We want to the last Biostatistics seminar for the fall series. It’s my great pleasure to welcome our speaker, Dr. Liangyuan Hu. Dr. Hu is an Assistant Professor of Biostatistics in the Department of Population Health Sciences and Policy at Mount Sinai School of Medicine. She received her PhD in Biostatistics from Brown University. Her methods research focuses on causal inference with complex longitudinal and survival data and Bayesian machine learning. Her independent research has been funded by NIH and Patient Centered Outcomes Research Institute. And her paper in Biometrics has been selected to receive the 2019 Outstanding Statistical Application Award by the American Statistical Association. Today, she’s going to share with us her recent work on developing a continuous time marginal structure of models for complex survival outcomes. Liangyuan, the floor is yours. Well, thank you Li Fan. Thank you so much Fan for your introduction, for the invite also. Let me just share my slides full screen.
I’m really excited to be here today to talk about some of the projects I’ve been working on in the causal inference field, namely, how do we use marginal structure models for more complex comparative effectiveness research questions involving continuous-time treatment and censored survival outcomes.

So I’d like to first acknowledge my colleagues, especially Doctors Hogan and Daniels who had been instrumental to me during the time I was working on this project. And let me just shift to the bar a little if I can.

So this is just for those who aren’t very familiar with causal inference, and simple slide to introduce the concept. Suppose we are interested in estimating the causal effect of a binary treatment A on some outcome Y. Using the potential outcomes framework, we can define the average treatment effect as the difference between the mean of the two sets of potential outcomes. So, Y1 here is the potential outcome that would have been observed if everyone in the population received the treatment.

Similarly, Y0 here is the potential outcome had everyone in the population received the treatment.
that would have been observed had no one in the population received the treatment.

To estimate the causal effect, the gold standard is the randomized controlled file. So in an RCT, we would randomly allocate patients to receive either treatment or the control or placebo, the randomization would make the two groups of patients more or less very similar in terms of their characteristics. So in a sense that these two groups are exchangeable, so that an individual’s potential outcome to either treatment or control would not depend on which treatment group this person was assigned to. But just depends on how the treatment works. And this way we can simply look at the difference and the mean of the observed outcome between the two treatment groups and just to estimate the causal effect. But in many, many situations, we cannot conduct an RCT and we have to rely on observational data to get the causal inference about treatment effects. So in these situations, the independence between the potential outcome
and treatment assignment would no longer hold.

Because there might be a confounder that is predictive of the outcome, such that the probability of receiving the treatment depends on the confounder.

For example, younger patients may be more likely to receive the treatment. So in this case, if you take the difference in the average of the observed outcome between the two groups, then this estimate would not bear a causal interpretation because the difference might be confounded by age.

So we would have to use specialized causal inference techniques to remove the confounding.

And there are just many, many techniques out there, but today I’m just gonna focus on marginal structure model, because it is simple to implement. It has good statistical properties, and it is versatile enough to accommodate many, many complications posed by observational data that I’ll talk about later.

So we can propose a marginal structure model.
relating the potential outcome to the treatment assignment.

And here theta one would capture the causal effect.

But in reality, we can only fit a model to the observer data. And as I talked earlier, the parameter estimator beta one here would not bear a causal interpretation, it just measures association. But we can get to causation if by solving the weighted estimating equation, using the weight as W inverse of conditional probability.

So that’s just a simple example to illustrate the use of marginal structure model.

And traditionally treatment assignment, treatment is assigned at baseline and it’s time fixed.

So it means that the treatment doesn’t change over time, but with increased availability of healthcare data sets, there are increased demands for more refined causal inference methods to evaluate complex treatment regimens.
So one example is that treatment initiation can actually depend on time, so it changes over time.

In this case, it would just be impractical to conduct RCTs because there are just simply too many treatment initiation time points.

So I’m going to use two motivating examples in this talk.

The first example is about timing of treatment initiation for patients who present both HIV and TB, tuberculosis.

For these patients, TB treatment will be initiated immediately after the diagnosis, but during the TB treatment,

when is the optimal time to initiate the HIV treatment or ART, anti-retroviral therapy?

That is a very important question to answer, because if you initiate the treatment too early, there might be drug interactions, drug toxicity, but if you delay the treatment too much, then there’s also increased the mortality associated with AIDS.

The second example is timing of HIV treatment for adolescents.

The timing now is defined with respect to the evolving value of a biomarker CD4.

And this is also an important question to answer
because the WHO guideline is in the form of treat this person when the person’s CD4 cell count drops below 350, for example,
and for the population of adolescents currently there’s no concrete evidence for supporting the optimal threshold.
So to statistically formulate these two motivating examples,
the first one, when is the best time to initiate a treatment?
So this is actually a static treatment regimen with respect to time,
and the initiation can occur on the continuous timescale.
And second example is actually a dynamic treatment regimen.
It’s dynamic because it depends on the evolving history
of treatment and a biomarker,
but initiation can also occur on the continuous timescale.
So marginal structure models are suitable for addressing a time dependent treatment,
but in order to use the models, we have to overcome some statistical challenges.
The first challenge is that we need to estimate the causal effect of the actual timing,
not compare protocols defined by some specific intervals,
which is a lot of existing studies did.
And also a lot of RCT reported these kinds of results. Because as I said earlier, it’s just impractical for RCTs to report continuous time causal effects. We would also need to address complications posed by observational data. This is something I’ll talk about later. And also we are dealing with censored survival outcomes that adds another layer of complexity.

