When To Start Antiretroviral Therapy

James Robins, Miguel Hernan, Lilliana Orellana, Andrea Rotnitzky

with colloboration of the investigators from

Multicenter AIDS Cohort Study (MACS, REF) and Women's Interagency HIV Study (WIHS), French Hospital Database on HIV (FHDH)

particularly Steve Cole, Dominique Costaggliola

Motivation

2004 International Aids Society Recommendations in JAMA for asymptomatic patients

RCT only in those with CD4 counts ≤ 200

Must rely on observational data for CD4 >200

Start

·when the CD4 is below 350 but above 200 and plasma HIV RNA exceeds 100,000 copies/ml or CD4 decline has exceeded 100/ul cells per year.

· otherwise start at CD4 count 200

Trade Off: Protect immune system from declining versus more time to develop drug resistance and side effects

Non-trivial To Figure Out How to Use Observational Data to Justify this regime as near-optimal

A Main Paper Leading to above decision rule:Paella et al.AnnInternal Med 2003

Simplified Description:

Outcome: Minimum of time to AIDS or death. Accrual of subjects 1996.

·Subjects who were already between 300 and 350 were followed from that measurement

·Those who started above 300 were regarded was group 1

·Those who started below 200 were regarded as group 2

·Those who never started were excluded.

Analysis better than many since follow up for both groups started at same time.

·Think if follow-up started at time that treatment began.

Then survival of those starting with CD4 $>$200 will be better than those with CD4 $<$200 even under null that "CD4 when started" makes no difference.

·Reason: Lead time bias- Must usually fall to below 200 to develop AIDS (old clinical def) or death

Nonetheless Paella Analysis has potential for severe bias

·Suppose prescribing trends increased with time as is the case. Take extreme example.

First year noone above 200 treated.

Second year every one below 350 treated

·Start below 200 (group 2) implies CD4 cell count fell below 200 within a year, so subjects were very sick and possibly died before 1 year.

·Start above 300 (group 1) guaranteed to have at least a year survival and included healthier subjects whose CD4 count was stable for a year.

·Under the null of no difference when you start (or even no treatment effect),

recomendation based on this analysis: start above 300

.

·Reason: Treatment above 300 is a surrogate for slowly declining counts and for survival itself (immortal person time bias) , so clearly it looks preferable.

·We will see Paella is still biased even if prescribing trends had not changed with time.

·Want to compare the 2 simple strategies:

start HAART when CD4 falls below 350 **versus** start HAART when CD4 falls below 200.

(assume no one starts above 350)

·Take subjects with CD4 counts exceeding 350

·when subject's CD4 first falls below 350 call it start of followup (time 0)

·randomize with p=1/2 to start immediately;

p= 1/2 to start when their CD4 first fell below 200.

·Compare survival using Cox PH

Such a trial has not been conducted so we must rely on observational data.

Samuel Beckett-like structure to talk

1.It can be done

2.It can't be done

3.It can be done

4.It can't be done

and so on ad infinitum

Because of time limitations we cover only point 1 and point 2,

Observational Analog: :

·Take subjects with CD4 counts exceeding 350.

·When a subject first falls below 350 he becomes on test at time 0.

·Regarded as group 1 (start below 350) if he initiates HAARTat time zero

·Otherwise in the group 2 (start below 200)

·The 2 groups are then compared using a standard COX PH model for time to AIDS or death that adjusts for the baseline (time 0) confounding factors such as HIV RNA, calendar date of entry, ethnicity, past rate of CD4 count decline, etc.

·However any group 2 patient who begins HAART at time t prior to falling below 200 or fails to start treatment at the time t he falls below 200 has failed to follow group 2 'protocol'

·That group 2 patient should be considered censored at t,

but **dependently censored** .

·No group 1 patient censored

·Patient who has never started treatment and has CD4 count $>$200 at end of follow-up counts in every risk set as a non-failure in group 2 since never censored!!

·Compare Paella. Excluded subjects who never started. Paella biased even were no trends in prescribing with time

Meaning of dependent censoring :

Consider all Group 2 subjects with CD4 cell count still greater than 200 cells/mL at time t after entry.

