DAVID SONTAG: A three-part lecture today, and I'm still continuing on the theme of reinforcement learning. Part one, I'm going to be speaking, and I'll be following up on last week's discussion about causal inference and Tuesday's discussion on reinforcement learning. And I'll be going into sort of one more subtlety that arises there and where we can develop some nice mathematical methods to help with.

And then I'm going to turn over the show to Barbra, who I'll formally introduce when the time comes. And she's going to both talk about some of her work on developing and evaluating dynamic treatment regimes, and then she will lead a discussion on the sepsis paper, which was required reading from today's class. So those are the three parts of today's lecture.

So I want you to return back, put yourself back in the mindset of Tuesday's lecture where we talked about reinforcement learning. Now, remember that the goal of reinforcement learning was to optimize some reward.

Specifically, our goal is to find some policy, which I can note as pi star, which is the arg max over all possible policies pi of v of pi, where just to remind you, v of pi is the value of the policy pi. Formally, it's defined as the expectation of the sum of the rewards across time.

So the reason why I'm calling this an expectation with like the pi is because there's stochasticity both in the environment, and possibly pi is going to be a stochastic policy. And this is summing over the time steps, because this is not just a single time step problem.

But we're going to be considering interventions across time of the reward at each point in time. And that reward function could either be at each point in time or you might imagine that this is 0 for all time steps, except for the last time step.

So the first question I want us to think about is, well, what are the implications of this as a learning paradigm? If we look what's going on over here, hidden in my story is also an expectation over x, the patient, for example, or the initial state. And so this intuitively is saying, let's try to find a policy that has high expected reward, average [INAUDIBLE] over all patients.

And I just want you to think about whether that is indeed the right goal. Can anyone think about a setting where that might not be desirable? Yeah.

AUDIENCE: What if the reward is the patient living or dying? You don't want it to have high ratings like saving two patients and [INAUDIBLE] and expect the same [INAUDIBLE].

DAVID SONTAG: So what happens if this reward is something mission critical like a patient dying? You really want to try to avoid that from happening as much as possible. Of course, there are other criteria that we might be interested in as well.

And both in Frederick's lecture on Tuesday and in the readings, we talked about how there might be other aspects about making sure that a patient is not just alive but also healthy, which might play into your reward functions. And there might be rewards associated with those.

And if you were to just, for example, put a positive or negative infinity for a patient dying, that's a nonstarter, right, because if you did that, unfortunately in this world, we're not always going to be able to keep patients alive. And so you're going to get into an infeasible optimization problem. So minus infinity is not an option.

We're going to have to put some number to it in this type of approach. But then you're going to start trading off between patients. In some cases, you might have a very high reward for-- there are two different solutions that you might imagine, one solution where the reward is somewhat balanced across patients and another situation where you have really small values of reward for some patients and a few patients with very large values and rewards.

And both of them could be the same average, obviously. But both are not necessarily equally useful. We might want to say that we prefer to avoid that worst-case situation.

So one could imagine other ways of formulating this optimization problem, like maybe you want to control the worst-case reward instead of the average-case reward. Or maybe you want to say something about different quartiles. I just wanted to point that out, because really that's the starting place for a lot of the work that we're doing here.

So now I want us to think through, OK, returning back to this goal, we've done our policy iteration or we've done our Q learning, that is, and we get a policy out. And we might now want to know what is the value of that policy? So what is our estimate of that quantity?

Well, to get that, one could just try to read it off from the results of Q learning by just computing that the pi-- what I'm calling v pi hat-- the estimate is just equal to now a maximum over actions a of your Q function evaluated at whatever your initial state is and the optimal

choice of action a.

So all I'm saying here is that the last step of the algorithm might be to ask, well, what is the expected reward of this policy? And if you remember, the Q learning algorithm is, in essence, a dynamic programming algorithm working its way from the sort of large values of time up to the present. And it is indeed actually computing this expected value that you're interested in. So you could just read it off from the Q values at the very end.

But I want to point out that here there's an implicit policy built in. So I'm going to compare this in just a second to what happens under the causal inference scenario. So just a single time step in potential outcomes framework that we're used to.