So these are four sensory patterns observed in our data. So our goal is to estimate the causal effect of A, treatment initiation time and T, death time. And we have almost 5,000 patients and only a very small proportion of patients have both observed A and T. A lot of patients don’t have observed T. So their death time is censored by C. And we have about 20% of our patients, they don’t even have observed A. Their treatment initiation time can be censored by death time or censored by C, dropout, for example. So our goal is to estimate effect of A on T, but we only have about 300 patients have complete information. Most of the patients we have incomplete information on either A or T or both.
How do we probably use these incomplete information to draw causal inference about A on T, the effect of A on T, that’s a problem we solve in this project. So three challenges. First one, treatment initiation time, this is observational data, so it’s not randomly allocated. We don’t know the actual functional form of causal effect of initiation timing or mortality rate. And we see that, Oh, there’s incomplete information on either exposure or outcome or both. The general solutions we proposed that we first formulate a flexible structural causal hazard model that can capture the effects of both timing and duration. Using the model outputs we can estimate the functional form of the causal relationship between our initiation timing and mortality. So some notation before we introduce our approach, note that we have three time to events in our study,
we have treatment initiation time, death time, censoring time. We'll use \( T_{\text{cap A}} \) to denote death time associated with the actual treatment time. And potential outcomes \( T_{\text{small A}} \), this is the death time. If treatment initiated at time \( A \), and we use \( T_{\text{infinity}} \) to denote death time if treatment is initiated beyond sometime point. Because of all the censoring, all the three time to events can be censored by one another. We use \( T^* \) to denote the minimum of \( T \) and \( C \). Delta \( T \) is a corresponding event indicator. So \( A^* \) is the minimum of the three time to events. Delta \( A \) is a corresponding event in the data. Adopting the convention in the causal inference literature, we use \( \overline{L} \) of \( T \) here is a covariate history up to a time \( T \). Putting everything together, we have a set of observed data. Now back to the censoring patterns. In case one, we observed both \( A \) and \( T \). So we would observe \( A \), we would observe \( T_{\text{sub A}} \). Case two \( T \) is censored by \( C \), so we observe \( A \), we just know \( TA \).
251 00:13:27.660 --&gt; 00:13:29.390 is going to be greater than C.
252 00:13:29.390 --&gt; 00:13:30.502 Case three,
253 00:13:30.502 --&gt; 00:13:32.180 we will observe A,
254 00:13:32.180 --&gt; 00:13:34.710 but we know A is greater than TA.
255 00:13:34.710 --&gt; 00:13:38.170 And case four we don’t observe A, we don’t observe T
256 00:13:38.170 --&gt; 00:13:41.723 but we know A is greater than C and TA is greater than C.
257 00:13:42.850 --&gt; 00:13:43.683 Okay.
258 00:13:43.683 --&gt; 00:13:46.020 So now we propose a structural causal
259 00:13:46.020 --&gt; 00:13:47.840 proportional hazards model
260 00:13:48.970 --&gt; 00:13:52.493 to capture the survival effect of treatment initiation time.
261 00:13:53.520 --&gt; 00:13:55.690 Lambda AT here is a hazard function
262 00:13:55.690 --&gt; 00:13:58.510 for the potential outcome T sub A,
263 00:13:58.510 --&gt; 00:14:01.320 we start from lambda infinity T right here.
264 00:14:01.320 --&gt; 00:14:04.514 This is a reference hazard for T infinity.
265 00:14:04.514 --&gt; 00:14:06.610 So we start from here.
266 00:14:06.610 --&gt; 00:14:10.040 Once the treatment is initiated at A,
267 00:14:10.040 --&gt; 00:14:13.910 there is an instantaneous effect of treatment initiation
268 00:14:13.910 --&gt; 00:14:16.870 captured by the G1 function here,
269 00:14:16.870 --&gt; 00:14:19.990 and the effect of staying on the treatment
270 00:14:19.990 --&gt; 00:14:22.360 at any given time point T,
271 00:14:22.360 --&gt; 00:14:27.000 is captured by the G2 function of ART duration.
272 00:14:27.000 --&gt; 00:14:30.260 And the G3 function here captures the interaction
273 00:14:30.260 --&gt; 00:14:34.323 between treatment initiation and treatment duration.
274 00:14:35.470 --&gt; 00:14:40.470 So we leave this structural model relatively flexible.
275 00:14:40.790 --&gt; 00:14:44.360 First, the reference hazard is left unspecified
and the 3G functions, we also left them unspecified smooth function of treatment initiation time duration and their interaction. So now we can parametrize these three functions using natural cubic splines, and by rewriting the risk function of our structural model, we can use beta this parameter to include the causal effects of ART initiation time on mortality hazard. The problem here now, our goal is to how do we obtain a consistent estimate of beta using observed a data? Once we have obtained that, we could also estimate the causal dose-response of initiation time versus mortality by relating the survival function to the hazard function. We can derive this from our structural model. And now we can also estimate the model-based optimal initiation time that will lead to the maximal survival probability at say 52 weeks after diagnosis.
Okay, how to obtain a consistent estimate of beta.