·The subgroup who initiate treatment (and become censored) at t will have, on average faster CD4 cell count decline form 0 to t and lower HIV RNA at t than those who do not initiate treatment (and remain uncensored) at time t.

·So individuals censored at t will have a worse prognosis than those who remain uncensored.

·In data, among those with a CD4 cell count less than 200 cells/mL for the first time at t, all started treatment as supposed to do.

Conclusion: A standard Cox analysis (which assumes independent censoring) will show death rate of the group 2 subjects artificially lowered

compared to what would have been seen in the above ideal randomized trial in which all group 2 subjects will have waited till 200 to start.

·So if better to start at 350, standard Cox may fail to detect this.

·**Method: Inverse probability of censoring weighting (IPCW) to adjust for dependent censoring**

.

·Solution use of inverse probability of censoring weighting (IPCW) on each group subject in each risk set.at time u.

·Each group 2 subject uncensored in risk set at time u divide by the probability of having remained uncensored to time u.

·This will fully adjust for dependent censoring due to measured time -dependent covariates, Here HIV RNA counts and rate of decline of $CD4$ cell count before u.

·Suppose a group 2 subject at risk at u with a CD4 count falling linearly from 350 to 250 from 0 to u has a probability of $1/4$ that he would not have started treatment by u.

·Then he counts for 4 people: himself and the three other similiar people who did start therapy.

·That is his $IPCW$ is 4 for the risk set at u.

·Subject CD4 falling 350-300 from 0 to u has probability of $1/2$ uncensored so

only counts for 2 persons

·In contrast if one adds CD4 count as a time-dependent covariate in the Cox model (rather than using IPCW weights), then

·bias due to dependent censoring remains because you have adjusted for a post randomization variable possibly affected by treatment.

·$Bias$ by such adjustment even if independent censoring

Ex: Suppose $people$ start HAART at random so no dependent censoring and HAART starting at 350 good.

1.$CD4$ on the causal pathway between HAART and AID/death

2.Among people at risk at u with same CD4 count history say 250 at u,

HAART does not predict failure at u.

So those who started at 350 same risk as those who have not started.

Falsely conclude no advantage to starting at 350

3. HAART helps because many more people have high CD4 counts at u if in group started at 350 than in thos who have not started..

·Conclusion:

No effect of HAART given CD4 count.

But HAART greatly improves survival by increasing CD4 count

.

For each group 2 subject in risk set at time u mutiply by the inverse of the estimated probability of having remained uncensored to time u , say $\widehat{W}(u)$.

$$\begin{aligned}
&\log it\left\{pr\left[A\left(j\right)=1|A\left(j-1\right)=0,Z\left(j-1\right)\right]\right\} \\
&= \alpha^{T}Z\left(j-1\right), \\
&\text{where } Z\left(j-1\right)=\left(RCD4\left(j-1\right),HIV\left(j-1\right)\right)
\end{aligned}$$

$$\begin{aligned}
\widehat{W}\left(u\right) &= \frac{1}{\prod_{j=1}^{u}\widehat{pr}\left[A\left(j\right)=0|A\left(j-1\right)=0,Z\left(j-1\right)\right]} \\
&= \prod_{j=1}^{u}\left\{1+e^{\widehat{\alpha}^{T}Z(j-1)}\right\} \\
&= \frac{1}{\prod_{j=1}^{u}\widehat{f}\left[A\left(j\right)\ |A\left(j-1\right),Z\left(j-1\right)\right]}
\end{aligned}$$

$$\widehat{W}(u) = \frac{1}{\prod_{j=1}^{u} \widehat{f}\left[A(j) \mid A(j-1), Z(j-1)\right]}$$

$\cdot f\left[A(k)|\bar{A}(k-1), \bar{L}(k)\right]$ is the probablity of having your observed treatment at time k given you past treatment and covariate history.

·Ex: 50 people have my same $\bar{A}(9), \bar{L}(10)$ through week ten.

Of those 50 people, 19 had $A(10) = 1$.

I have $A(10) = 0$.

So for me $f\left[A(10)|\bar{A}(9), \bar{L}(10)\right] = 31/50$

·Different from propensity score.