Notice that the value of this policy, the reason why it's a function of pi is because the value is a function of every subsequent action that you're taking as well. And so now let's just compare that for a second to what happens in the potential outcomes framework. So there, our starting place-- so now I'm going to turn our attention for just one moment from reinforcement learning now back to just causal inference.

In reinforcement learning, we talked about policies. How do we find policies to do well in terms of some expected reward of this policy? But yet when we were talking about causal inference, we only used words like average treatment effect or conditional average treatment effect, where for example, to estimate the conditional average treatment effect, what we said is we're going to first learn, if we use a covariate adjustment approach, we learn some function f of x comma t, which is intended to be an approximation of the expected value of your outcome y given x comma-- I'll say y of t. There. So that notation.

So the goal of covariate adjustment was to estimate this quantity. And we could use that then to try to construct a policy. For example, you could think about the policy pi of x, which simply looks to see is-- we'll say it's 1 if CATE or your estimate of CATE for x is positive and 0 otherwise. Just remind you, the way that we got the estimate of CATE for an individual x was just by looking at f of x comma 1 minus f of x comma 0.

So if we have a policy-- so now we're going to start thinking about policies in the context of causal inference, just like we were doing in reinforcement learning. And I want us to think through what would the analogous value of the policy be? How good is that policy? It could be another policy, but right now I'm assuming I'm just going to focus on this policy that I show up

here.

Well, one approach to try to evaluate how good that policy is, is exactly analogous to what we did in reinforcement learning. In essence, what we're going to say is we evaluate the quality of the policy by summing over your empirical data of pi of xi. So this is going to be 1 if the policy says to give treatment 1 to individual xi.

In that case, we say that the value is f of x comma 1. Or if you gave the second-- if the policy would give treatment 0, the value of the policy on that individual is 1 minus pi of x times f of x comma 0.

So I'm going to call this sort of an empirical estimate of what you should think about as the reward for a policy pi. And it's exactly analogous to the estimate of v of pie that you would get from a reinforcement learning context. But now we're talking about policies explicitly.

So let's try to dig down a little bit deeper and think about what this is actually saying. Imagine the story where you just have a single covariate x. We'll think about x as being, let's say, the patient's age. And unfortunately there's just one color here. But I'll do my best with that. And imagine that the potential outcome y0 as a function of the patient's age x looks like this.

Now imagine that the other potential outcome y1 looked like that. So I'll call this the y1 potential outcome. Suppose now that the policy that we're defining is this. So we're going to give treatment one if the condition of our treatment effect is positive and 0 otherwise.

I want everyone to draw what the value of that policy is on a piece of paper. It's going to be-- I'm sorry-- I want everyone to write on a piece of paper what the value of the policy would be for each individual. So it's going to be a function of x.

And now I want it to be-- I'm looking for y of pi of x. So I'm looking for you to draw that plot. And feel free to talk to your neighbor. In fact, I encourage you to talk to your neighbor.

[SIDE CONVERSATION]

Just to try to connect this a little bit better to what I have up here, I'm going to assume that f-- this is f of x1, and this is f of x0. All right. Any guesses? What does this plot look like? Someone who hasn't spoken in the last one week and a half, if possible. Yeah?

AUDIENCE: Does it take like the max of the functions at all point, like, it would be y0 up until they intersect

and then y1 afterward?

DAVID SONTAG: So it would be something like this until the intersection point.

AUDIENCE: Yeah.

DAVID SONTAG: And then like that afterwards. Yeah. That's exactly what I'm going for. And let's try to think through why is that the value of the policy? Well, here the CATE, which is looking at a difference between these two lines as negative-- so for every x up to this crossing point, the policy that we've defined over there is going to perform action-- wait. Am I drawing this correctly?

Maybe it's actually the opposite, right? This should be doing action one. Here. OK. So here the CATE is negative. And so by my definition, the action performed is action 0. And so the value of the policy is actually this one.