So first let’s assume if A is randomly allocated and both A and T are observed, then we can write the partial likelihood score function.

And this is a sample average of score function is an unbiased estimator of the expectation under the randomized treatment assignment.

So E sub R here is the expectation under the randomized treatment assignment. So this would be an unbiased estimator function, and solving this unbiased estimating equation would give us a consistent estimator of beta. Now, if A is still randomly allocated, but T can occur before A, so A may be censored by T. In this case, we would need to break the mean of an individual score contribution into two parts.

In one part A is observed. The second part is A is not observed. And then we can apply the law of total expectation to the second part.

The inner expectation would be conditioning on the observed information. Then using this strategy and taking in account the survival hazard structure,
we can revise the estimating equation. And by solving this to obtain a consistent estimate of beta.

In the case of non random allocation of treatment,

then if we want to estimate the causal effect of A on T,

then we would have to make a key assumption, ignore ability assumption.

Essentially the assumption says that the initiation of treatment at any given time T

is sequentially randomized in the sense that as a potential outcome beyond this time

is independent of treatment initiation.

Conditioning on all covariate history up to T.

So with this assumption,

we will be able to use observed data to derive the causal effect.

So say PR is the data distribution under randomized A,

and PO is the data distribution.

And they’re not random allocation of A.

Note that in both settings,

there is a same set of observed data.

And as long as the observed data under PR is absolutely continues with the observed data under PO.

Now we can derive a random-nikodym derivative.

And so Murphy’s 2001 paper developed a version of R-N derivative.
that connects the distribution of the observed data under PR and under PO for discrete time and ordinary GEE score.

Johnson’s 2005 paper extended this version of R-N derivative to continuous time still for ordinary GEE score. In this paper we extended the R-N derivative for time to event setting. So this is a version of R-N derivative for survival data. The reason why we wanted to use R-N derivative is that we can then use it to derive an unbiased estimating equation using some weighted version of the observed data. So we can estimate the causal effect. So now we want to use this R-N derivative for survival data. We want to apply that to Cox score and to derive S rated estimating equation. That’s a little bit more complex than the GEE score, but we can observe that the Cox score can essentially be represented in three averages. The one in blue, the one in orange and the whole average. And each average converges to its expectation. And as I showed earlier, we can always break the expectation into two parts.
In one part A is observed, second part is not observed. For the second part, we can apply the total law of expectation, the law of total expectation, and recognizing the survival structure to derive the second part. And then we can apply the R-N derivative for survival data to each piece separately, to construct the unbiased score equation. So after some derivation, we would arrive at the weights and actually the weights come down in a very neat form. Essentially, it suggests that for patients who have initiated treatment by time T, we would weight them by the marginals density function of A divided by the conditional density of A given their covariate history after time T. And for those who are censored, so not initiated by the time T, we would weight them by some survival function of the treatment initiation process. And then by applying this weighting scheme, we will be able to derive a weighted estimating equation. And just a note that we have to apply the same weighting scheme to the people who are still in the risk set at any time T.
And so now that said, previously we have assumed there's no censoring. Now with censoring, we need to assume another similar assumption, similar to the ignore ability assumption, and then using the similar strategy to derive another set of weight for censoring. For those who stay, remain in the study, we would weight them by the survival function for censoring. Censoring contributes information about the parameter in two ways, FC is observed as the person is actually censored. It contributes to the risk set up to C. If C is not observed, so C could be censored by T. If death's occurred, then it contributes to the individual partial likelihood to weight for C but evaluated at death time. Okay, now we know how to weight. Back to the four censoring patterns. The first one, both A and T are observed. We would weight them by the first set of weight for A evaluated at A, T occurred, so the weight for C but evaluated at T.
Second case, $T$ is not observed, $A$ is observed.
So first set of weight for $A$ evaluated at $A$ and $C$ just contributes information to the risks set.

Third case, $A$ is not observed, so second weight for $A$ evaluated at $T$.
And weight for $C$, censoring evaluated at $T$.
The fourth case or final case, $A$ is not observed, again, second set of weight for $A$, but evaluated at $C$, and $C$ also contributes to the risks set.
Okay, so now we know how to weight.
We would have to estimate the weights.
The approach we used in the paper is that we model the intensity processes associated with the two counting processes, one for $A$, one for $C$.
And then when we fit Cox proportional hazards models for the two intensity processes, we use fitted hazard to estimate the weights.
We use empirical cumulative hazards to estimate the conditional density and function.
And for the marginal density function, we use some nonparametric Nelson-Aalen estimator, and use similar fashion to estimate rates for censoring.
Then we apply our methods to the AMPATH data.
AMPATH is a large HIV care program based in West Kenya, our data has almost 5,000 patients and for covariates, we have demographic information and some disease-specific information. Some of them are time varying like, weight, the CD4, these are time varying variables. We categorize the baseline CD4 subgroups into two groups, the less than, or below 50 group, this is the highest risk group. So CD4 the higher, the better. And between 200 and 350, there’s relatively healthy patients. The reason we categorize them into three groups is because the program guidelines are based on these subgroups and RCT is reported results for below 50 group. We want to compare our results to our CT findings. So this plot shows the three estimated G functions. The G1 A here suggests that the instantaneous effect of a treatment initiation has a U shape, achieving maximum benefit, or the lowest mortality hazard at just about 10 weeks.
And after that, the longer the treatment is delayed, the less the benefit of the treatment initiation. And this is the effect of duration, in general, it says that the longer you stay on the treatment, the more benefit you get. There’s an upward trend for the interaction effect. Essentially suggesting that delayed treatment initiation would reduce the benefit associated with long ART duration. And so the net causal effect of treatment initiation is summarized in this plot. Top panel shows the mortality rate at one year versus treatment initiation time. Bottom panel compares immediate initiation versus delayed initiation at A. So we can see that the benefit of early initiation is most pronounced for the CD4 below 50 group, or the highest risk group. And the curves here are pretty flat, suggesting that there’s not much benefit of early initiation for relatively healthy patients. Several advantages for this approach. It’s easy to get optimal initiation time based on the model outputs. And we could also use the model outputs.
to emulate comparisons between regimens reported in RCTs.