IPCW valid (ie reproduces results of the RCT) if

·$Z\,(j-1)$ all joint risk factors for treatment (censoring) and failure

· logisitc model for treatment (censoring) correct

·can be made doubly robust: ie either model for censoring or model for AIDS-free survival given $Z\,(j-1)$ (roughly)

Actual Analysis:

2344 HIV-infected subjects included in the French Hospital Database on HIV (FHDH) who had their first CD4 cell count measurement below 500 cells/mL (not 350) between 1 January 1996 and 30 June 2004, and who had never received antiretroviral therapy before that measurement.

We followed these subjects from their first CD4 cell count measurement below 500 cells/mL (baseline) until a diagnosis of AIDS, death, or June 2004, whichever occurred earlier.

Data on HAART use, as well as on time-dependent covariates (e.g., CD4 cell count) were recorded throughout the follow-up.

·There were 131 subjects in group 1,

·2217 in the group 2.

·655 subjects in group 2 were censored

·when they either started HAART before their CD4 cell count dropped below 200 cells/mL

or failed to start HAART first time below 200.

- RR 0.9 (95% confidence interval: 0.4, 1.8) just straight Cox.
- RR 0.5 (95% confidence interval: 0.2,1.1) IPCW:
- Appears better to start at 350 than 200

In practice want to answer:

·Find the optimal CD4 count x at which to start.

·That is want to compare all x in the candidate set $\{500, 499, ..., , 200\}$ rather than just 2 dynamic regimes.

·Would like to use expected utilities Y rather than Cox PH

·With K end of FU time, need something like

$Y = T \; if$ AIDS or death before K

$Y = K + 4\frac{CD4}{500}$ if survives to K (or possibly other quality of life measures)

·Let Y_x be utility if start HAART first time CD4 falls below x a particular regime.

·Find x_{opt} that maximizes $E[Y_x]$.

·If randomized trial with full compliance where different subjects with baseline CD4>500 randomized to different regimes x.

$\widehat{E}[Y_x]$ is average of Y among subjects randomized to regime x (even if never took HAART because CD4 always above x)

What if few subjects randomized to any one x.

·Fit ITT flexible polynomial (say 5th order) regression by least squares

$$Y = \beta_0 + \sum_{k=1}^{5} \beta_k x^k + \epsilon \tag{1}$$

to the n study subjects.

·Use first year calculus to find the value $\widehat{x}_{opt}$ of x where the fitted polynomial $\widehat{E}[Y_x] = \widehat{\beta}_0 + \sum_{k=1}^{5} \widehat{\beta}_k x^k$ obtains its maximum.

How to mimic in an observational study: (method based on dynamic marginal structural models).

Data $\left(Y, \bar{L}(K), \overline{A}(K)\right)$ among subjects with baselien CD4>500.

$\overline{A}(k) = \{A(0), ..., A(k)\}$ treatment history

Temporal Ordering: $L(m)$ before $A(m)$.

$\left(\bar{L}, \overline{A}\right) = \left(\bar{L}(K), \overline{A}(K)\right)$

Once started on treatment considered always on treatment .

·Estimate by pooled logistic regression

$$pr\left[A\left(k\right)=1|\overline{A}\left(k-1\right)=0,\bar{L}(k)\right]$$

$$\begin{gathered}\log it\left\{pr\left[A\left(k\right)=1|\overline{A}\left(k-1\right)=0,\bar{L}(k)\right]\right\} \\ =\alpha^{T}Z\left(k\right), \\ \text{where } Z\left(k\right)=z\left(\bar{L}(k)\right)=\left(k,CD4\left(k\right),RCD4\left(k\right),HIV\left(k\right)\right)\end{gathered}$$

·Consider a subject who started anti-retroviral therapy at a CD4 of 250 in week t whose lowest prior CD4 counts was 300. This subject followed all of the regimes $x = 251, 252, ..., 300$.

·Consider a subject who never started therapy and whose lowest CD4 count was 225. This subject followed regimes $x = 200, 201, ..., 225$.