[INTERPOSING VOICES]

DAVID SONTAG: Oh. Wait. Oh, good. [INAUDIBLE]. Because this is the graph I have in my notes. Oh, good. OK. I was getting worried. OK. So it's this action, all the way up until you get over here. And then over here, now the CATE suddenly becomes positive. And so the action chosen is 1. And so the value of that policy is y1.

So one could write this a little bit differently for-- in the case of just two policies, and now I'm going to write this in a way that it's really clear. In the case of just two actions, one could write this equivalently as an average over the data points of the maximum of fx comma 0 and f of x comma 1.

And this simplification turning this formula into this formula is making the assumption that the pi that we're being evaluated on is precisely this pi. So this simplification is only for that pi. For another policy, which is not looking at CATE or for example, which might threshold CATE at a gamma, it wouldn't quite be this. It would be something else.

But I've gone a step further here. So what I've shown you right here is not the average value but sort of individual values. I have shown you the max function. But what this is actually looking at is the expected reward, which is now averaging across all x.

So to truly draw a connection between this plot we're drawing and the average reward of that

policy, what we should be looking at is the average of these two functions, which is we'll say something like that. And that value is the expected reward.

Now, this all goes to show that the expected reward of this policy is not a quantity that we've considered in the previous lectures, at least not in the previous lectures in causal inference. This is not the same as the average treatment effect, for example.

So I've just given you one way to think through, number one, what is the policy that you might want to derive when you're doing causal inference? And number two, what is one way to estimate the value of that policy, which goes through the process of estimating potential outcomes via covariate adjustment?

But we might wonder, just like when we talked about in causal inference where I said there are two approaches or more than two, but we focused on two, using covariate adjustment and doing inverse propensity score weighting, you might wonder is there another approach to this problem all together? Is there an approach which wouldn't have had to go through estimating the potential outcomes? And that's what I'll spend the rest of this third of the lecture focused talking about.

And so to help you page this back in, remember that we derived in last Thursday's lecture an estimator for the average treatment effect, which was 1 over n times the sum over data points that got treatment 1 of yi, the observed outcome for that data point, divided by the propensity score, which I'm just going to write as ei.

So ei is equal to the probability of observing t equals 1 given the data point xi minus a sum over data point i such that ti equals 0 of yi divided by 1 minus ei.

And by the way, there was a lot of confusion in class why do I have a 1 over n here, a 1 over n here, but right now I just took it out all together, and not 1 over the number of positive points and 1 over the number of 0 data points. And I expanded the derivation that I gave in class, and I posted new slides online after class. So if you're curious about that, go to those slides and look at the derivation.

So in a very analogous way now, I'm going to give you a new estimator for this same quantity that I had over here, the expected reward of a policy. Notice that this estimator here, it made sense for any policy. It didn't have to be the policy which looked at, is CATE just greater than 0 or not? This held for any policy. The simplification I gave was only in this particular setting.

I'm going to give you now another estimator for the average value of a policy, which doesn't go through estimating potential outcomes at all. Analogous to this is just going to make use of the propensity scores. And I'll call it R hat.

Now I'm going to put a superscript IPW for inverse propensity weighted. And it's a function of pi, and it's given to you by the following formula-- 1 over n sum over the data points of an indicator function for if the treatment, which was actually given to the i-th patient, is equal to what the policy would have done before the i-th patient.

And by the way, here I'm assuming that pi is a deterministic function. So the policy says for this patient, you should do this treatment. So we're going to look at just the data points for which the observed treatment is consistent with what the policy would have done for that patient. And this indicator function is 0 otherwise. And we're going to divide it by the probability of ti given xi.

So the way I'm writing this, by the way, is very general. So this formula will hold for nonbinary treatments as well. And that's one of the really nice things about thinking about policies, which is whereas when talking about average treatment effect, average treatment effect sort of makes sense in the comparative sense, comparing one to another.

But when we talk about how good is a policy, it's not a comparative statement at all. The policy does something for everyone. You could ask, well, what is the average value of the outcomes that you get for those actions that we're taking for those individuals? So that's why I'm writing a slightly more general fashion already here. Times yi obviously.

