Dr. Roychoudhury, He had 15 years of extensive experience. He served as the (indistinct) co-chair workshop. And he’s serving as co-chair for DIA, FDA, biostatistics (indistinct). Dr. Roychoudhury was exacted to be a panel of the American (indistinct) Association. The country (indistinct), international (indistinct) society in 2019. So let’s welcome Dr. Roychoudhury. Now I’m yours. Thank you. Thanks a lot Dr. (indistinct) for the nice introduction. Can you all hear me well? And thank you for the opportunity to present here. and having a chance to interact with all of you. So today, I’m gonna talk about a problem that is many of the recent drug development are facing, and we try to talk about better interpretation
29 00:01:54.813 --> 00:01:57.007 of clinical trial data,
30 00:01:57.007 --> 00:01:59.024 especially bringing the different perspective 
31 00:01:59.024 --> 00:02:01.617 as some of you know and some of you don’t. 
32 00:02:01.617 --> 00:02:04.927 FDA actually specifically started 
33 00:02:04.927 --> 00:02:07.890 a patient oriented drug development program, 
34 00:02:07.890 --> 00:02:11.550 which kind of how to make the data more understandab 
35 00:02:11.550 --> 00:02:13.950 to the patient, more reachable to the patient. 
36 00:02:13.950 --> 00:02:17.030 So this was kind of on that theme mostly. 
37 00:02:18.360 --> 00:02:21.031 Before I begin, just I wanted to mention 
38 00:02:21.031 --> 00:02:24.515 the standard disclaimer, this my own view, 
39 00:02:24.515 --> 00:02:27.900 and not necessarily reflect the view of the Pfizer. 
40 00:02:29.700 --> 00:02:32.670 So I think many of you have work 
41 00:02:32.670 --> 00:02:36.180 on survival analysis as a coursework 
42 00:02:36.180 --> 00:02:40.710 or maybe analyzing trial data, research perspe 
43 00:02:40.710 --> 00:02:44.070 So often, in critical trial setting, 
44 00:02:44.070 --> 00:02:46.710 we call that analysis over time to event data. 
45 00:02:46.710 --> 00:02:51.172 Because we look into the data that up to time, 
46 00:02:51.172 --> 00:02:54.390 so we analyze the time up to a certain event 
47 00:02:54.390 --> 00:02:56.577 or sensor the data looked at. 
48 00:02:56.577 --> 00:02:59.175 And the standard way of analyzing such data, 
49 00:02:59.175 --> 00:03:01.320 these are the, more or less, a standard tool. 
50 00:03:01.320 --> 00:03:04.161 I’m sure you all have done or are going to do 
51 00:03:04.161 --> 00:03:07.083 in your coursework, that looking into Kaplan marker 
52 00:03:07.083 --> 00:03:09.887 which looks into the survival function, 
53 00:03:09.887 --> 00:03:12.180 the effect over time. 
54 00:03:12.180 --> 00:03:15.900 Log-rank test, which basically tests the differen 
55 00:03:15.900 --> 00:03:18.690 between the two curve, that if the treatment is better
than control or control is worse.

And then last, to summarizing treatment of it, right?

At the end of the day, we need to know how good is the treatment.

And one way to say that was basically hazard ratio

or something we call Cox regression.

Now low-rank p val-

These are very standard reporting techniques.

Any time to event data, if you pick up any medical journals,

are typically analyzed using three (indistinct).

But the question was, and based on some examples done,

we’ll dive into the details as we go along.

the question really came up,

are these really good metrics to analyze the data?

So just to give a quick introduction,

I’m sure all of you know it very well.

But just to understand the fundamental assumptions,

what the Cox regression means.

So Cox regression and hazard ratio are the very,

very popular method.

And it started, basically introduced

by Dr. DR Cox in 1972, one of the brilliant statistician.

