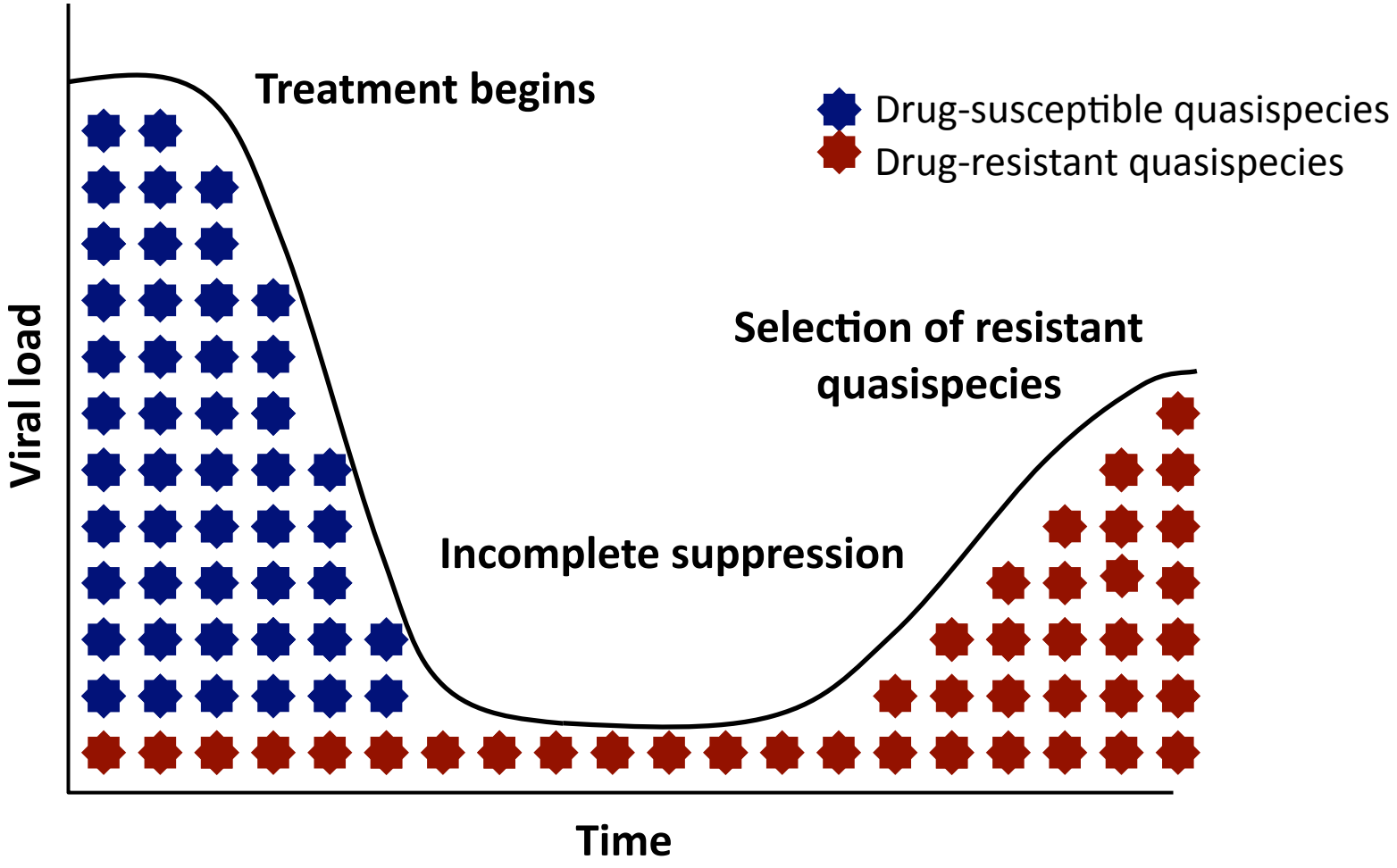
Divisions of Biostatistics and Epidemiology,
University of California, Berkeley

CAPS, October 19, 2010

Outline

1. What types of questions can causal inference methods help us to answer?
   - Example: Strategies for detecting and managing first line failure
2. When do standard analysis methods break down?
   - Time-dependent confounding of longitudinal treatments
3. What estimation tools are available?
   - Inverse Probability Weighting
   - Longitudinal G- computation
   - New methods coming…

HIV frequently develops resistance to antiretroviral drugs



Detecting and responding to virological failure

- Once virological failure develops, change regimen immediately to prevent additional resistance and disease progression

-> Standard of care: Regular viral load monitoring

  - Detect failure and initiate change in regimen

- Viral loads are expensive and require significant infrastructure

  - Burden of HIV greatest in resource-limited settings where viral loads often not available

When to switch? WHO 2010 Recommendations

Virological	Plasma viral load above 5000 copies/ml (confirmation recommended)
Immunological	• Fall of CD4 count to pre-therapy baseline (or below); or • 50% fall from the on-treatment peak value; or • Persistent CD4 levels below 100 cells/mm
Clinical	New or recurrent WHO Stage 4 condition (not IRIS; +/- some Stage 3)

Impact of Alternative WHO Failure Criteria?

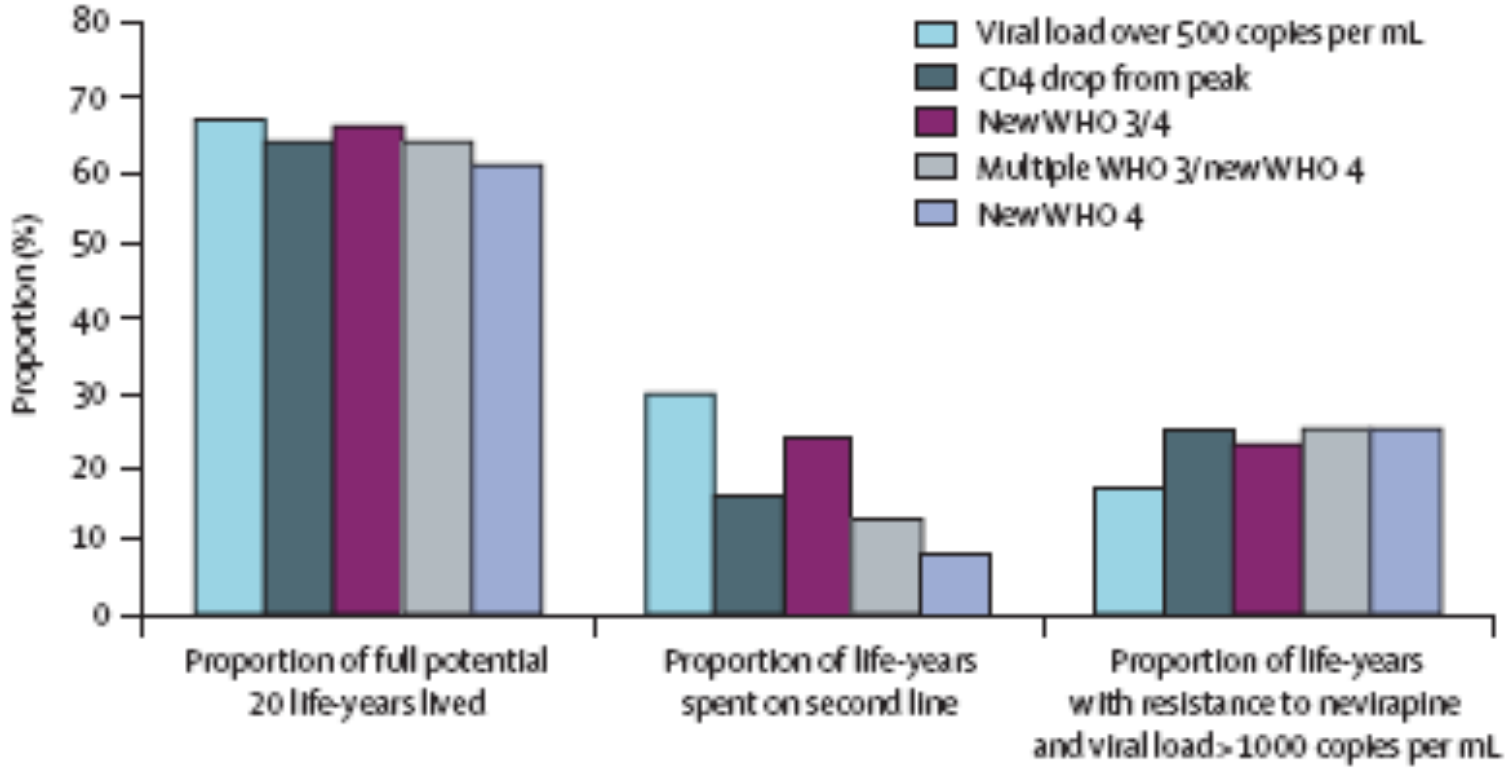
- Many studies: Immunologic criteria have poor sensitivity for detecting virological failure
  - Ex: Keiser (2009): Sensitivity 17% (7,32)
- ART-LINC (2009): Higher switch rates and switch at higher CD4 counts in programs that have VL monitoring
- Use of CD4 vs. VL-based criteria will result in delayed switch for many