So this is now a new estimator. I'm not going to derive it for you in class, but the derivation is very similar to what we did last week when we tried to drive the average treatment effect. And the critical point is we're dividing by that propensity score, just like we did over there.

So this, if all of the assumptions made sense, you had infinite data, should give you exactly the same estimate as this. But here, you're not estimating potential outcomes at all. So you never have to try to impute the counterfactuals. Here, all it relies on knowing is that you have the propensity scores for each of the data points in your training set or in a data set.

So for example, this opens the door to tons of new exciting directions. Imagine that you had a very large observational data set. And you learned a policy from it. For example, you might have done covariate adjustment and then said, OK, based on covariate adjustment, this is my new policy.

So you might have gotten it via that approach. Now you want to know how good is that. Well, suppose that you then run a randomized control trial. And then you run a randomized control trial, you have 100 people, maybe 200 people, and so not that many. So not nearly enough people to have actually estimated your policy alone.

You might have needed thousands or millions of individuals to estimate your policy. Now you're only going to have a couple individuals that you could actually afford to do a randomized control trial on.

For those people, because you're flipping a coin for which treatment they're going to get, suppose that were in a binary setting where the only two treatments, then this value is always 1/2 1/2. And what I'm giving you here is going to be an unbiased estimate of how good that policy is, which one can now estimate using that randomized control trial.

Now, this also might lead you to think through the question of, well, rather than estimating the policy through-- rather than obtaining a policy through the lens of optimizing CATE, of figuring how to estimate CATE, maybe we could have skipped that all together.

For example, suppose that we had that randomized control trial data. Now imagine that rather than 100 individuals, you had a really large randomized control trial with 10,000 individuals in it.

This now opens the door to thinking about directly maximizing or minimizing, depending whether you want this to be large or small, pi with respect to this quantity, which completely bypasses the goal of estimating the condition of average treatment effect.

And you'll notice how this looks exactly like a classification problem. This quantity here looks exactly like a 0 1 loss. And the only difference is that you're weighting each of the data points by this inverse propensity.

So one can reduce the problem of actually finding an optimal policy here to that of a weighted classification problem, in the case of a discrete set of treatments. There are two big caveats to that line of thinking. The first major caveat is that you have to know these propensity scores.

And so if you have data coming from randomized control trial, you will know this propensity scores or if you have, for example, some control over the data generation process. For example, if you are an ad company and you get to choose which ad to show to your

customers, then you look to see who clicks on what, you might know what that policy was that was showing things. In that case, you might exactly know the propensity scores.

In health care, other than in randomized control trials, we typically don't know this value. So we either have to have a large enough randomized control trial that we won't over-fit by trying to directly minimize this or we have to work within an observational data setting.

But we have to estimate the propensity scores directly. So you would then have a two-step procedure, where first you estimate these propensity scores, for example, by doing logistic regression. And then you attempt to maximize or minimize this quantity in order to find the optimal policy.

And that has a lot of challenges, because this quantity shown in the very bottom here could be really small or really large in an observational data set due to these issues of having very small overlap between your treatments.

And this being very small implies then that the variant of this estimator is very, very large. And so when one wants to use an approach like this, similar to when one wants to use an average treatment effect estimator, and when you're estimating these propensities, often you might need to do things like clipping of the propensity scores in order to prevent the variants from being too large. That then, however, leads to a biased estimate typically.

I wanted to give you a couple of references here. So one is Swaminathan and Joachims, J-O-A-C-H-I-M-S ACML 2015. In that paper, they tackle this question. They focus on the setting where the propensity scores are known, such as do it half from a randomized controlled trial. And they recognize that you might decide that you prefer something like a biased estimator because of the fact that these propensity scores could be really small.

And so they use some generalization results from the machine learning theory community in order to try to control the variants of the estimator as a function of these propensity scores. And they then learn, directly minimize the policy which is what they call counterfactual regret minimization, in order to allow one to generalize as best as possible from the small amount of data you might have available.

A second reference that I want to give just to point you into this literature, if you're interested, is by Nathan Kallus and his student, I believe Angela Zhou, from NeurIPS 2018. And that was a paper which was one of the optional readings for last Thursday's class. Now, that paper they

also start from something like this, from this perspective.