·Consider a subject who started anti-retroviral therapy at a CD4 of 250 in week t whose lowest previous CD4 counts was less than 250. This subject failed to follow any regime. in the candidate set $\{500, 499, ..., 200\}$

·Consider the data set $(Y_i, x_{i1}), (Y_i, x_{i2}), ..., \left(Y_i, x_{i\Gamma_i}\right)$, where the x_{ik}, $k = 1, ..., \Gamma_i$ denote the Γ_i regimes followed by subject $i, i = 1, ..., n$

·Let $\widehat{W}_{ik}$be the inverse of the estimated probability that subject i followed regime x_{ik}, without being censored (i.e starting when fell below x_{ik} without starting HAART at a value above x_{ik}).

·$\widehat{W}_{ik}$=$\widehat{W}_i$ $=$ the inverse of the probability that subject i followed his own observed treatment history.

$$
\begin{aligned}
\widehat{W}_i &= \prod_{k=0}^{K} \frac{1}{f\left[A_i(k)|\bar{A}_i(k-1), \bar{L}_i(k); \widehat{\alpha}\right]} \\
&= \frac{I\left\{K_{start} \leq K\right\}}{\exp it\left[\widehat{\alpha}^T Z\left(K_{start}\right)\right]} \prod_{k=0}^{\min(K_{start}, K, Death)} \frac{1}{1 - \exp it\left[\widehat{\alpha}^T Z\left(k\right)\right]} \\
&= \frac{I\left\{K_{start} \leq K\right\}}{\exp it\left[\widehat{\alpha}^T Z\left(K_{start}\right)\right]} \prod_{k=0}^{\min(K_{start}, K, Death)} \left\{1 + \exp\left[\widehat{\alpha}^T Z\left(k\right)\right]\right\}
\end{aligned}
$$

Including $k = 0$ (baseline) controls confounding by baseline factors.

·Fit the flexible polynomial (say 5th order) regression

$$Y = \beta_0 + \sum_{k=1}^{5} \beta_k x^k + \epsilon \tag{2}$$

by weighted least squares applied to the data set (Y_i, x_{ik}) with weights $\widehat{W}_i, k = 1, ..., \Gamma_i, i = 1, ..., n$

·Use first year calculus to find the value $\widehat{x}_{opt}$ of x where the fitted polynomial $\widehat{E}[Y_x] = \widehat{\beta}_0 + \sum_{k=1}^{5} \widehat{\beta}_k x^k$ obtains its maximum.

Mathematically the above analysis consistent for $E[Y_x]$ and thus for x_{opt} if no unmeasured confounders for treatment

$$f\left[A(k)|\bar{A}(k-1),\bar{L}(k),Y_x\right] = f\left[A(k)|\bar{A}(k-1),\bar{L}(k)\right]$$
$$Y_x \text{ ind } A(k)|\bar{L}(k),\bar{A}(k-1)$$

and positivity for treatment

$$pr\left[A(k)=1|\bar{A}(k-1)=0,\bar{L}(k)\right] > 0$$

Applied this method to the MACS-WIHS data and obtained

$\widehat{x}_{opt} = 289$ with CI $(266, 312)$

We can allow for the fact that the optimal treatment regime in our candidate set may differ depending on a subject's measured pretreatment variables V such as gender, ethnicity, HIV risk group (eg IV drug users vs homosexual contact), genetic profile (SNPs)

·For instance we might replace the model $Y = \beta_0 + \sum_{k=1}^{5} \beta_k x^k$ with $Y = \beta_0 + \sum_{k=1}^{5} \beta_k x^k + \gamma_0^T V + \sum_{k=1}^{5} \gamma_k^T V x^k$

·Use first year calculus to find the value $\widehat{x}_{opt}(V)$ of x where the fitted polynomial $Y = \widehat{\beta}_0 + \sum_{k=1}^{5} \widehat{\beta}_k x^k + \widehat{\gamma}_0^T V + \sum_{k=1}^{5} \widehat{\gamma}_k^T V x^k$ obtains its maximum.

The method can be made robust by replacing the assumption of a linear main efffect $\gamma_0^T V$ of V by an arbitrary function $h(V)$

A more complex set of candidate regimes can be optimized in both randomized and observational studies..

·If current HIV RNA is greater than z, start HAART if the current CD4 count is less than x

·If current HIV RNA is not greater than z, start HAART if the current CD4 count is less than q.