Impact of Alternative WHO Failure Criteria?

- Systematic review
  - Chang et al (Cochrane 2010)
- HBAC (Tororo,Uganda): RCT comparing clinical/CD4/VL
  - Published in abstract only (CROI 2008)…
- Three ongoing RCTs…
  - Zambia (Saag): routine vs. targeted VL
  - Thailand (Lallemant): CD4 vs. VL
  - Cameroon (Laurent): CD4/VL vs. clinical
  - Multi-country (PENPACT1): different VL thresholds in kids

Mathematical model of alternative monitoring strategies in Africa

• Monitoring based on CD4 instead of viral load results in longer delay until treatment modification, but little long-term mortality difference



Phillips et. al. *Lancet* 2008, 20 year outcomes

Impact of alternative monitoring strategies

- Will delayed regimen modification after loss of viral suppression increase mortality? How much?
- Will use of CD4 (or clinical) rather than VL-based switching criteria increase mortality? How much?
- Which CD4 criteria will give the best patient outcomes?

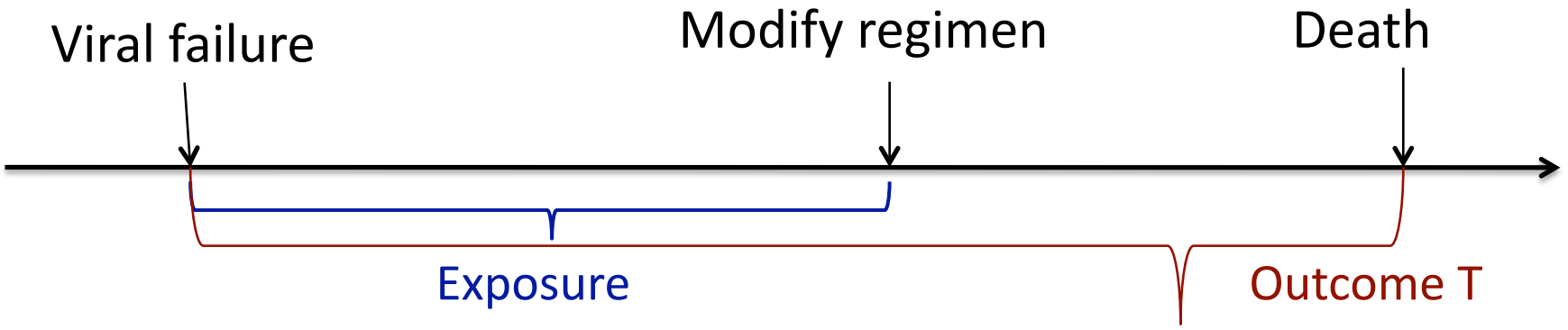
Hypothetical Randomized Trial

- Will delayed regimen modification following viral failure increase long-term mortality?

Subjects enrolled at time of viral failure

Subjects remain on failing therapy for a randomly assigned duration

Follow until death



Counterfactuals

- T_{switch} : counterfactual survival time under a specified switch time
  - Example: T_0 = an individual's survival if she were to have switched immediately
- Ideally, observe the survival time of everyone in the population under each possible delay time
  - Compare survival under different delay times
- Example: $E(T_{switch\ at\ 0\ mo} - T_{switch\ at\ 9\ mo})$
  - Difference in mean survival time if everyone in population had switched immediately versus after a 9 month delay

Marginal Structural Models

- Model on the counterfactual outcome
  - Ex. Model Discrete Hazard

$$P(T_{switch} = t | T_{switch} \geq t) = m(t, switch|\beta)$$

  - One possible model:

$$logit(m(t, switch|\beta)) =$$

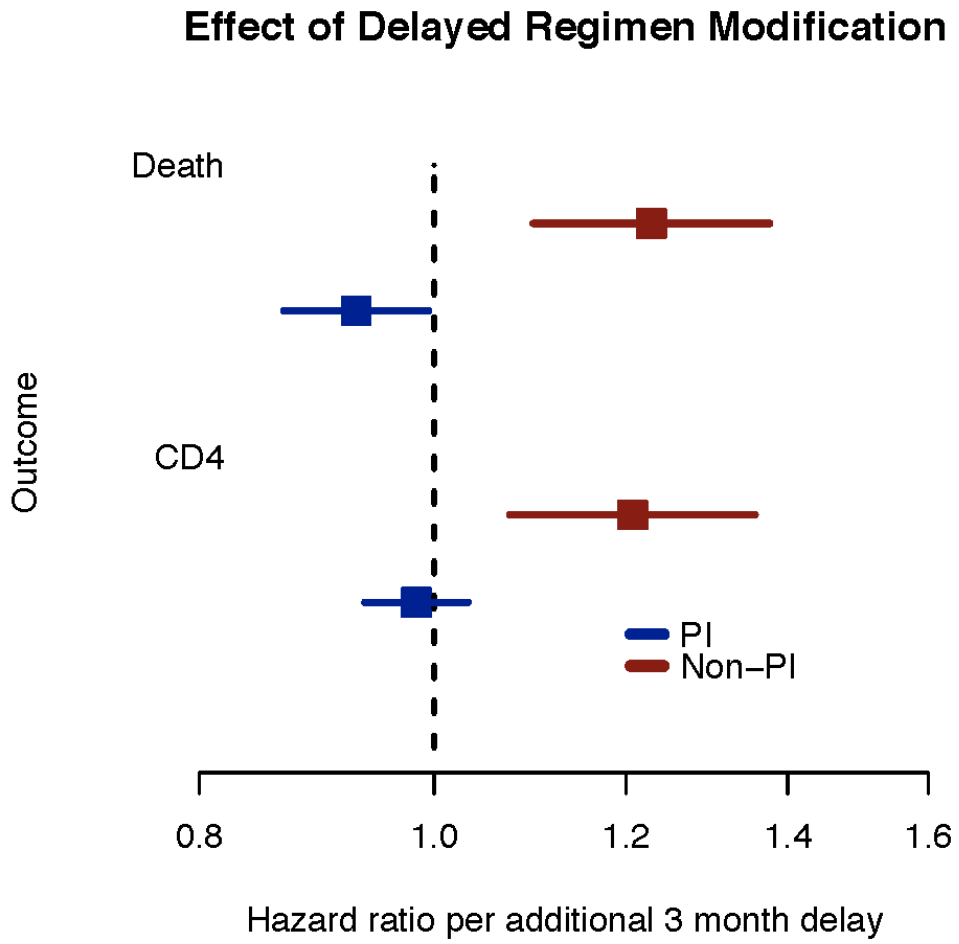
$$\beta_0 + \beta_1 \min(t, switch) + \beta_2 t + \beta_3 \min(t, switch) \times t$$