And they say that, oh, now that we're working in this framework, one could think about what happens if you have actually unobserved confounding. So there, you might not actually know the true propensity scores, because there are unobserved confounders that you don't observe. And that you can think about trying to bound how wrong your estimator can be as a function of how much you don't know this quantity.

And they show that when you try to-- if you think about having some backup strategy, like if your goal is to find a new policy which performs as best as possible with respect to an old policy, then it gives you a really elegant framework for trying to think about a robust optimization of this, even taking into consideration the fact that there might be unobserved confounding. And that works also in this framework.

So I'm nearly done now. I just want to now finish with a thought, can we do the same thing for policies learned by reinforcement learning? So now that we've sort of built up this language that's returned to the RL setting.

And there one can show that you can get a similar estimate for the value of a policy by summing over your observed sequences, summing over the time steps of that sequence of the reward observed at that time step times a ratio of probabilities, which is going from the first time step up to time little t of the probability that you would actually take the observed action t prime, given that you are in the observed state t prime, divided by the probability-- this is the analogy of the propensity score, the probability under the data generating process-- of seeing action a given that you are in state t prime.

So if, as we discussed there, you had a deterministic policy, then this pi, it would just be a delta function. And so this would just be looking at-- this estimator would only be looking at sequences where the precise sequence of actions taken are identical to the precise sequence of actions that the policy would have taken.

And the difference here is that now instead of having a single propensity score, one has a product of these propensity scores corresponding to the propensity of observing that action given the corresponding state at each point along the sequence.

And so this is nice, because this gives you one way to do what's called off-policy evaluation. And this is an estimator, which is completely analogous to the estimator that we got from Q

learning. So if all assumptions were correct, and you had a lot of data, then those two should give you precisely the same answer.

But here, like in the causal inference setting, we are not making the assumption that we can do covariate adjustment well. Or said differently, we're not assuming that we can fit the Q function well.

And this is now, just like there, based on the assumption that we have the ability to really accurately know what the propensity scores are. So it now gives you an alternative approach to do evaluation. And you could think about looking at the robustness of your estimates from these two different estimators.

And this is the most naive of the estimators. There are many ways to try to make this better, such as by doing w robust estimators. And if you want to learn more, I recommend reading this paper by Thomas and Emma Brunskill in ICML 2016.

And with that, I want Barbra to come up and get set up. And we're going to transition to the next part of the lecture. Yes.

AUDIENCE: Why do we sum over t and take the project across all t?

DAVID SONTAG: One easy way to think about this is suppose that you only had a reward of the last time step. If you only had a reward of the last time step, then you wouldn't have this sum over t, because the rewards in the earlier steps would be 0. You would just have that product going from 0 up to capital T of last time step.

The reason why you have it up to at each time step is because one wants to be able to appropriately weigh the likelihood of seeing that reward at that point in time. One could rewrite this in other ways. I want to hold other questions, because this part of the lecture is going to be much more interesting than my part of the lecture.

And with that, I want introduce Barbra. Barbra, I first met her when she invited me to give a talk in her class last year. She's an instructor at Harvard Medical School-- or School of Public Health.

She recently finished her PhD in 2018. And her PhD looked at many questions related to the themes of the last couple of weeks. Since that time, in addition continuing her research, she's been really leading the way in creating data science curriculum over at Harvard. So please

take it away.

BARBRA DICKERMAN: Thank you so much for the introduction, David. I'm very happy to be here to share some of my work on evaluating dynamic treatment strategies, which you've been talking about over the past few lectures.

So my goals for today, I'm just going to breeze over defining dynamic treatment strategies, as you're already familiar with it. But I would like to touch on when we need a special class of methods called g-methods. And then we'll talk about two different applications, different analyses, that have focused on evaluating dynamic treatment strategies.

So the first will be an application of the parametric g-formula, which is a powerful g-method to cancer research. And so the goal here is to give you my causal inference perspective on how we think about this task of sequential decision making and then with whatever time remains, we'll be discussing a recent publication on the AI clinician to talk through the reinforcement learning perspective.