We can use the same methods to jointly find $\left(z_{opt}, x_{opt}, q_{opt}\right)$.

Classic Problem: Find the optimal regime given data $L_0 A_0 L_1 A_2, ..., L_K A_K L_{K+1}$ where A_k is HAART at time k. L_k covariates at time k.

$\overline{d}_{opt} = \left\{ d_{opt,k} \left(\overline{a}_{k-1}, \overline{l}_k \right), k = 0, 1, ..., K \right\}$ that maximizes the mean of $Y_{\overline{d}}$ over all regimes $\overline{d}$.

New statistical models for estimating $\overline{d}_{opt}$ from observational data, that are a clever empirical twist on dynamic programming (backward induction) first by Susan Murphy (2003) and then by myself (2004).

Benefits: Can find optimal regime in a much bigger class of regimes than the classes looked at using dynamic marginal structural models.

Limitations: Optimal regime can be very (too) complicated. But dynamic programming methods cannot simplify the class too much since must allow any possible past history of treatment and at least one covariate in $\bar{l}_k$ say CD4 count.

Estimated Optimal Regime: Initiate HAART at t if the following exceeds zero:

21.4 - 4.1 times current CD4 count - 3.2 times the rate of decline of CD4 + 5.6 times the square of the rate of decline of CD4 + 2.1 times the product of the current CD4 count and the rate of decline.

If we want something simple like first time CD4 falls below x_{opt} so easy to implement in many places.then use previous method: find optimal regime in a prespecified set such as first time CD4 below x

Problem with MACS WHIS

·Seen only every 6 months

·Treatment Decisions made by treating physicians based on unknown CD4,HIV RNA clinical responses at time of decison (except for 6 month interval)

Possible Confounding.

Solution: Continuous time records where all physician visits, labs, and pharmacy records are obtained eg HMO, so we know what the prescribing physician knew.

FHDH approximates this but many new problems arise.

Consider again regimes "begin HAART the first time t the measured CD4 count falls below x," $x \in \mathcal{X} = \{201, ..., 500\}$.

Let x_{opt} be the $x \in \mathcal{X}$ for which the expected utility $E[Y_x]$ is a maximum.

x_{opt} will depend on the frequency with which subjects have their $CD4$ count measured.

Consider particular population. $CD4$ counts are obtained very frequently, say every 3 weeks, and x_{opt} would be 340.

Second population CD4 obtained only every 6 months. Then x_{opt} would presumably exceed 340:

Second population cannot change frequency for financial or logisitical reasons.

Many subjects who were well above 340 at the time of their last CD4 blood test but were not started on HAART would have CD4 count well below 340 6 months later (say 250).

They would have been started three months earlier if in first population if 3 months ago they were already below 340.

The Extrapolation Problem:

·Can we use the data from the first population to estimate x_{opt} in the second pop g

if populations "biologically" the same. Discussed below.

Data: $\left(Y, \bar{L}(K), \overline{A}(K), \overline{T}(K)\right)$.

Ordering $L(m) = (W(m), I(m))$,$A(m), T(m)$.

Observed Data: $L^*(m) = (W(m), I^*(m)), I^*(m) = I(m) T(m-1)$

Yes, if we can estimate x_{opt} in the first population ie

no unmeasured confounders for treatment and positivity for treatment

$$\begin{aligned} & pr\left[A(k)=1|\bar{A}(k-1)=0,\overline{T}(k-1),\bar{L}^{*}(k),Y_{x}\right] \\ = \; & pr\left[A(k)=1|\bar{A}(k-1)=0,\overline{T}(k-1),\bar{L}(k)\right] \\ > \; & 0 \end{aligned}$$

and

(a) both CD4 testing in the first and second population have no unmeasured confounders

$$
\begin{aligned}
& pr\left[T(k)=1|\bar{A}(k)=0,\overline{T}(k-1),\bar{L}^{*}(k),Y_{x}\right] \\
= \;& pr\left[T(k)=1|\bar{A}(k)=0,\overline{T}(k-1),\bar{L}(k)\right] \\
> \;& 0
\end{aligned}
$$

and with pr_g replacing pr.