  - For a given time *t*, linear summary (on log odds scale) of the cumulative effect of time on failing therapy on the discrete hazard of death at *t*

Prospective clinical cohort data

- 2 cohorts of HIV-infected patients
  - Johns Hopkins ; Univ. North Carolina, Chapel Hill
- Eligible when experience virologic failure on HAART (*t=0*)
  - 982 subjects, 3414 person-years follow-up, 93 deaths
  - 742 failed a PI-based regimen; 240 fail a non-PI based regimen
- Primary outcome: All-cause mortality
  - Secondary outcome: immunological failure
- Exposure: Delay from virological failure to switch
- Covariates: Demographics, risk group, CD4/CD8, Viral Load, AIDS-related DX, ARV history

Delayed modification following failure of a HAART regimen without a protease inhibitor increases mortality



Petersen et al, AIDS 2008

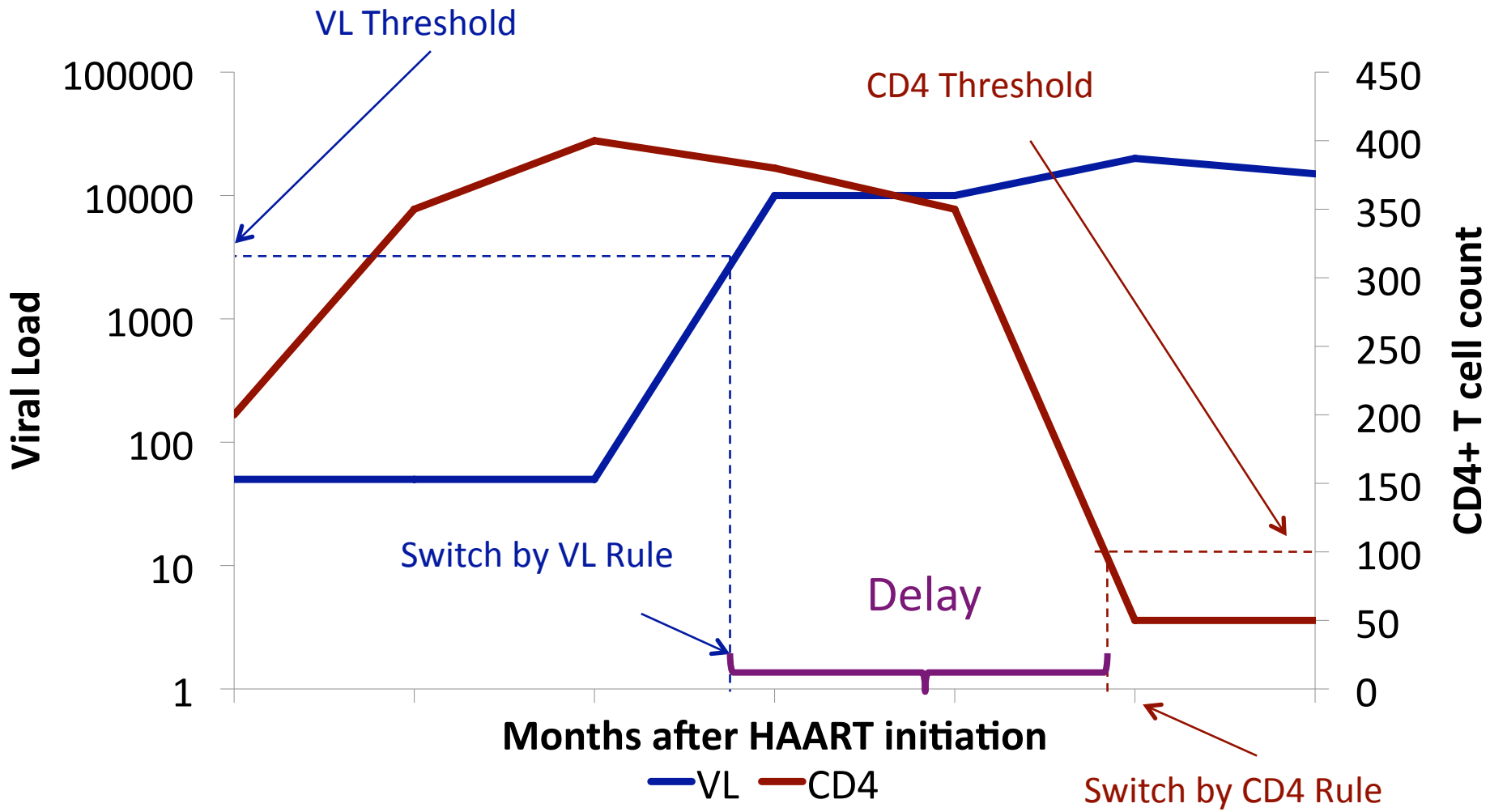
How much does this tell us about the importance of viral load monitoring in Africa?

- Subjects from North America
  - Different health care systems, patient populations….
- A historical cohort, not representative of current first line failures
  - Included failures between 1996 and 2006
  - Roughly 50% of sample exposed to ARVs prior to 1996
    - Not powered to exclude these individuals
- Does not directly address the question: What impact will lack of viral load monitoring have on mortality?