And it’s closely related to mathematically

with the log-rank test,
There mathematically, the score function, the score test for Cox regression is equal value to the log-rank test under two sample case. But it has an assumption also. It assumes that the treatment effect basically is constant over time, so it emerges at the beginning of the trial and remain kind of a constant over time. Which is a problem when this is not true, because not all drugs work the same way. Sometimes, maybe some patient get benefited after getting treated for longer time. On such case, such a measure has a ratio, start to lack its interpretation. And also there is some problem regarding causal inference perspective. Let’s look into two examples, two, three examples, real life examples, and try to understand where the problem was. So first, let’s start with more simpler when treatment effect emerges at the beginning and remained at that (indistinct) over the trial. So the trial on... And they’re all real trials, and I added the references in case you want to look into later. The one on the left is basically a trial for non-small cell lung cancer where gemcitabine.
and gemcitabine plus erlotinib combination has been looked into, which is showed the blue curve. There is experimental combination drug, which shows superiority over the standard of care, gemcitabine. The one on the right is basically on refractory multiple myeloma disease, which is a very fatal disease. And they’re basically triple combo, so kRd is a triple combination drug, where Rd is basically a double combination drug. I’m not going into the details of that. Looking into basically kRd, was the experimental drug of this showing the superiority? In both cases, the proportionality has assumption, looks valid, because the effects started and it continued. But that’s not always the case. Here are the very few, very recent examples, especially if you are looking into the newspapers as well. And I think two years or three years back, the person from MD Anderson actually got the Nobel Prize for looking into the immunotherapy, one of the fundamental research from that area. So when it does this particular type of therapy, basically putting people immune system,
basically and they kind of train them to fight against their cancer.

And CheckMate 141, which is squamous cell carcinoma on head and neck,

nivolumab is one of the immunotherapies of the first (indistinct) cluster for their client.

When they looked into the actual treatment effect,

started to emerge pretty late.

So that means it take the time for the immune system to actually work.

And then the question is,

the treatment effect is no more constant, right?

Immunity is at three month.

It’s even more problematic coming into the example in IM211 trial, which is urothelial carcinoma and well atezolizumab, a compound, and Roche was actually looking into against chemotherapy.

The compound was detrimental at the beginning and slightly at least compared to the standard of care,

the chemotherapy that how the people are treated.

And then it showed benefit, definitely.

But there is something more interesting in the second ex-

I will concentrate and each of them are interesting,

each of them have a story of its own.
But I just focus on this IM211 trial to bring in the patient perspective a little bit. So if you look into this trial, of course the trial doesn’t have a significant amount, right? The log-rank test, which tested the superiority of the curve basically said it’s nonsignificant. But if you actually wanna look into this survival curve a little bit more details, one can see the survival effect is really emerging. You come in carcinoma. So this is kind of an 8% survival benefit is for this class of patient is quite a meaningful benefit on that context. So the question is, are we evaluating this drug correctly by these two standard metrics like Cox, hazard ratio, which doesn’t look good here, 0.9, close to one. So it has a ratio less than one because we are comparing treatment versus control. Less than one means treatment is better and bigger than one means treatment is worse. So basically doesn’t really reflect these angles. And also, if you look into the hazard function plots, or hazard ratio plots,
you can quickly see they’re not constant over time.
They’re actually emerging or varying there.
I mean, there has been a working group,
there has been a number of workshop here,
and even FBA and other regulators got interested.
There has been a lot of discussions started in 2016
when nivolumab, the first class
immunotherapy compound came in in a development phase.
And there has been a lot of research.
People were looking into different tests, okay,
log-rank is not good, because it is most powerful test
under proportionality hazard assumption.
We don’t have it, especially for situation like this,
CheckMate or IM211 trial;
or we should do some other test.
This is not powerful, some better test.
So there has been a zoo of tests being came in.
You may have heard about some of them weighted,
so instead of log-rank test, weighted log-rank test
where we selectively weight that separate areas
of the Kaplan-Meier curve.
Then some a little bit more robust, like Max-Combo,
weighted Kaplan-Meier test, restricted mean survival time,
and there are many, many. I mean, and they try to look into how to handle different potential. Because one thing we need to understand, proportional hazard is a very specific character. When we say proportional hazard that these two hazard plots are (indistinct). But once we set non-proportional hazard, there could be many, many possibilities. Drugs can be crossed, drugs can separate the link. Drugs can first separate and then emerge back. So it’s a no more unique (indistinct). So we need a set of methodology that’s sort of robust across this different class of alternatives. But the problem is even having a P value, suppose once we call about, okay, there’s a good test and we got a good P value for such a curve. But one can say, oh how can a P value is meaningful? The beginning of the curve, the patients are getting worse. Is a P value really meaningful in this context? So rejecting the null is really a less informative. You need to know more information to understand these kind parts of compounds better. And that’s also been looked into. There are multiple, multiple things has been looked into
with simulation, with data, a lot of meta analysis