So we could mimic random allocation of treatment initiation time to specific intervals by assuming a distribution for treatment initiation time A, that is independent of covariates and outcome and compare interval specific mortality rates and draw inferences about treatment initiation.

But with the continuous time marginal structure model, we'll also be able to conduct a higher resolution analysis that can potentially generate new insights in relation to a randomized control trial. For the sake of timing, I just gonna briefly talk about the simulation. We conduct simulation to examine the finite-sample properties of weighted estimators, we evaluate sensitivity of our estimators to the violations of the ignore ability, or no unmeasured confounding assumption, but we only considered confounding at baseline. So the sensitivity analysis strategy for time-varying confounding, especially with the censored survival outcome is kind of very complex topic, and we were still working on this project right now,
In this paper we just consider confounding at baseline.

Under random allocation of treatment, our estimator produced a new zero bias and nominal coverage probability, in the presence of measured confounding, it eliminated nearly all the biases and provided close to nominal coverage probability, but in the presence of unmeasured confounding, there was bias in our estimator. And the biases were in proportion to the degree of measured confounding.

Okay, so moving to the second example, this is a continuous time dynamic treatment regimen of the form, initiate treatment when a biomarker crosses a threshold. It’s dynamic treatment regimen because it depends on evolving history of treatment and a tailoring variable. So in our case, CD4 is a tailoring variable. That means we make our treatment decision based on this variable. A little bit different from our previous motivating example. The outcome interest is different. This is a pediatric data. So for the kids, the mortality rate is very low.
and our data I think it’s around 3%. And for kids, we’re also interested in their CD4 measurements, because CD4 is important marker of immune system function.

557 00:30:21.300 --> 00:30:24.260 and both outcomes, both mortality rate and CD4.

558 00:30:24.260 --> 00:30:26.333 are sparsely measured in our data, but we are interested in both.

560 00:30:29.620 --> 00:30:32.790 Other than that, we also have complications posed by observational data.

562 00:30:36.250 --> 00:30:41.250 So this is a picture of nine randomly selected individuals from our data,

564 00:30:42.670 --> 00:30:45.900 X axis here, follow-up time in days,

565 00:30:45.900 --> 00:30:49.440 Y axis here square root of CD4,

566 00:30:49.440 --> 00:30:53.220 purple line is end of follow-up,

567 00:30:53.220 --> 00:30:56.950 two gray lines here mark one year

568 00:30:56.950 --> 00:30:59.023 and two years post diagnosis.

569 00:31:00.270 --> 00:31:03.950 Empty circles here mean that the patient has not been treated.

570 00:31:03.950 --> 00:31:06.010 has not been treated.

571 00:31:06.010 --> 00:31:09.310 Solid circles, mean that they’re on the treatment.

572 00:31:09.310 --> 00:31:11.920 So we can see that there’s a lot of variability in terms of the treatment initiation time.

574 00:31:15.940 --> 00:31:19.620 And some people are followed much longer than some other patients.

576 00:31:22.290 --> 00:31:27.290 And the follow-up time is pretty irregularly spaced.

577 00:31:29.370 --> 00:31:33.880 and overall the CD4 measurements are quite sparse,

578 00:31:33.880 --> 00:31:36.440 and there’s also incomplete information for example, these two they either died
or were lost to follow up
before they even got a chance to be treated.
So there’s also a lot of complication in the
data.
There’s a continuous time measurement
of the treatment initiation.
It just happens all over the place.
The longitudinal outcome of interest are
sparsely measured,
leading to incomplete data.
There’s also a censoring due to dropout or
deaths.
So our general solution is that we’ll use weight-
ing to handle time-varying confounding.
And will show how to derive a continuous
time versions
of the weights.
For the missing outcomes
that is caused by sparse measurement and
censoring
we’ll use imputations from a model of the
joint distribution
of CD4 and mortality.
And because we’re interested in both mortality
status
and CD4, we’ll develop a composite outcome.
So our general approach is to emulate a ran-
domized trial
in which we would randomize individuals
to follow specific DTR Q.
And Q equals zero means never treated,
because CD4 can never drop below zero.