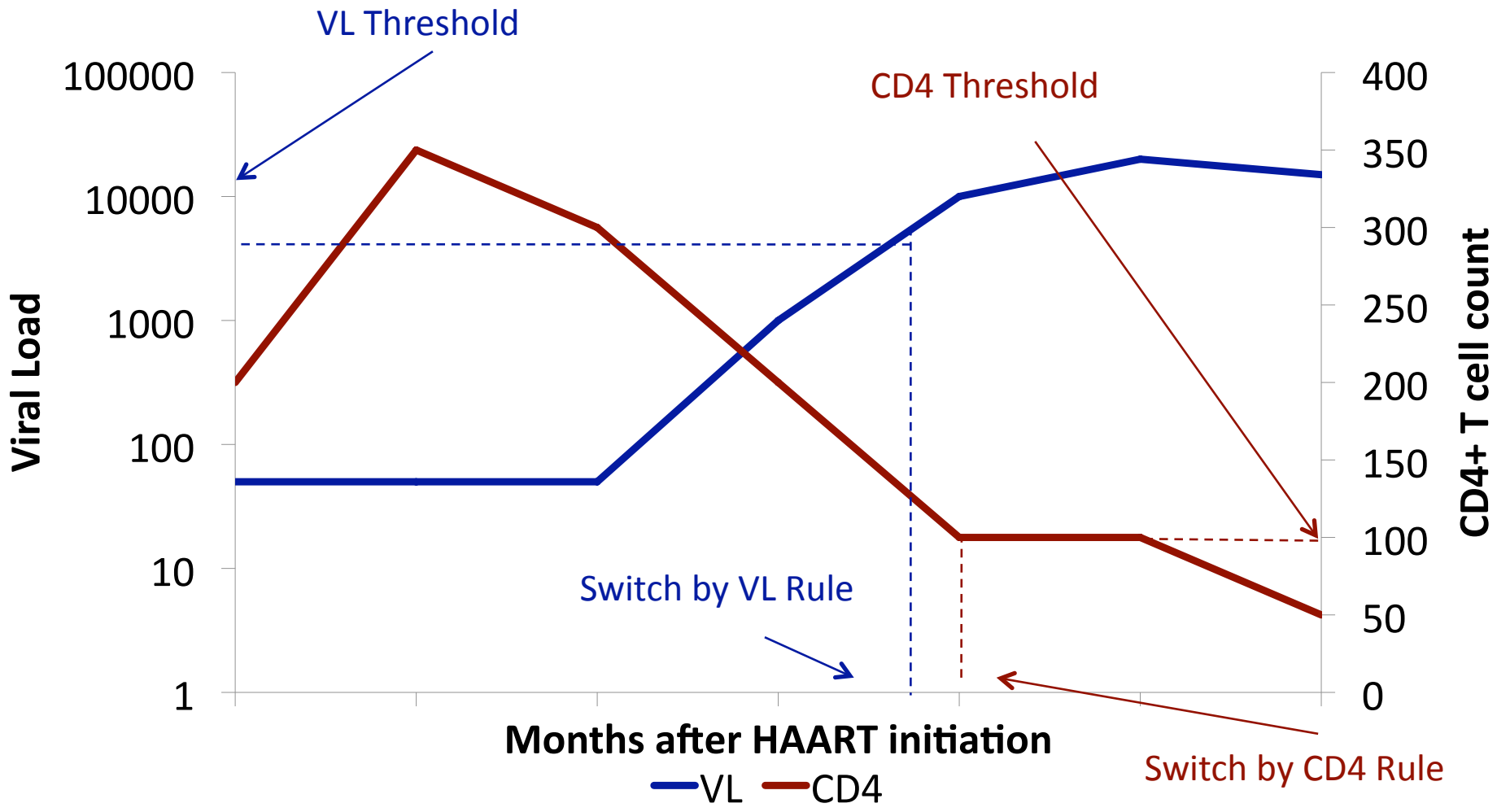
Refining the Question

- Say we find that delayed modification after viral failure does increase mortality
- Substantial data suggest that use of CD4 counts rather viral loads will increase the average delay time between viral failure and switch
- Does this imply that use of a CD4-based switching strategy will increase mortality?

Subject #1: Delayed switch after failure; CD4 elevated for much of the delay time



Subject #2: Minimally delayed switch after failure



Refining the Question

- Say we find that delayed modification after viral failure does increase mortality
- Substantial data suggest that use of CD4 counts rather viral loads will increase the average delay time between viral failure and switch
- Does this imply that use of a CD4-based switching strategy will increase mortality?

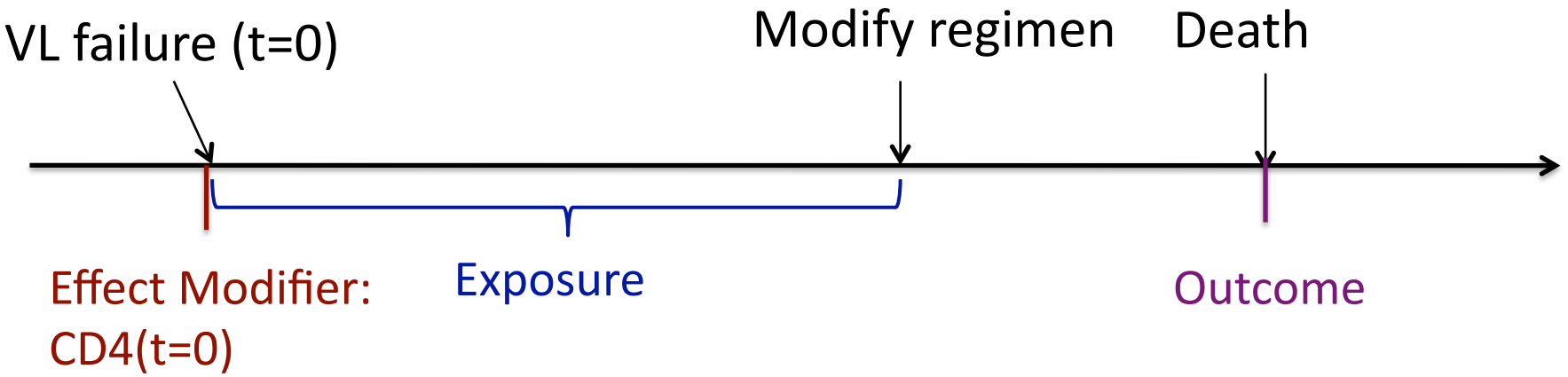
Not necessarily….

- Subject 1 meets CD4 failure criteria months after virologically failing
  - Substantial delay in modification if CD4 versus viral load strategy is used for patients like this
- Subject 2 meets CD4 failure criteria shortly after virologically failing
  - Minimal delay in modification if CD4 versus viral load strategy is used for patients like this
- *Is delay harmful for subject 2, or just for subject 1?*

Defining the Target Parameter in Terms of a Series of Hypothetical Randomized Trials

- Hypothesis: The effect of delayed regimen modification differs depending on a subject's current CD4 count

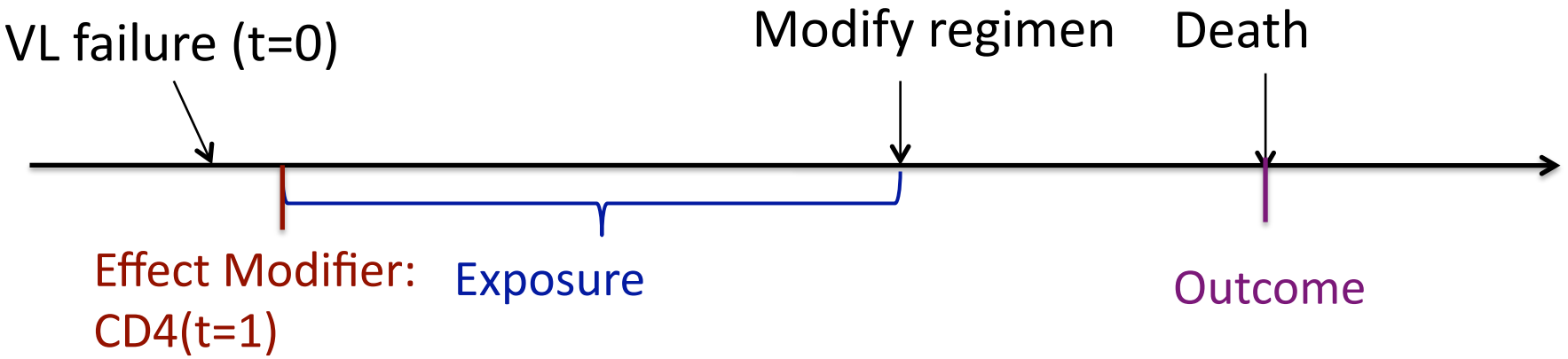
At t=0, randomly assign delay conditional on (within strata of) current CD4 count