with different data looking into the percentile, right?

I mean one part of we’re looking, why don’t we look into the ratio of the median

Or we look into the (indistinct) time or percentile, right?

Things are separating at percentile.

Maybe we can look into percentile over the time ratios of those.

Then milestone survival and coming into a moment,

which is a sort of a very meaningful
to a patient’s restricted means survival.

Basically you average over the area under the curve,

which has a ratio, because we are doing with a test,

a dual estimator of that is weighted hazard ratio,

piecewise hazard ratio.

Cox model, now, the initial Cox model doesn’t have

a time component in it (indistinct).

Only the time component is in the baseline hazard.

We can introduce a time component into that

to make sure the treatment effect is sort of time dependent.

And then there are other things like net benefit.

I won’t dive every one of details,
but I just wanted to mention that hazard, so weighted hazard ratio means hazard ratio, we were doing a Cox regression. We were looking into basically the regression coefficient corresponding to the treatment using a partial likelihood method. And now, the whole idea was similar to like testing where we are weighting different part of Kaplan-Meier differently, we weight different part of the partial likelihood differently. Or there are other type of weightings as well, like average weights and others. So basically, the whole idea was not to treat each event similar, but differently based on their interest. And that’s the (indistinct) of course, treatment emerges at the end. We may be interested more towards the end, but it adds a subjective choice, right? And it’s not very easy for non-clinicians to understand on that aspect. So now then, of course in 2005, people also talk quite a bit about piecewise hazard ratios and now still piecewise hazard is still very important. Like you divide the whole time axis into different intervals and you basically calculate local hazard ratios. Which is very meaning ’cause you can look
The hazard ratio's close to one, then the hazard ratio emerges. And there's a natural extension of that was basically using a regression using a time factor. But really, the power and the performance really depends on the function that you choose, which is, again, is difficult to interpret in a practical sense that if results really value on such a choice. But one thing was kinda after all, among all this discussion, one thing was when we talked to non-statisticians, specifically clinician, one thing was very clear, one measure we found. The measure is improvement in survival. That means they're very clear about certain metric that, okay, what is the survival gain at five year or eight year? Those metrics seems to be very intuitive, very clear to to non-statistician patient and all other stakeholders. But the only problem is at the beginning of the file, you just don’t know when the curve is gonna be separate. You don’t know that.
So (indistinct) finds such point, can be very dangerous for you.
The last measure I mentioned was this, been over and over discussed,
which is called residual, meaning lifetime,
basically (indistinct) lifetime,
residual means survival time.
Which has been over and over discussed recently
in clinical literature, as well as in statistical,
restricted mean people.
When people looking into restricted mean,
that means they're looking for supposed up to a time Tau.
You cut the curve, you compare the area under the curve,
which is in this block.
And you see if you can look into the difference,
you can look into the ratio,
and try to see if the area under the curve is better
on that context.
But remember, the problem is here,
the comparison also very much depends on the choice
of the cutoff, where you choose it, basically.
The Tau and risk of the censoring pattern
of the Kaplan-Meier plays a very big role
in such a such compass,
especially in a setting for such metrics.
Especially, it's very problematic
when in a cancer setting like metastatic,
the one I showed you earlier, because there is a, the disease is very fit on it, happens very quickly. Patient gets multiple therapies. So censoring patterns are much more aggressive, whereas there are other disease setting where RMST seems to be very meaningful. Because you get lot of follow up much more uniformly. Now the problem was, okay, all these measures differently. We apply, we get different results, we get, okay, one is good right here, right? We talked about IM211 at beginning. We saw survival benefit. We see that the survival benefit into the 12 month zone, but RMST was still showing, okay, there is no confirmation of such an effect. So there has been a interesting paper I came across last year. It was a patient voice survey in UK. They basically did a survey on patients as well as the health practitioners who are participating in the TACT trial. Taxotere is a adjuvant chemotherapy drug. It’s a very big trial. And they surveyed that what patients really want, how they want the trial results are for.
I mean, because at the end of the day, the doctors tried to discuss these results with the patient before they choose as a therapy. And then, it’s not surprising that actually, it’s coming in that the patients are really want to understand these results in a very simple term. I mean, it’s not like, oh, your drug may not be decreased event rate, but it expected survival going to get really bigger. But that’s not the question a patient is asking, right? I mean, it’s not type of mindset a patient has. Then of course, the patients are one who have the results as soon as possible. It’s a very interesting article I suggest to read. It’s a very fun reading article to look into the patient voice, how patient wanted... Because most of the time patients are not being always communicated about trial results and how they wanted. It came out to be very nicely in this, especially most of them actually want still comfortable to have the results from their nurse or doctors. Because they’re more comfortable to discuss with them.
rather than having a big seminar or formal paper.