and either

(b) we have two way positivity for testing in the first population relative to the second

$$\begin{aligned} pr_g\left[T(k)=1|\bar{A}(k)=0,\overline{T}(k-1),\bar{L}(k)\right] &> 0, \text{then} \\ pr\left[T(k)=1|\bar{A}(k)=0,\overline{T}(k-1),\bar{L}(k)\right] &> 0 \end{aligned}$$

and

$$\begin{aligned} pr_g\left[T(k)=0|\bar{A}(k)=0,\overline{T}(k-1),\bar{L}(k)\right] &> 0, \text{then} \\ pr\left[T(k)=0|\bar{A}(k)=0,\overline{T}(k-1),\bar{L}(k)\right] &> 0 \end{aligned}$$

or

(c1)we have one way positivity for testing in the first population relative to the second

$$\begin{aligned} pr_g\left[T(k)=1|\bar{A}(k)=0,\overline{T}(k-1),\bar{L}(k)\right] &> 0, \text{then} \\ pr\left[T(k)=1|\bar{A}(k)=0,\overline{T}(k-1),\bar{L}(k)\right] &> 0 \end{aligned}$$

and

(c2) CD4 tests have no direct effect on the outcome $Y_{\overline{a},\overline{t}} = Y_{\overline{a}}$

·(a) True in our example. Every 6 and 3 months are deterministic so clearly unconfounded

·(c2) true.

Diagnostic or prognostic tests are only useful in so far as they help a physician to determine the optimal dosing strategy, by providing information on both the current health state and the prognosis of a patient because, in contrast to drug therapies, these tests have no direct causal effect on disease progression.

(Actually not true in our example if change in treatment possible based on CD4 after starting HAART. Would really need all treatments recorded in $\overline{a}$, not just initial treatment)

·(c1) True in our example: Every 6 months in g every 3 months in study pop

·(b) false

·Bio log $ically$ the same means the distribution of $Y_{\overline{a},\overline{t}} = Y_{\overline{a}}$ the same in both populations

·If (a) and (b) true nothing new except weights now reflect treatment and testing history.

·Restrict previous analysis to subjects in the observed data who followed the testing history of population g subjects but with

$$\widehat{W}_i = \prod_{k=0}^{K} \frac{1}{f\left[A_i(k)|\bar{A}_i(k-1), \overline{T}_i(k-1), \bar{L}_i(k); \widehat{\alpha}\right]} \times \prod_{k=0}^{K} \frac{1}{f\left[T_i(k)|\bar{A}_i(k), \overline{T}_i(k-1), \bar{L}_i(k); \widehat{\alpha}\right]}$$

the inverse of the probability that subject i followed his own observed treatment and testing history.

In our example no such subjects

Method (a) and (c) true :

·Throw away all CD4 test data from first population except for 6 monthly tests.

·Now find the optimal x_{opt} from the modified data as above except with weighting by the inverse of the estimated probability of having the treatment they did indeed have based on the unmodified data.

·Note the regimes followed by a subject in the modified data will differ from those in the raw data.

·Data on person i, CD4 at 0,3,6,9 month 500, 375, 330, 500. i started at month 3 after CD4 375. Then in raw data followed regimes $x = 500, ..., 376$ but in modified data followed no regime.

If i started month 6 followed regimes $x = 375, ..., 331$ in raw data.

modified data followed $x = 500, ..., 331$ in

· Weights must depend on original raw data since must control confounding and doctor may have looked at all CD4 counts to decide when to stat therapy.

·Note this method only works because no direct effect of testing as we use people with tests every 3 months to represent people with tests every 6 months with same treatment history

·What if some people in study population are tested less frequently than

every 6 months but (a) and (c) true.

·Can we still use the data from the first population to estimate x_{opt} in the second

if populations biologically the same.

·Yes

Method:

·Keep only subjects in study population who had tests more frequently than every 6 months.

·Throw away all CD4 test data from first population except for 6 monthly tests.