Defining the Target Parameter in Terms of a Series of Hypothetical Randomized Trials

- Hypothesis: The effect of delayed regimen modification differs depending on a subject's current CD4 count

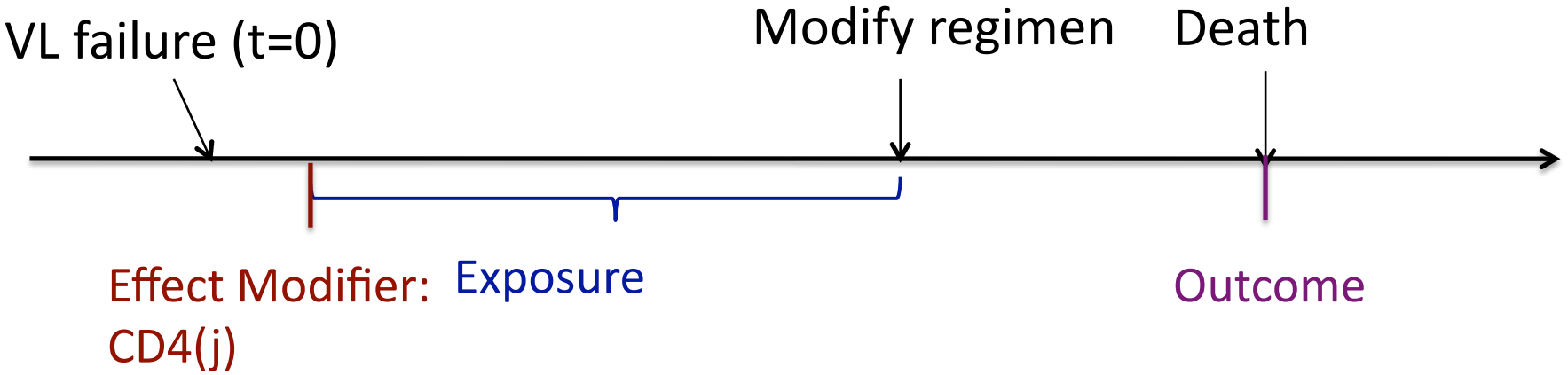
At t=1, randomly assign delay time to subjects who have not yet switched conditional on current CD4 count



Defining the Target Parameter in Terms of a Series of Hypothetical Randomized Trials

- Series of trials, beginning at successive times j
  - Effect of additional delay until regimen modification on future outcome
  - Modification of this effect by CD4 count at time j

At time *j*, randomly assign delay time to subjects who have not yet switched conditional current CD4 count



History-Adjusted Marginal Structural Models (HA-MSM)

- Generalization of standard MSM
  - Different parameter of interest
    - Indexed by interventions beginning at different time points
  - HA-MSM assume a standard MSM beginning at each time point
    - For counterfactuals indexed by treatment after time j
    - Conditional on some subset of the observed history up till time j
- HA-MSM allow us to assume common parameters across time points

Defining the Target Parameter Using a History-Adjusted Marginal Structural Model

- Ex: Model the counterfactual probability of survival at least 3 years among subjects who have not yet switched therapy as a function of future switch time and current CD4 count

$$E(Y_{Orig(j)=1,switch}(j+3)|Orig(j)=1,CD4(j)) =$$
$$\beta_0 + \beta_1 \min(j+3, switch) + \beta_2 CD4(j)$$
$$+\beta_3 \min(j+3, switch) \times CD4(j)$$

Defining the Target Parameter Using a History-Adjusted Marginal Structural Model

- Allows us to test the hypotheses

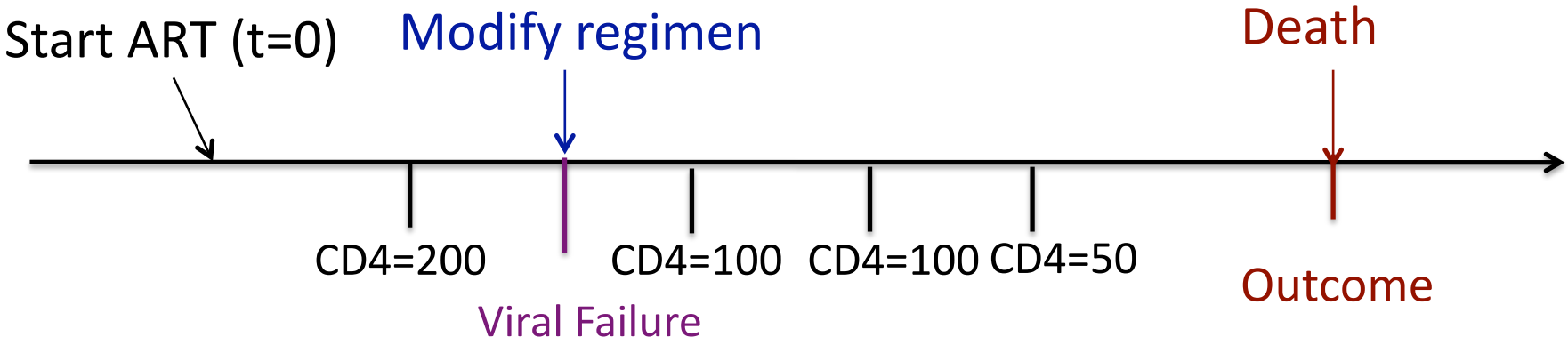
1. Effect of delayed regimen modification following virological failure differs depending on a subject's CD4 count during failure
2. Delayed regimen modification remains harmful for subjects who maintain CD4 counts about the WHO switching threshold
   - These are the subjects who will actually be subjected to delays
   - Evidence for importance of VL monitoring

Impact of alternative monitoring strategies

- Will delayed regimen modification after loss of viral suppression increase mortality? How much?
- Will use of CD4 (or clinical) rather than VL-based switching criteria increase mortality? How much?
- Which CD4 criteria will give the best patient outcomes?