So I think it'll be a really interesting discussion, where we can share these perspectives, talk about the relative strengths and limitations as well. And please stop me if you have any questions.

So you already know this. When it comes to treatment strategies, there's three main types. There's point interventions happening at a single point in time. There's sustained interventions happening over time. When it comes to clinical care, this is often what we're most interested in. Within that, there are static strategies, which are constant over time. And then there's dynamic strategies, which we're going to focus on. And these differ in that the intervention over time depends on evolving characteristics.

So for example, initiate treatment at baseline and continue it over follow up until a contraindication occurs, at which point you may stop treatment and decide with your doctor whether you're going to switch to an alternate treatment. You would still be adhering to that strategy, even though you quit.

The comparison here being do not initiate treatment over follow up, likewise unless an indication occurs, at which point you may start treatment and still be adhering to the strategy. So we're focusing on these because they're the most clinically relevant.

And so clinicians encounter these every day in practice. So when they're making a

recommendation to their patient about a prevention intervention, they're going to be taking into consideration the patient's evolving comorbidities.

Or when they're deciding the next screening interval, they'll consider the previous result from the last screening test when deciding that. Likewise for treatment, deciding whether to keep the patient on treatment or not. Is the patient having any changes in symptoms or lab values that may reflect toxicity?

So one thing to note is that while many of the strategies that you may see in clinical guidelines and in clinical practice are dynamic strategies, these may not be the optimal strategies. So maybe what we're recommending and doing is not optimal for patients. However, the optimal strategies will be dynamic in some way, in that they will be adapting to individuals' unique and evolving characteristics. So that's why we care about them.

So what's the problem? So one problem deals with something called treatment confounder feedback, which you may have spoken about in this class. So conventional statistical methods cannot appropriately compare dynamic treatment strategies in the presence of treatment confounder feedback.

So this is when time varying confounders are affected by previous treatment. So if we kind of ground this in a concrete example with this causal diagram, let's say we're interested in estimating the effect of some intervention A, vasopressors or it could be IV fluids, on some outcome Y, which we'll call survival here.

We know that vasopressors affect blood pressure, and blood pressure will affect subsequent decisions to treat with vasopressors. We also know that hypotension-- so again, blood pressure, L1, affects survival, based on our clinical knowledge.

And then in this DAG, we also have the node U, which represents disease severity. So these could be potentially unmeasured markers of disease severity that are affecting your blood pressure and also affecting your probability of survival.

So if we're interested in estimating the effect of a sustained treatment strategy, then we want to know something about the total effect of treatment at all time points. We can see that L1 here is a confounder for the effect of A1 on Y so we have to do something to adjust for that. And if we were to apply a conventional statistical method, we would essentially be conditioning on a collider and inducing a selection bias. So an open path from A0 to L1 to U to Y.

What's the consequence of this? If we look in our data set, we may see an association between A and Y. But that association is not because there's necessarily an effect of A on Y. It might not be causal. It may be due to this selection bias that we created.

So this is the problem. And so in these cases, we need a special type of method that can handle these settings. And so a class of methods that was designed specifically to handle this is g-methods.

And so these are sometimes referred to as causal methods. They've been developed by Jamie Robins and colleagues and collaborators since 1986. And they include the parametric g-formula, g-estimation of structural nested models, and inverse probability weighting of marginal structural models.

So in my research, what I do is I combine g-methods with large longitudinal databases to try to evaluate dynamic treatment strategies. So I'm particularly interested in bringing these methods to cancer research, because they haven't been applied much there. So a lot of my research questions are focused on answering questions like, how and when can we intervene to best prevent, detect, and treat cancer?

And so I'd like to share one example with you, which focused on evaluating the effect of adhering to guideline-based physical activity interventions on survival among men with prostate cancer. So the motivation for this study, there's a large clinical organization, ASCO, the American Society of Clinical Oncology, that had actually called for randomized trials to generate these estimates for several cancers.