580 00:31:41.490 --> 00:31:43.830 or were lost to follow up
581 00:31:43.830 --> 00:31:47.410 before they even got a chance to be treated.
582 00:31:47.410 --> 00:31:51.382 So there’s also a lot of complication in the
data.
583 00:31:51.382 --> 00:31:53.640 There’s a continuous time measurement
584 00:31:53.640 --> 00:31:55.380 of the treatment initiation.
585 00:31:55.380 --> 00:31:57.725 It just happens all over the place.
586 00:31:57.725 --> 00:32:02.600 The longitudinal outcome of interest are
sparsely measured,
587 00:32:02.600 --> 00:32:04.710 leading to incomplete data.
588 00:32:04.710 --> 00:32:08.570 There’s also a censoring due to dropout or
deaths.
589 00:32:08.570 --> 00:32:11.410 So our general solution is that we’ll use weight-
ing
590 00:32:11.410 --> 00:32:13.800 to handle time-varying confounding.
591 00:32:13.800 --> 00:32:16.830 And will show how to derive a continuous
time versions
592 00:32:16.830 --> 00:32:18.820 of the weights.
593 00:32:18.820 --> 00:32:21.400 For the missing outcomes
594 00:32:21.400 --> 00:32:24.470 that is caused by sparse measurement and
censoring
595 00:32:24.470 --> 00:32:27.920 we’ll use imputations from a model of the
joint distribution
596 00:32:27.920 --> 00:32:30.150 of CD4 and mortality.
597 00:32:30.150 --> 00:32:33.200 And because we’re interested in both mortality
status
598 00:32:33.200 --> 00:32:36.363 and CD4, we’ll develop a composite outcome.
599 00:32:37.970 --> 00:32:42.460 So our general approach is to emulate a ran-
domized trial
600 00:32:42.460 --> 00:32:45.000 in which we would randomize individuals
601 00:32:45.000 --> 00:32:47.900 to follow specific DTR Q.
602 00:32:47.900 --> 00:32:50.950 And Q equals zero means never treated,
603 00:32:50.950 --> 00:32:54.170 because CD4 can never drop below zero.
Now, Q equals infinity means treat immediately.

So after randomization,

all the individuals will be followed for a fixed amount of time,

at which point, say T star,

both their mortality status.

And among those who are alive at T star,

their CD4 count will be assessed.

So what define a composite outcome XQ,

that is the product of the test indicator and the potential CD4.

So the cumulative distribution of this composite outcome is a useful measure of treatment utility,

because it has appointments at zero corresponding to mortality rate.

Thereby capturing both mortality status and CD4 count among survivors at T star.

So for example, the probability of a positive XQ,

that’s the survival fraction,

and the probability of XQ greater than X,

that’s the fraction of survivors with CD4 above X.

Okay, so similar to the first motivating example,

we again have three timed events.

Death time, censoring time, treatment initiation time.

And now we have a tailoring variable, CD4 count.
So the CD four process is defined for all continuous time, but it’s just measured at discrete times. And we also have a P by one covariate process. Using a convention in the DTR literature, we assume that the treatment decision is always made after observing the covariate history and the CD4 count.

Putting everything together, we have a history information indicator. For each individual, we’ll have an observed data process. And just note that each person can have a different lens of followup at different time points.

Our goal is to evaluate the effect of DTRs, but we’re dealing with observational data, so we’ll have to map the observed treatment regimen to specific DTRs that we are interested in evaluating.

Essentially we’ll follow the deterministic function to create the mapping. Essentially there are three rules. First rule says not to treat the person if the person has not yet initiated treatment and their CD4 has not fallen below Q, or has not been observed. Second rule says, treat this person if their time T, CD4 has fallen below Q for the very first time.
Once treated, always treat them. Following these three rules, we'll be able to create a regimen specific compliant process for each individual in the data. So essentially if the rule says treat, and if the person is actually treated by the time T, then this person is compliant at time T. If the rule says do not treat, and the person was not treated at the time T, then this person is still compliant to the rule. And so we'll be able to observe a compliant process for each person. Here a simple example to show you how to create the mapping. For example, we're interested in Q equals 350. This person came in at baseline, had a measurement 400 above the threshold. The rule says do not treat, the person was not treated, still compliant at this point. Third visit, the person’s CD4 drops to 330, which is below the threshold for the very first time, the rules are start treating this person, the person was actually treated.
So compliant at this point.

Next visit the rule says once treated always treat them.

the person kept being treated.

So this person was compliant with the rule all throughout his or her followup.

Next example, the first two rows are the same.

The third visit, the person’s CD4 jumps to 450, which is above the threshold.

The rule says do not treat, but on the contrary, the person was actually treated.

and kept being treated.

So from this time point onward,

the person was not compliant with this rule.

Okay, so that’s just some simple example to show how to create the mapping.

With missing outcomes for those alive at the target measurement time $T^*$, the observed outcome $XI$ is the CD4 measurement at $T^*$.

But because of CD4 is sparsely measured and irregularly spaced, Z of $T^*$ is directly observed only when the person’s followup time is exactly at $T^*$.