Hypothetical Randomized Trial

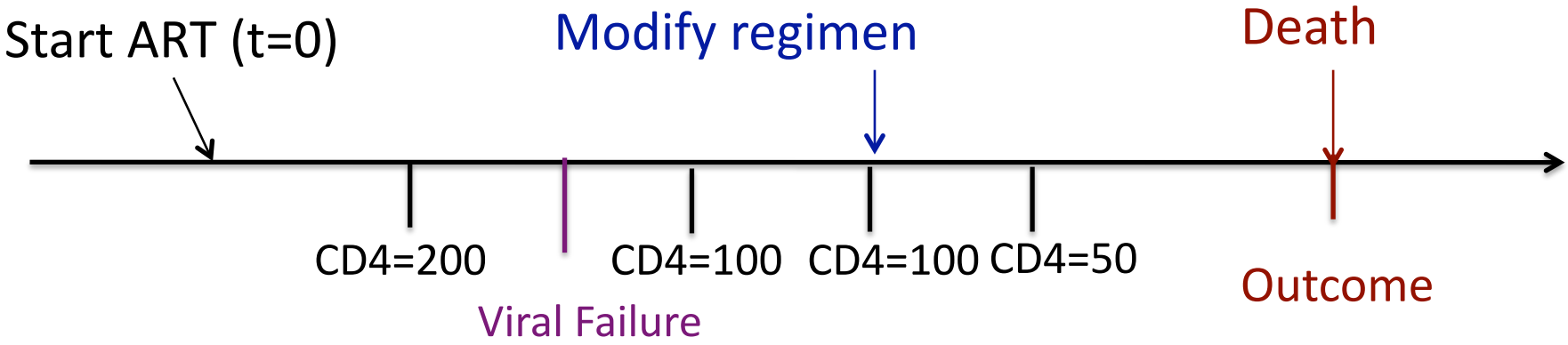
- At time of ART initiation, randomly assign patients to VL versus CD4 monitoring
  - Rule=I (assigned to VL arm)
- Stay on first line regimen until meet failure criteria for your assigned strategy
- Measure Outcome



Exposure=What strategy are you assigned to? ("Rule")
Ex: Rule=1 (assigned to VL arm)

Hypothetical Randomized Trial

- At time of ART initiation, randomly assign patients to VL versus CD4 monitoring
  - Rule=I (assigned to VL arm)
- Stay on first line regimen until meet failure criteria for your assigned strategy
- Measure Outcome



Exposure=What strategy are you assigned to? ("Rule")
Ex: Rule=0 (assigned to CD4 arm)

Dynamic versus static regimes

- A different type of intervention
- Rather than assigning a fixed treatment, assign treatment according to a strategy or rule
  - What treatment a given subject gets depends on that subject's covariate values
  - Ex. VL trajectory, CD4 trajectory…
- This is an example of a dynamic regime question
  - Different counterfactuals
  - Different estimation methods

Dynamic Counterfactuals

- T_{rule} : counterfactual survival time under specific rule for deciding when to switch
  - Ex. $T_{rule=1}$ = a subject's survival time if he had followed a VL-based switching rule
  - Ideally, compare survival distribution under VL versus CD4 switching rules
- Example: $E(T_{rule=1} - T_{rule=0})$
  - Difference in mean survival time if whole population had followed a VL-based versus CD4-based switching strategy

Dynamic Marginal Structural Models

- Model on counterfactuals indexed by dynamic rules
  - Ex: Model for Discrete Hazard

$$P(T_{rule} = t | T_{rule} \geq t) = m(t, rule|\beta)$$

  - One possible model:

$$logit(m(t, rule|\beta)) =$$

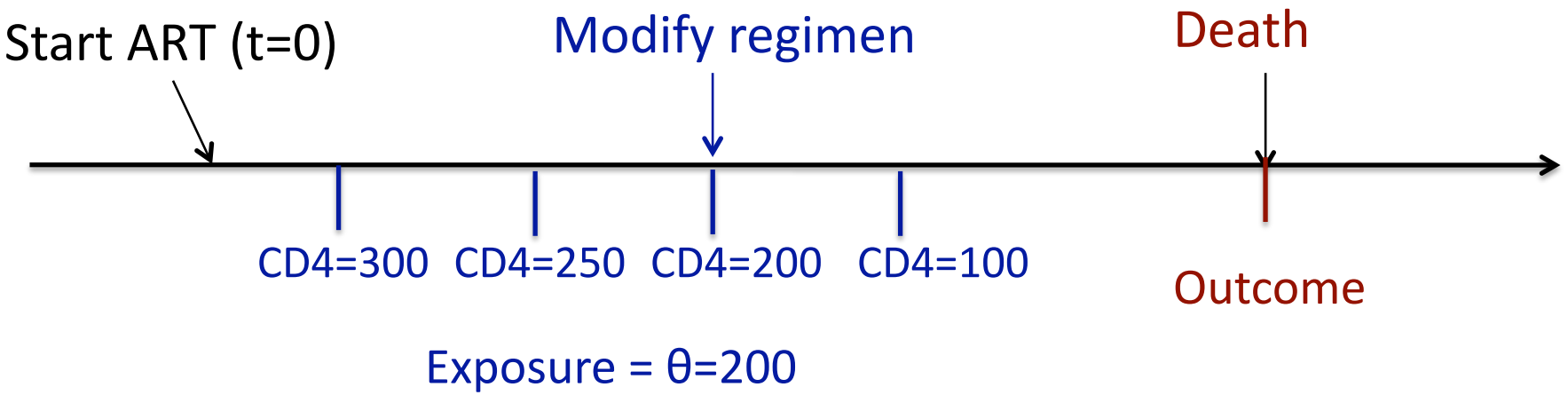
$$\beta_0 + \beta_1 rule + \beta_2 t + \beta_3 rule \times t$$

Impact of alternative monitoring strategies

- Will delayed regimen modification after loss of viral suppression increase mortality? How much?
- Will use of CD4 (or clinical) rather than VL-based switching criteria increase mortality? How much?
- Which CD4 criteria will give the best patient outcomes?

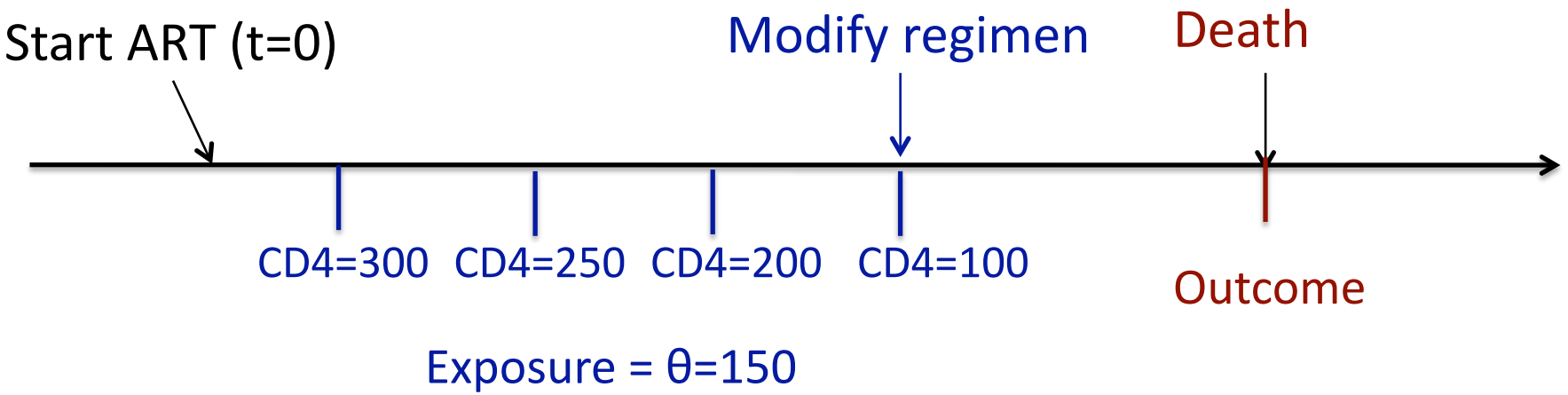
Hypothetical Randomized Trial

- At time of ART initiation, randomly assign CD4 modification threshold θ
- Stay on first line regimen until CD4 count meets this criteria, then modify immediately (or within some allowed window)
- Measure Outcome