But really, let’s think about the patient perspective,

what question they really ask.

The question they really ask are these, right?

I mean, if we put ourself onto that shoes,

the question is does this drug really work?

What are my chances that I will do better

in terms of survival or in terms of painful quality

of my life if on the new drug compared to no treatments?

So these are the much more simple question.

Unfortunately, many of the well known methods are unable

to address because I mean, you can indirectly address

those question or you may have the study shows you have

a benefit of pharma.

But that really doesn’t answer the question

from that patient perspective.

Which actually motivates us a little bit,

that can we dig in?

Can we try to see if we can use

our modern statistical analytic methods

to address some of that question,

especially from peer’s perspective,

from a kind of practitioner’s perspective.

So one of the thing that is more easier for anybody

to understand the visual graphics, right?
Kaplan-Meier plot is something which is more easy to comprehend compared to many other metrics.

So that was kind of how our motivation (indistinct) became.

And then the question is really, it's not like what trial shows.

Now the question, if a new patient is entering the trial,

can we predict their benefits based on the available data we have?

And this could be an additional metrics. Of course we are not saying these metrics is gonna change the trial or the trial practice,

but this is definitely another metric which can help patient much better for making their decisions.

So what we introduced is basically this quantity which is called individual effect, Y minus X.

What is Y?

So Y is a survival time at the treatment arm.

I'm just making simpler, instead of progression survival,

let's have the conversation on survival,

because it's just easier to understand.

So Y is basically if a patient receives treatment and control, Y means their survival time in treatment and X is the survival time in control.

But in a real trial, there was never been one patient
who received both treatment and control. So this quantity is actually counterfactual. We cannot have that data. So we somehow have to predict this. So in order to do that, we first need to understand the marginal distribution of Y and X. And then also understand the need to take account the correlation in Y and X. Let’s dive in what we need to do in this step, and I’ll go into details of the statistics in next of my few slides, what technically it means. Basically, we are trying to find, so what we are trying to do, we are trying to find this predictive patient level effect of a drug. So what we need to have for that? We need to have a marginal distribution of survival for both, if a patient is independently goes to treatment or in control, basically. We need the marginal distribution first. Then we somehow need to calculate the difference by considering the association between them. Because Y and X are coming from a same patient. So first thing was interest, first thing was kind of how to do the marginal distribution.
That part is easier. That part is technically maybe intrigued, but easier.

So suppose any trial data we have, we can have multiple trial data. We look into the treatment, and we look into the control. I just, for simplicity, suppose we have one trial in our hand. And if there are multiple, we can definitely add more layers to that and adding trial specific event. So we can fit a...