·Now find the optimal x_{opt} from the modified data

(with weighting by the inverse of the probability of having the

treatment they do indeed have

$\times$

probability of having had a test history every 6 months,

both based on the unmodified data)

$$\widehat{W}_i = \prod_{k=0}^{K} \frac{1}{f\left[A_i(k)|\bar{A}_i(k-1), \overline{T}_i(k-1), \bar{\bar{L}}_i(k); \widehat{\alpha}\right]} \times \prod_{\{k;k/6 \text{ is whlole}\}}^{K} \frac{1}{pr\left[T(k)=1|\bar{A}_i(k), \overline{T}_i(k-1), \bar{\bar{L}}_i(k); \widehat{\alpha}\right]}$$

Can use these methods to extrapolate to a a population with any

$$pr_g\left[T(k)=1|\bar{A}(k)=0,\overline{T}(k-1),\bar{L}(k)\right]$$

, not just every 6 months.

In practice the testing schedule will not be known for population g and our specification of $pr_g\left[T(k)=1|\bar{A}(k)=0,\overline{T}(k-1),\bar{L}(k)\right]$ will reflect a best guess, which we treat as known in the analysis.

However, in practice, the selected g could be varied in a sensitivity analysis or be replaced by an empirical estimate if appropriate data were available on population g.

Suppose that in a particular population on which we have collected longitudinal data we do have the ability to increase at some financial cost, or decrease with some financial savings, the frequency of CD4 tests and we desire to use the data to determine jointly the optimal CD4 testing schedule and the optimal time to start HAART.

This issue is the well-known issue of assessing the "value of information" in sequential decision making; a CD4 test has no direct biological effect on a patient whatsoever. Its net value, if any, is that the information supplied by the test can be used to fine tune the time to start HAART, leading to a net increase in expected utility, even when the costs of the test are included in the utility function. Of course to include the costs of a CD4 test in the utility implies that our utility function is in dollars and that we must place a monetary value, usually adjusted for quality of life, on each additional year of life...

Exact same methods can be used.

Suppose we have data from a comprehensive HMO data base that records all patient visits.

HIV-infected patients come to clinic to be seen by a physician at a time t either because of acute symptoms or because of a scheduled follow-up appointment.

Data on the reason for a given visit are often not recorded for data analysis.

This random clinic visit process, unlike the regular monthly visit process we have assumed results in an association between visits and risk since patients who are sick are generally more likely to return to the clinic at frequent intervals both spontaneously and to keep frequent appointments (although it is possible moderately ill patients who are not getting better come infrequently and miss many scheduled visits.).

The visit process will result in non-ignorable missing data unless we not only collect data at t on the health status of those who return to the clinic at t (which we typically do) but also of those who do not come to the clinic at t (which we almost never do).

Suppose we accept that the visit process is nonignorable and therefore cannot be adjusted for even by inverse probability weighting without additional strong model-based assumptions.

In contrast, we feel it is often reasonable to assume that among patients coming to the clinic at t, the decison to treat or not can be viewed as effectively randomized (i.e., ignorable) given the past including data on the health staus measured at a visit at time t. Under this assumption, we can still estimate the optimal CD4 count at which to start HAART in the study population (provided

the utilty Y is always observed) but extrapolation to another population of biologically similiar individuals with a different visit process becomes problematic.

Without loss of generality, we consider the following stylized extrapolation problem where the extrapolation in question is temporal rather than spacial. The mathematics apply equally to either type of extrapolation.

Suppose, after a clinic visit (whether scheduled or emergent), the HMO's policy is to schedule the next clinic visit in roughly 12 weeks, unless a problem was found at the current visit, in which case the next visit is scheduled for two weeks.

Suppose the HMO decides to offer more intensive health care to better compete with its rivals by cutting the 12 week interval to 6 weeks and the two week interval to one week. CD4 counts are measured at every visit. Then, as argued previously, after the schedule change, the optimal CD4 count at which to start HAART will change.

However, even were the prechange visit process ignorable, data obtained before schedule change could not be used to estimate the optimal post-change CD4 count at which to start, if, as is likely, the maximum number of visits per

month of any patient after the schedule change exceeds the maximum before the change for posotivity problems

Suppose now the HMO decides to save money by increasing the 12 week interval to 24 and the 2 week interval to 4.