Hypothetical Randomized Trial

- At time of ART initiation, randomly assign CD4 modification threshold θ
- Stay on first line regimen until CD4 count meets this criteria, then modify immediately (or within some allowed window)
- Measure Outcome



Dynamic marginal structural models (2)

- What is the "best" threshold?
- Ex. Model the probability of survival for at least 7 years as a function of modification threshold θ and baseline CD4 count

$$P(T_{rule_\theta} \geq 7) = m(\theta|\beta)$$

- One possible model:

$$logit(m(t, \theta|\beta)) = \beta_0 + \beta_1\theta + \beta_2\theta^2$$

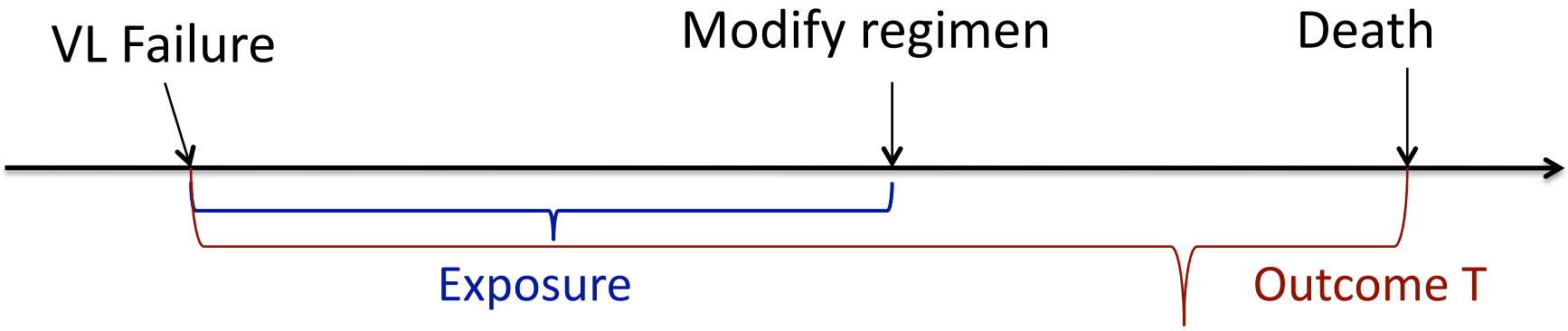
- Solve for θ_{opt} that maximizes probability of surviving at least 7 years….

Outline

1. What types of questions can causal inference methods help us to answer?
2. When do standard analysis methods break down?
   - Time-dependent confounding of longitudinal treatments
3. What estimation tools are available?
   - Inverse Probability Weighting
   - Longitudinal G- computation
   - New methods coming…

For Illustration, focus on first question

- Will delayed regimen modification after loss of viral suppression increase mortality? How much?



Simple Confounding

- Patients who are sicker (eg have lower CD4 counts/OD) at time of VL failure are more likely to modify immediately
  - Clinicians less likely to delay second line
  - These patients also more likely to die
  - Unadjusted analysis will underestimate harm from delayed switch
- Could control for this confounding by adjusting for baseline CD4 count (and other prognostic markers)

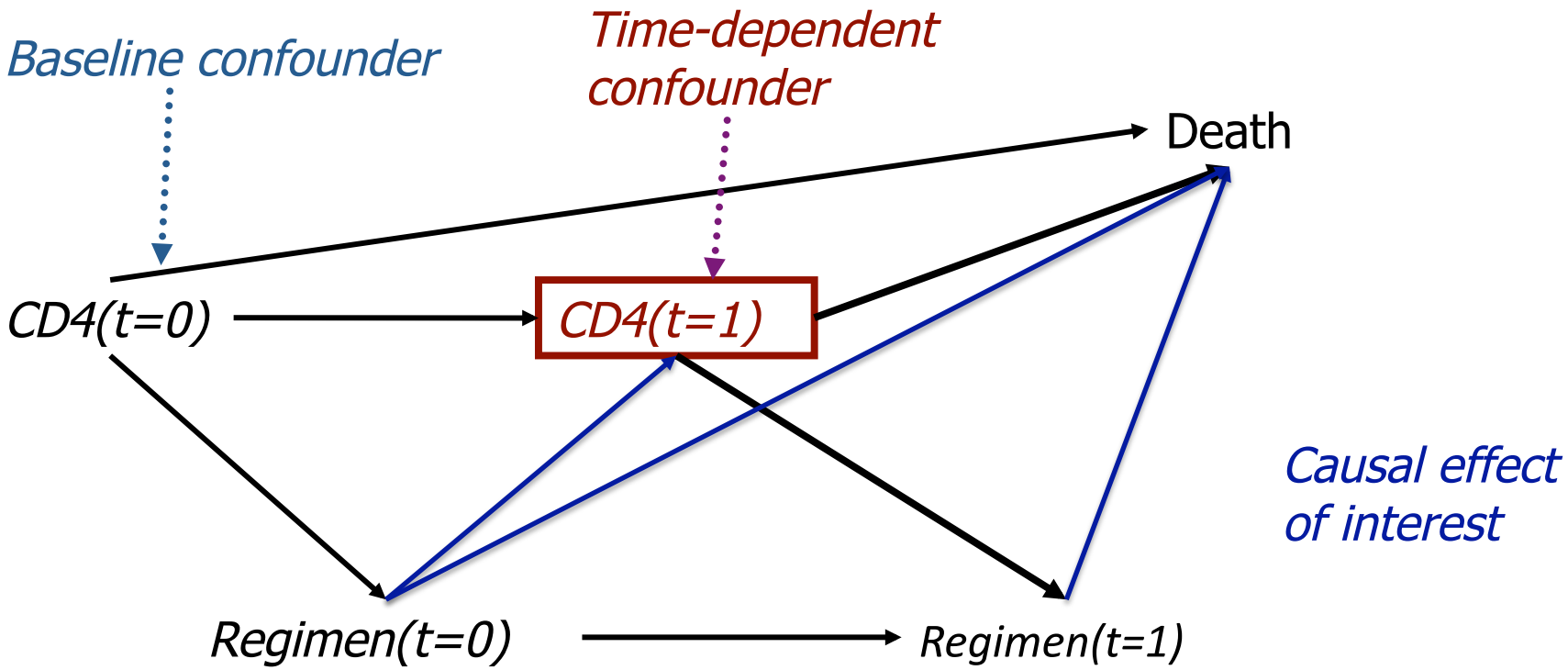
Time-dependent confounding

- Confounding by covariates after baseline
  - A subject whose CD4 counts decline rapidly following failure is more likely to modify therapy
  - A subject with rapidly declining CD4 has a worse prognosis
- Subjects who modify early are likely to be sicker than those who modify late
  - Even after controlling for baseline differences between the two groups
- Associations adjusted only for baseline covariates will still tend to underestimate the harm of delayed modification

Limitations of Standard Approaches

- Why can't we just adjust for these post-baseline differences between subjects who modify early and those who modify late?
  - Ex. Regress vital status on delay time, baseline CD4, and time-updated CD4?
- Post baseline confounders may be affected by the exposure!
  - Part of the causal pathway we are interested in
  - Adjusting for them will bias our estimates

Time-dependent confounding



- Can't control for *CD4(t=1)* in standard analyses: on causal pathway!