So in this case, it is pretty common to redefine a interval and capture the CD4 that is measured at the time closest to the target measurement time.
But even using this strategy, there's still a possibility that there is no measurement in predefined interval. Then we say this person has a missing outcome. And it's also possible that the person dropped out before TA. And so in this case, the outcome is also missing. For these missing outcomes, our general strategy is to use multiple imputation. So we would specify and fit model for the joint distribution of the CD4 process and the mortality process. For those known to be alive, but without a CD4 measurement, we would impute the CD4 count from the fitted CD4 sub-model. And for those missing the CD4, because of right censoring, we would calculate the mortality probability from the fitted survival sub-model, and then impute the death indicator with this calculated probability. If the death indicator was imputed to be zero, then we further impute a CD4 count for this person. Otherwise we'll set X to be zero. And again, we would have to assume
some standard causal inference assumptions in order to draw causal effects about the DTRQ using observational data. And we can estimate and compare DTRs along a continuum. We can formulate a causal model for the smooth effect of Q on the task quantile of XQ. This is our composite outcome with separate parameters capturing the effect of treat immediately, and the effect of never treat. And then we can parametrize the model using splines of Q for the third term here, to gain statistical efficiency. And we can obtain a consistent estimator of effect of Q by solving the weighted quantile regression estimating equation. So what should be the weights? First, we assume there's no dropout or death prior to the target measurement time. In the discrete time setting with common time point, the form of the weights have already been done. It has been derived in several papers. Essentially, the denominator of the weight is this conditional probability. It's a conditional probability of the person being compliant.
all throughout the follow up, given the covariate history.

So if we have a common set of discrete time points,
it’s a cumulative a product of the conditional probability
of this person being compliant at every time point.
And essentially if the rule says treat,
it’s a condition of probability of the person
actually being treated at this time point,
if the rule says not treat,
as a conditional probability of this person not treated
by this time point.
So in order to estimate this probability,
we just need to model the observed
treatment initiation process among those regimen compliers,
but this is for discrete time setting.
What would be the continuous time weights?
We note that the occurrence of treatment initiation
in a small time interval T and T plus TD
is actually a Bernoulli trial with outcome DNA of T.
So then we can rewrite this probability,
this probability here,
in the form of individual partial likelihood
for the counting process of A.
And now we note that when DT becomes smaller and smaller,
this finite product approaches a product integral.
So then this finite product can be rewritten as a final product over jump times of the counting process for \( A \) times the survival function. And then by recognizing that each individual had at most one jump at exactly \( A_i \). Now we can further reduce this probability to this form. Which suggests weighting scheme. Essentially it says for those who have been treated by a \( T \) star, we would weight them by the conditional density function of \( A \). For those who haven’t been treated by the time \( T \) star, we would weight them by the survival or function of \( A \). So if you recall the weighting scheme for the first motivating example, this is exactly the same, the same rating scheme, but we took different approaches. The first example, we use a random Aalen derivatives to derive the weighting scheme. The second project we derive the limit of the finite product, but using different approaches, we arrive at the same weighting scheme. And so similarly we modeled the intensity process of treatment initiation. We estimate the weights.
So if there was a censoring or death prior to target measurement time, we would have to assume once lost to follow up at a time prior to T star, the treatment and regimen status remain constant.

And this way we will just estimate the weights up to a time point CI, and if the person died before T star, then we would only evaluate compliance and treatment initiation processes up to time TI.

Okay, so for missing outcomes, we propose a joint modeling approach. We specify a two-level model for the observed CD4 process. The first level, the observed CD4 process is a true CD4 trajectory plus some arrow process. The second level, we relate the true CD4 trajectory to baseline characteristics and treatment initiation time, and some subject specific random effects, capturing subject-specific deviations from the mean trajectories.

And now we propose a hazard model for deaths uses the true CD4 trajectory as a covariate linking the two processes, linking the death process and linking with a CD4 process.
Now we use the joint model to impute the missing outcomes and estimate the variance of the target estimator using Rubin’s combination wall. So we applied this method to the IeDEA dataset. IeDEA is another HIV consortium based in West Kenya. So we have almost 2000 data. We see that the CD4 is pretty sparsely measured and death rate is low around three and 4%. Most of patients have been treated by one year and we have a set of covariates. Some of them are time varying, some of them are time fixed. We proposed three target estimators, so first we’re interested in mortality proportion. We’re also interested in the median of the distribution of the composite outcome XQ. We also looked at CD4 among survivors, but this estimator does not have a causal interpretation because it conditions on having survived two T star. So it only measures association, but the first two estimators have causal interpretations. So we first look at the effectiveness of five specific regimens for both one year and two years.
after diagnosis.

We can see that the immediate treatment initiation lead to significant lower mortality rate and significantly higher median values of the composite alcohol compared to delayed treatment.

And the never treat initiation will lead to a significantly higher mortality probability.

And for those who do survive to T star, their CD4 count is higher.

So resulting higher to theta Q2 and higher theta Q3 compared to other delayed treatment regimen.

So this may suggest that those who do survive to T-star without any treatment, maybe they are relatively healthier at the beginning of the followup.

Okay, and then we also plot the dose response curve of the median value of the composite outcome versus DTR Q,

also suggests that the immediate treatment also would lead to significantly higher median values of XQ,

and also as illustration

of the gained statistical efficiency

by modeling the smooth effect Q on the quantile of the XQ.