Piecewise exponential models are more, kind of a flexible model. So of course, one can use (indistinct), other parametric family.

The reason we are moving into parametric because we wanted to extrapolate. We wanted to see the survival in the future. So that’s why we chose basically piecewise exponential graph where each of the within, the whole time axis is divided into different time span, different time points, and then we assume the hazard response within that. But just not, we don’t assume any hazard ratio.

We are (indistinct) treatment effect responsive.
We are fitting this piecewise exponential separately for treatment and control, basically. So there is no proportional hazard option. And then here, we assume that alpha and beta (indistinct) and using the, basically, the gamma prior for each of the interval specific hazard. And we assume non-informative practice. Most often we only have the trial data, but if they have more information from early trials, you can use informative (faintly speaking). But one of the major challenge always for piecewise exponential is how you choose cut points there. There’s like the cut choice of cut points, people often do eyeballing, right? They look into the plot, eyeball the times, but those are mostly subjected. That means one prediction to another prediction, it can vary there, which is problematic. We need a a little bit more uniformal way of selecting these cut offs. The second more easier one can think, okay, I use the person notes there. But here the problem is when we fixed the personnel, maybe from one plot to another, there may be intervals which doesn’t have much event.
So (indistinct) may not be basically calculated in a right way. So not stairwise. There is no event within that interval. So what we did, we basically looked into a optimal cutoff points searching algorithm. So we basically first divided the cut, the axis based on the person health. So you have 10 intervals to... So basically, at most, 10 intervals. So then we consider all possible models. So one component, two component, to the power 10 component model. Of course, if some of intervals doesn’t have very few event, we basically merge them so that you have a reasonable estimate for (indistinct) basically. And we chose the best model based on the DIC among those two to the power, N number of models. Of course, there’s not always 10 because some of the intervals may be empty, so we plot them. And one can actually do a k-fold cross validation as well, which is we also looked into kind of a giving more or less similar result. But if you have a long term data, one can also do a k-fold validation in order to choose the hard points.
Now the second part is prediction, right? So prediction, so what you got from the model now is that distribution of (indistinct). That’s parameter. But when you go to prediction, we are talking about sampling space now. So we are talking about a new patient’s survival time. So that needs to take account uncertainty of the new sample. That’s the beauty of the Bayesian distribution, the posterior predictive distribution automatically does that. And the setting, the reason we took this setting, because the predictive posterior parameter distribution is again, a piecewise (indistinct) distribution, which is basically closed form, that first hour computation quite a bit. And then we basically use, as I say, it’s can easily done with this Bayesian computation. But if you want to do more complex model, we’ll one, need to move into a little bit more MCMC algorithm or kind of writing and way of sampling it. Now the third aspect, which was actually the more, most interesting aspect. Now we got the marginal distribution. We predicted the marginal survival times.
Now the question is, it’s the same patient, right, we are talking. So they’re correlated, X and Y. How we most are correlation structure, which is meaningful? Now that thing was actually the idea came from a very old paper by Lehmann and Doksum in 1974, which we were looking for a scale shifting distribution. I’m going into a moment. Which actually brings up a very important property. And that property is very important because this methodology, it depends on that property. And we can talk about in the setting where you think this property is not true. So the property is basically a rank preserving property. What that means, that means if a two population, so if one fail earlier in control, they will fail also earlier in treatment. That’s a rank preserving property. That’s a very important property for this. But one can also question that, okay, for targeted therapy, that may not be true. Somebody with a biomarker that order may change, right? So we need to do appropriate adjustments to make that assumption.
I’m going into that. That was one of the referee comments, by the way.

But before going into that, this paper was really fascinating. I mean, it was very simple paper in actually 1974 and also of statistics, very simply written paper.

And what they were saying that they were looking into distance between two normal curves there. And the property they introduced was scale shifting, a shift function, which basically makes that, so your X and Y, what you just mentioned, we said X plus delta X, which is one can interpret as gain also, and Y kind of have a same distribution.

That was the main property. But it’s very, basically, that means you project this curve to other curve. There is a path to project all. And the solution of delta X is, I mean, what can easily cost start that it’s not unique.