Outline

1. What types of questions can causal inference methods help us to answer?
2. When do standard analysis methods break down?
   - Time-dependent confounding of longitudinal treatments
3. What estimation tools are available?
   - Inverse Probability Weighting
   - Longitudinal G- computation
   - New methods coming…

Inverse probability of treatment weighting (IPTW)

- Confounding can be viewed as a problem of biased sampling
- Because the decision when to switch is not randomized, sicker patients are over-represented in the group that switches earlier
- Goal of IPTW: reweight the sample to remove these imbalances

Inverse probability of treatment weighting (IPTW)

- IPTW: Creates a re-weighted dataset in which the exposure is randomized
  - Up-weight individuals whose observed exposure is rare given their covariates
  - Count the experience of rare individuals more than once to make up for people like them who are missing
- Ex. Sicker patients that delay modification get bigger weights

Implementation of IPTW (1)

1. Model the probability of modifying therapy (exposure) given covariate history among subjects who remain on their failing therapy
   - E.g. Using multivariable logistic regression
2. Use this model to calculate predicted probability of each subject having his observed exposure history (switch time)

Implementation of IPTW (2)

3. Weight= 1/ predicted probability of having observed switch time, given observed covariate history
   - Can also use stabilized weights
4. Fit a weighted regression of survival on time spent on failing therapy
   - Using specified MSM
   - Ex. Pooled logistic regression to estimate discrete hazard ratios

Inverse weighting is just one possible way to adjust for time-dependent confounders…

- IPTW: Control for confounding based on estimate of the probability of treatment over time given the past

- G-computation: Control for confounding based on estimate of the distribution of time-dependent confounders (and survival) over time given the past

Longitudinal G-computation (1)

1. Model the distribution of each covariate at each time point given past covariates and past treatment
   - Ex. Using series of multivariable pooled logistic and/or linear regression models

2. Model hazard of death at each time point given past covariates and past treatment
   - Ex. Using multivariable pooled logistic regression model

Longitudinal G-computation (2)

3. Use these models to simulate distribution of survival times under various delay times
   - Monte Carlo Simulation
   - Generate predicted survival times for each subject under each possible switch time

4. Using simulated data, fit a regression of survival on switch time
   - Using specified MSM, such as pooled logistic model

Model Specification

- Consistency of IPTW estimator depends on consistently estimating how treatment assignment depends on confounders
- Consistency of G-computation estimator depends on consistently estimating how covariates (and outcome) depend on the past
- Getting the specification of models to estimate either can be very challenging
- Data-adaptive algorithms
  - Cross-validation to choose best bias-variance tradeoff
  - Example: Superlearner

Key Assumptions (1)

1. No unmeasured confounders
   - Key determinants of survival that also affect the decision when to modify are measured
   - Plausible?
     - Unrecorded comorbidities
     - Adherence

- May be reasonable to argue that residual confounding expected to be negative

-> Estimate of harm due to delayed switch conservative…

Key Assumptions (2)

2. Adequate variability in exposure
   - "Experimental Treatment Assignment/Positivity"
- Need variability in delay time
- Beyond this- need that covariates do not deterministically predict switch
  - Ex. If subjects with low CD4 always switch, no information about experience of subjects with low CD4 that do not switch

Estimator Comparison: IPTW

- Pros
  - Well-established for studying the effect of longitudinal treatments
    - Track record in HIV
  - Easy to implement
- Cons
  - Inefficient (high variability)
  - Substantial finite sample bias when some covariate/treatment combinations are rare

Estimator Comparison: G-computation

- Pro
  - More efficient
    - Parametric MLE-based estimation
- Cons
  - Harder to implement (but doable)
  - Requires modeling dependence of all time-varying covariates and survival on the past
    - Getting model specification right can be a challenge!
    - With African cohort data, limited number of longitudinal covariates and time points

A Third Option…
Double Robust Estimators

1. Augmented- IPTW
2. Targeted Maximum Likelihood estimation (TMLE)

- More robust and efficient
- TMLE software for point treatment effects available
- TMLE software for longitudinal effects: under development…pilot coming…

Current Research

- Implementation of above with
  - SA-IeDEA, some East Africa cohorts
  - CNICs (North America)
- Application of TMLE to longitudinal data/ dynamic regime estimation
  - Comparison with standard IPTW approaches
- Prediction of viral failure using MEMs data
  - Machine-learning for high dimensional prediction

Current Research

- Analysis of RCT data
  - Methods described can be used to estimate improved “as treated” effect
  - Secondary analysis to look at additional questions than those initially targeted by the trial
    - Ex. DART: RCT comparing clinical vs. CD4 arms
    - How would impact of CD4 monitoring have differed if CD4 monitored at different frequency?
- Design and analysis of studies to investigate community based interventions
  - CRTs/ Combination prevention programs in Africa

References: NPSEM/Counterfactuals/ G-computation formula

- Pearl J. (2009). “Causal inference in statistics: An overview," *Statistics Surveys,* 3:96—146.

- Pearl J. (2009). Causality: Models Reasoning and Inference. Cambridge University Press, Cambridge UK.

- Robins JM. (1987). “A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods”. *Journal of Chronic Disease* , 2:139s-161s.

References: Standard and HA- MSM

- Robins JM. (1999). “Association, causation, and marginal structural models”. *Synthese*, 121:151-179.

- Petersen et. al. (2007) "History-Adjusted Marginal Structural Models for Estimating time-varying Effect Modification” *AJE*

- Van der Laan, et al. (2005). "History-Adjusted Marginal Structural Models and Statically-Optimal Dynamic Treatment Regimens ", *IJB*: http://www.bepress.com/ijb/vol1/iss1/4

References: Dynamic MSM

- Hernan et. al. (2006) Comparison of dynamic treatment regimes via inverse probability weighting. *Basic Clin. Pharm Tox*; 98:237–242

- van der Laan & Petersen. (2007) Causal Effect Models for Realistic Individualized Treatment and Intention to Treat Rules. *Int. J. Biostat. 2007* 3.1: 3.

- Robins, et. al. (2008). Estimation and extrapolation of optimal treatment and testing strategies. *Stat Med.* Oct 15;27(23):4678-721

References: Data-adaptive Estimation

- Eric Polley's Website: www.stat.berkeley.edu/~ecpolley/
  - Super Learner Test R code and papers
- Van der Laan, Polley and Hubbard (2007) Super Learner. SAGMB; 6(25)
- Van der Laan and Dudoit (2003). Unified Cross-Validation Methodology For Selection Among Estimators … http://www.bepress.com/ucbbiostat/paper130