The thing with prostate cancer is it's a very slowly progressing disease. So the feasibility of doing a trial to evaluate this is very limited. The trial would have to be 10 years long probably. So given that, given the absence of this randomized evidence, we did the next best thing that we could do to generate this estimate, which was combine high-quality observational data with advanced EPI methods, in this case parametric g-formula. And so we leveraged data from the Health Professionals Follow-up Study, which is a well-characterized prospective cohort study.

So in these cases, there's a three-step process that we take to extract the most meaningful and actionable insights from observational data. So the first thing that we do is we specify the protocol of the target trial that we would have liked to conduct had it been feasible.

The second thing we do is we make sure that we measure enough covariates to approximately adjust for confounding and achieve conditional exchangeability. And then the third thing we do is we apply an appropriate method to compare the specified treatment strategies under this assumption of conditional exchangeability.

And so in this case, eligible men for this study had been diagnosed with non-metastatic prostate cancer. And at baseline, they were free of cardiovascular and neurologic conditions that may limit physical ability.

For the treatment strategies, men were to initiate one of six physical activity strategies at diagnosis and continue it over followup until the development of a condition limiting physical activity. So this is what made the strategies dynamic. The intervention over time depended on these evolving conditions. And so just to note, we pre-specified these strategies that we were evaluating as well as the conditions.

Men were followed until diagnosis, until death, and to followup 10 years after diagnosis or administrative end to followup, whichever happened first. Our outcome of interest was all cause mortality within 10 years. And we were interested in estimating the per protocol effect of not just initiating these strategies but adhering to them over followup. And again, we applied the parametric g-formula.

So I think you've already heard about the g-formula in a previous lecture, possibly in a slightly different way. So I won't spend too much time on this. So the g-formula, essentially the way I think about it is a generalization of standardization to time varying exposures and confounders.

So it's basically a weighted average of risks, where you can think of the weights being the probability density functions of the time varying confounders, which we estimate using parametric regression models. And we approximate the weighted average using Monte Carlo simulation.

So practically how do we do this? So the first thing we do is we fit parametric regression models for all of the variables that we're going to be studying. So for treatment confounders and death at each followup time.

The next thing we do is Monte Carlo simulation where essentially what we want to do is simulate the outcome distribution under each treatment strategy that we're interested in. And

then we bootstrap the confidence intervals.

So I'd like to show you kind of in a schematic what this looks like, because it might be a little bit easier to see. So again, the idea is we're going to make copies of our data set, where in each copy everyone is adhering to the strategy that we're focusing on in that copy.

So how do we construct each of these copies of the data set? We have to build them each from the ground up, starting with time 0. So the values of all of the time varying covariates at time 0 are sampled from their empirical distribution. So these are actually observed values of the covariates.

How do we get the values at the next time point? We use the parametric regression models that I mentioned that we fit in step 1. Then what we do is we force the level of the intervention variable to be whatever was specified by that intervention strategy. And then we estimate the risk of the outcome at each time period given these variables, again using the parametric regression model for the outcome now.

And so we repeat this over all time periods to estimate a cumulative risk under that strategy, which is taken as the average of the subject-specific risks. So this is what I'm doing. This is kind of under the hood what's going on with this method.

DAVID SONTAG: So maybe we should try to put that in language of what we saw in the class. And let me know if I'm getting this wrong. So you first estimate the markup decision process, which allows you to simulate from the underlying data distribution.

So you know that probability of this sort of next sequence of observations, given the previous sequence and action and previous actions, and then with that, then you could then intervene and simulate the forms. Because that was, if you remember Frederick gave you three different buckets of approaches. Then he focused on the middle one. This is the left-most bucket. The right?

AUDIENCE: Yes.

DAVID SONTAG: So we didn't talk about it.

AUDIENCE: No, [INAUDIBLE] model based on relevance.

BARBRA Yeah. Yes.

DICKERMAN:

DAVID SONTAG: But it's very sensible.

AUDIENCE: Yeah. But it seems very hard.

BARBRA DICKERMAN: What's that?

AUDIENCE: Sorry. Oh, it seems very hard to model this [INAUDIBLE].