The variance in the one year outcome associated with Q equals 350, achieved about 15% reduction
compared to that from the regimen specific estimates.
So we gain a bit of our statistical efficiency by modeling the smooth effect.
So there are several strands of continuous time marginal structure model.
We see that we can derive, using different approaches,
closed form of weights for continuous-time treatment.
It can handle complex dataset on its own terms
without having to artificially align measurement times,
which could possibly lead to loss of information.
It is amenable to many different outcomes.
We’ve used the survival outcomes,
we’ve used composite outcomes.
You can also handle many data complications introduced by various censoring patterns.
within the same marginal structure model.
So these are the strengths,
but there are also limitations with this approach of course.
One notable limitation is extreme ways, which could possibly lead to unstable estimates.
So how to address this issue,
especially for time varying confounding.
with censored outcome, this would be a challenging task,
but if we can solve this issue,
it might be a very important contribution to the field.
So this is something my colleagues and I have been thinking about and working on for some time.
Another limitation is that we know that weighting-based estimator is less efficient than the so-called G methods.
The G computation, G estimation, and both G methods require integrating over the space of longitudinal confounders.
So the G methods are computationally much, much more expensive than the marginal structure model-based methods.
And as far as I know, currently there’s no continuous time version of the G computation methods.
Judith Lok has a paper, back in 2008. She developed theory for continuous time G-estimation,
but I have yet to see a practical implementation of this method.
So this could be another avenue for future research.
how to increase efficiency of the continuous time weighting-based methods.
And here’s some key references.
Thank you.
Thank you Liang yuan for this very interesting
and comprehensive presentation.

Let’s see if we have any questions from the audience.

If there’s any questions, please feel free to unmute yourself and speak or type in the chat.

- [Donna] Thanks, it was a very interesting talk.

This is Donna Spiegelman.

- Hi, Donna.

Yeah, hi.

I was wondering I might’ve missed it, but did you say much about estimating the variance?

I see you have (indistinct) around the curve, so you must derive the variance.

So I’m wondering if you could say a little bit about that or a little more about that if I missed what you did say.

Sure, sure, sure.

So for this one, this is the second example, for this one we have multiple amputation and we also have weighting.

So with weighting part, the variance was estimated using bootstrap.

So then we use the Rubin’s combination role to estimate the total variance.
For the first example, we actually used a bootstrap, and the coverage probability was actually okay. It's good for the estimator. Did you think about asymptotic variants derivations? - I did. It was a very difficult task, there's a story about our first paper found that about it. It was first submitted to Jaza and then they asked about the asymptotic variants about the estimator. And it's quite complex because they involve the splice and involves the survival data. And we have already approved as a consistency, and it also involves optimization. So it's just comes to - What's the optimization piece. Oh, it's the model based optimal treatment initiation time that will lead to the maximum survival at predefined time points. Right, so they are interested in the optimization. So the inference about the optimized treatment initiation time. We did some empirical evidence for like the largest sample convergence rate,
979 00:54:08.217 --> 00:54:13.217 but we weren’t successful at deriving asymptotic variants.
980 00:54:14.450 --> 00:54:18.170 So that’s another piece, I think maybe,
981 00:54:18.170 --> 00:54:19.003 I don’t know.
982 00:54:19.003 --> 00:54:21.280 We had this discussion among colleagues
983 00:54:21.280 --> 00:54:23.720 and also my advisor at the time,
984 00:54:23.720 --> 00:54:27.840 we just not sure about whether it’s worth the effort
985 00:54:27.840 --> 00:54:29.573 to go and do that route.
986 00:54:30.420 --> 00:54:32.220 - It’s probably way more complex
987 00:54:32.220 --> 00:54:34.130 than just the usual derivation.
988 00:54:34.130 --> 00:54:36.470 ‘Cause you do have like two weighting models,
989 00:54:36.470 --> 00:54:39.590 which are also survival models,
990 00:54:39.590 --> 00:54:41.900 and also the derivation that these variances
991 00:54:41.900 --> 00:54:44.440 sometimes can be specific to the choice
992 00:54:44.440 --> 00:54:45.800 of these (indistinct) models.
993 00:54:45.800 --> 00:54:48.590 And so if you have a variance and the cup’s model,
994 00:54:48.590 --> 00:54:51.696 it does not apply to other forms of models,
995 00:54:51.696 --> 00:54:54.290 I guess it’s really a trade-off right?
996 00:54:54.290 --> 00:54:57.343 - Yeah, it is a trade off.
997 00:54:58.420 --> 00:55:03.330 It’s still an open question and nobody had done it yet,
998 00:55:03.330 --> 00:55:06.240 but just, whether you’re thinking it’s was the effort
999 00:55:06.240 --> 00:55:10.020 just to devote a couple of years to work on that.
1000 00:55:10.020 --> 00:55:14.780 - So was bootstrap time consuming for these datasets,
1001 00:55:14.780 --> 00:55:18.211 for this data analysis, or they’re pretty manageable.
1002 00:55:18.211 --> 00:55:19.780 - They’re pretty manageable.
1003 00:55:19.780 --> 00:55:23.070 And it looks complicated because we have to weight everybody
We also have to weight everywhere in the risk set at any time point. So it looks pretty complex, but still manageable. Another reason is because we use parametric models. If we wanted to, I'm not aware of any machine learning algorithm that can handle survival data, but also with time varying covariates, that's something I'm also thinking about. Like, if we use those algorithm might be more time consuming, but with just a parametric models, it's pretty manageable. And when you're bootstrapped, you go back to the weight models and refit the weight models every time? Yeah. But the variable is pre-determined. So that's what you mentioned, machine learning. So the variables are predetermined and they're functional forms in the model, but the coefficients that correspond to them are re estimated for each bootstrap. Very estimated. Right, right. Exactly. Yeah.
Great question.