Then, if the pre-change visit process were ignorable and the visit process had no direct effect on one's health status, a straightforward extension of the above methods of Section 6.3 could be used to estimate the optimal post-change CD4 count at which to start from the prechange data. We redefine $T(m)$ to be the visit indicator, redefine $I(m)$ to be $L(m)$, eliminate and restrict to treatment regimes that allow treatment change only when a visit occurs.]

However, we have argued above that the visit process is nonignorable. In that case the methods of Section 6.3 cannot be used.

What then?

What if the populations biologically different.

· Patients come to hospital (ie HMO) at random times based on symptoms, routine follow-up appt, missed visit.

·$Data$ on why a visit generally not available except perhaps in good HMO (even routine follow-up)

· Typical solution- censor at certain time from last visit but this is dependent censoring, usually nonignorable

· If patients sick, more (or less) likely to come back,

we have association of treatment and risk

· Only can control this confounding if data on health for people who return at time t and on those who do not: Noone collects such data

· Thus only reasonable to assume, among subjects who have visit at time t with same past history decision to treat or not treat is at random.

· But then $E[Y_x]$ means treat not first time CD4 falls below x , but rather first time CD4 falls below x at a clinic visit.

· May then be impossible to extrapolate to other populations with different reasons and frequency for coming to a clinic.

· Say data from US HMO: HIV patients come every 3 weeks and subjects are compliant

· Consider extrapolation to another patient population where subjects are scheduled only every six months and few comply

· In second pop, those who arrive at clinic at t with CD4 below x for the first time are a very different, often symptomatic, selected sample.

·For example may have been below x for many months so would have been treated months before if in HMO.

Problems of Extrapolation of 'When To Start results' Due to Medical Treatment Patterns in addition to the other issues.

·$Africa$ has Tb Malaria different nutrition , health care etc

·Different Distributions of Time Since Infection if not in Seroconverter Group

Suggests only fairly important differences in results on "when to start" should be taken seriously

Actually,as my coauthor's Andrea and Lilianna have pointed out , even the RCT described previously can be 'biased' for deciding whether better to treat all HIV patients with the regime "start when CD4 first falls below 350" versus "start when CD4 first falls below 200" when HAART is already in clinical use

Matters are very subtle when even the obvious RCT can be 'biased' !

Recall the RCT design.

RCT

·Take subjects with CD4 counts exceeding 350

· Exclude subjects who start when CD4 count $>$350

·when subject's CD4 first falls below 350 call it start of followup (time 0)

·randomize with p=1/2 to start immediately; p= 1/2 to start when their CD4 first fell below 200.

·Compare survival using Cox PH

Suppose the log rank test shows no difference between the two arms shows no difference

Possible 'Bias' attributable to HAART already being in clinical use: Consider the subjects who were not entered in the trial because their physicians had started them on HAART when there CD4>350. Likely these would among the subgroup of patients whose CD4 was falling rapidly (although still >350) and/or very high HIV RNA.

Suppose only this subgroup benefits from starting when CD4 first falls below 350 rather than waiting to 200. Then by excluding most of the subgroup from the trial, we failed to discover that the 350 arm would have done better than the 200 arm had all HIV subjects been entered.

Of course, a better answer response to this 'bias' is to recognize that treatment initiation should be based on rate of CD4 decline and HIV RNA level as well as current CD4 level (as suggested in the JAMA recommendations). Discussed above.

Possible 'Bias' attributable to short follow-up (time to adminstative censoring).:

How large is this bias?

Try to quantify by comparing estimates that for subject who starts at CD4 x_3 at time u_3 on his third hosp visit who had earlier visits at times u_1 with $CD4$ x_1 and u_2 with $CD4$ x_2

·weights W_{ik} of one over cond prob of not have started HAART at any of the previous visits but starting at this visit

versus

·weights W_{ik} of one over prob of not have started HAART at any of the previous visits

but starting at this visit times one over the probability of having visits at exactly times u_1, u_2, u_3.

· If these were very similiar some evidence that the subjects with a given visit pattern are not a self selected group whose risk differs from subjects with another visit pattern based on unmeasured factors.

· would allow one to extrapolate to another set of subjects for whom visits patterns are also not self-selected based on risk factors. So even if ok in no bias in US not extrapolable to the second group.