BARBRA DICKERMAN: Yeah. So that is a challenge. That is the hardest part about this. And it's relying on a lot of assumptions, yeah. So the primary results that kind of come out after we do all of this. So this is the estimated risk of all cause mortality under several physical activity interventions.

So I'm not going to focus too much on the results. I want to focus on two main takeaways from this slide. One thing to emphasize is we pre-specified the weekly duration of physical activity. Or you can think of this like the dose of the intervention.

We pre-specified that. And this was based on current guidelines. So the third row of each band, we did look at some dose or level beyond the guidelines to see if there might be additional survival benefits. But these were all pre-specified.

We also pre-specified all of the time varying covariates that made these strategies dynamic. So I mentioned that men were excused from following the recommended physical activity levels if they developed one of these listed conditions, metastasis, MI, stroke, et cetera. We pre-specified all of those. It's possible that maybe a different dependence on a different time varying covariate may have led to a more optimal strategy. There was a lot that remained unexplored.

So we did a lot of sensitivity analyses as part of this project. I'd like to focus, though, on the sensitivity analyses that we did for potential unmeasured confounding by chronic disease that may be severe enough to affect both physical activity and survival.

And so the g-formula is actually providing a natural way to at least partly address this by estimating the risk of these physical activity interventions that are at each time point t only applied to men who are healthy enough to maintain a physical activity level at that time. And so again in the main analysis, we excused men from following the recommended levels if they

developed one of these serious conditions.

So in sensitivity analyses, we then expanded this list of serious conditions to also include the conditions that are shown in blue text. And so this attenuated our estimates but didn't change our conclusions.

One thing to point out is that the validity of this approach rests on the assumption that at each time t we had available data needed to identify which men were healthy at that time enough to do the physical activity. Yeah.

AUDIENCE: Sorry, just to double-check, does excuse mean that you remove them?

BARBRA DICKERMAN: Great question. So because the strategy was pre-specified to say that if you develop one of these conditions, you may essentially do whatever level of physical activity you're able to do. So importantly-- I'm glad you brought this up-- we did not censor men at that time. They were still followed, because they were still adhering to the strategy as defined. Thanks for asking.

And so given that we don't know whether the data contain at each time t the information necessary to know, are these men healthy enough at that time, we therefore conducted a few alternate analyses in which we lagged physical activity and covariate data by two years. And we also used a negative outcome control to explore potential unmeasured confounding by clinical disease or disease severity.

So what's the rationale behind this? So in the DAGs below for the original analysis, we have physical activity A. We have survival Y. And this may be confounded by disease severity U. So when we see an association between A and Y in our data, we want to make sure that it's causal, that it's because of the blue arrow, and not because of this confounding bias, the red arrow.

So how can we potentially provide evidence for whether that red pathway is there? We selected questionnaire nonresponse as an alternate outcome, instead of survival, that we assumed was not directly affected by physical activity, but that we thought would be similarly confounded by disease severity.

And so when we repeated the analysis with a negative outcome control, we found that physical activity had a nearly null effect on questionnaire nonresponse, as we would expect, which provides some support that in our original analysis, the effect of physical activity on death was not confounded through the pathways explored through the negative control.

So one thing to highlight here is the sensitivity analyses were driven by our subject matter knowledge. And there's nothing in the data that kind of drove this.

And so just to recap this portion. So g-methods are a useful tool, because they let us validly estimate the effect of pre-specified dynamic strategies and estimate adjusted absolute risks, which are clinically meaningful to us, and appropriately adjusted survival curves, even in the presence of treatment confounder feedback, which occurs often in clinical questions. And of course, this is under our typical identifiability assumptions.

So this makes it a powerful approach to estimate the effects of currently recommended or proposed strategies that therefore we can specify and write out precisely as we did here. However, these pre-specified strategies may not be the optimal strategies.

So again, when I was doing this analysis, I was thinking there are so many different weekly durations of physical activity that we're not looking at. There are so many different time-varying covariates where we could have different dependencies on those for these strategies over time. And maybe those would have led to better survival outcomes among these men, but all of that was unexplored.