Yeah.

So a lot of open questions still.

So any other questions from the audience?

I have another comment.

So by getting back to this,

that you're estimated the coefficients

for the weight models.

So in sort of the standard marginal structural model,

the variability due to those weight models is ignored.

And the robust variance is used

and said to be an overestimate,

implying that if you took that variation into account,

you'd get a smaller variance

and you might see the same thing here with your bootstraps.

If you took the weight models as fixed,

you might find that you have a less efficient estimator,

which is kind of interesting

just in terms of say a methods paper to show,

because there's different ways to do bootstraps,

but here you're automatically taking the estimation

of the weight models into account,

which is not saying that say the classic paper

by Hernan in epidemiology, that's ignored and the robust variance is recommended.
1056 00:57:42.310 --> 00:57:43.460 - Hmm.
1057 00:57:43.460 --> 00:57:46.090 It’s a very great comment.
1058 00:57:46.090 --> 00:57:47.480 Something I have to think about.
1059 00:57:47.480 --> 00:57:51.860 So you’re saying that in each bootstrap,
1060 00:57:51.860 --> 00:57:54.810 when we estimate the weight model, we fix
the weight model.
1061 00:57:56.903 --> 00:57:59.763 So the coefficients from the weight model
stay fixed-
1062 00:58:00.850 --> 00:58:02.810 - Yeah, so you don’t even do a bootstrap for
that.
1063 00:58:02.810 --> 00:58:06.465 You basically hold the weight model as a
constant,
1064 00:58:06.465 --> 00:58:07.298 and then you’d-
1065 00:58:07.298 --> 00:58:08.570 - Robust variance.
1066 00:58:08.570 --> 00:58:10.950 - Yeah, you use the robust variance,
1067 00:58:10.950 --> 00:58:12.690 which I guess it’s a little tricky
1068 00:58:12.690 --> 00:58:14.520 because now you don’t have the robust vari-
ance
1069 00:58:14.520 --> 00:58:15.776 because you’re not using it,
1070 00:58:15.776 --> 00:58:20.776 but it seems the bootstrap analog of the
approach taken
1071 00:58:20.870 --> 00:58:24.230 would be to just fit the weight model once,
1072 00:58:24.230 --> 00:58:26.870 treat that fixed unknown,
1073 00:58:26.870 --> 00:58:31.190 and then only bootstrap on the outcome
model.
1074 00:58:31.190 --> 00:58:32.260 - Right, right.
1075 00:58:32.260 --> 00:58:33.093 Yeah.
1076 00:58:33.093 --> 00:58:34.618 - [Fan Li] Totally. Yeah.
1077 00:58:34.618 --> 00:58:36.170 - Interesting.
1078 00:58:36.170 --> 00:58:37.753 Take that in as a note.
1079 00:58:39.300 --> 00:58:41.590 - So I do have a question as well.
1080 00:58:41.590 --> 00:58:44.200 I think Liangyuan you had presented two
applications
At the HIV observational studies, do you see the application that these new methods to other areas as well to solve the other questions? - Yeah. Yeah, actually this is not pertaining to HIV area. It's actually in the public health areas. A lot of questions are involving this statistical formulation. So for example, I've been collaborating with an epidemiologist at Columbia. They are doing cardiovascular research. So one research question is that, I think it’s blood pressure lowering intervention. So blood lowering innovation is very useful for preventing cardiovascular diseases, but they don’t know. And there also a lack of randomized control trials. What is the optimal threshold to start giving the blood lowering treatment? So this is exactly the same form as our second motivating example. Like what is the optimal CD4 threshold to start the HIV treatment? And their question is what is the optimal threshold to start the blood lowering treatment? So I think there’s a lot of possibility as to apply these kinds of methods
in other health research area.
- Yeah, it’s a huge controversy in terms of the treatment of hypertension,
- what’s the optimal blood pressure to start antihypertensives.
- And I think there was a very large trial that showed that it was better to start it at a much earlier threshold than what current practices.
- And it’s very troublesome for people around the world because these medicines are expensive.
- And if you see now, like another like 40% of the population should now be initiated a antihypertensive medication,
- well, most countries can’t even afford that.
- So the implications of these different thresholds is a very big topic of sort of substantive research and debate right now.
- Well, that’s great to know, there’s urgent need for that.
- (indistinct) I think we are at the hour, so thanks Liangyuan again for your great presentation
- All right, I think we have any questions,
- I’m sure Liangyuan is happy to take any questions offline
- by emails.
And I think this is the final seminar of our fall series,
and I hope to see everyone next spring,
have a good holiday.
Thank you.
- Thank you.
- Bye.
- Bye.