But they actually, this is how they constructed it using the shift function. But what did that means for our case? Why we need that, right? So this means that, so if we,
if I know marginally, the (indistinct) tells of X,
I can project the same content into the Y distribution,
that basically we’ll coordinate ourselves.
So that means if we build, if the ordering of X remain,
ordering in Y will remain too.
That is what the rank preservation property’s coming in.
I’m sorry, I’m going to revisit the...
Sometimes simple papers give you very nice ideas as I find.
Now the job is simple, right?
Now we have marginal distribution in hand.
We get the predictive distribution,
how to predict the marginal, and now we know
how to link them using the quantile function
from one to another, how to project.
basically what we did,
we simulated from this posterior distribution
and we simulated this uniform numbers.
And then what we did in order to bring X and Y to related,
we basically from same quantile,
we obtained the X for given US,
we obtained the quantile XS and YS for each case.
So they’re related by that way.
We project it from X to Y
and then that gives us a pair XS YS for them
636 00:31:34.573 --> 00:31:37.053 and we can make the distribution.
637 00:31:41.029 --> 00:31:43.920 But of course, this is questionable, right?
638 00:31:43.920 --> 00:31:46.470 I mean, we are saying the order remains,
639 00:31:46.470 --> 00:31:48.300 which is not always true.
640 00:31:48.300 --> 00:31:51.542 Suppose this is a very classic example
641 00:31:51.542 --> 00:31:53.823 from the nivolumab trial.
642 00:31:54.690 --> 00:31:58.110 So it’s attacking the PD1 inhibitor.
643 00:31:58.110 --> 00:32:02.937 So people with PD1 expressed, it’s supposed to work better.
644 00:32:02.937 --> 00:32:05.070 So if you pick two subject,
645 00:32:05.070 --> 00:32:07.918 one is PD1 expressed and PD1 expressed not,
646 00:32:07.918 --> 00:32:12.330 if in control, PD1 expressed one actually works worse
647 00:32:12.330 --> 00:32:15.360 than the PD1 non-expressed treatment that can reverse.
648 00:32:15.360 --> 00:32:19.174 Because PD1 is the line, the target of the truck.
649 00:32:19.174 --> 00:32:21.030 So on such a case,
650 00:32:21.030 --> 00:32:24.270 our solution is basically divide the groups
651 00:32:24.270 --> 00:32:26.430 into the homogenous biomarker class.
652 00:32:26.430 --> 00:32:28.260 And then, the (indistinct) still works.
653 00:32:28.260 --> 00:32:31.590 And that can be easily done adding it as a regression
654 00:32:31.590 --> 00:32:34.323 into the model.
655 00:32:35.459 --> 00:32:37.570 So then finally, we summarize this
656 00:32:41.576 --> 00:32:45.270 with a survival gain and out the loss of tests,
657 00:32:45.270 --> 00:32:48.267 I mean basically using the posterior, sorry,
658 00:32:48.267 --> 00:32:52.140 the predictive distribution of Y minus X.
659 00:32:52.140 --> 00:32:56.160 And then also, we summarized the median
660 00:32:56.160 --> 00:32:59.460 and the 95% prediction intervals.
661 00:32:59.460 --> 00:33:03.330 And also to us, the Kaplan-Meier plot is still important
662 00:33:03.330 --> 00:33:05.670 because that’s where the whole story begin.
26
That's the data which generates our marginal distribution where we actually...

So we should compare that side by side on this.

Let's see how this method works then.

I mean, so we started, does it improves anything after all doing fancy things,

make things complicated, making lot of mathematical names.

Did you improve anything?

So did you gain more insight into this?

So let's go back to this trial where begin the urothelial cancer trial of atezolizumab,

which is basically the low-rank test says it was basically...

The survival curves are not separated.

Basically significance that doesn't reach.

The Cox has a ratio, upper bound is about one,

even the median, because it's very interesting.

Because the curve actually is separated after median.

So even somebody looking into the median, the treatment is worse than that for the standard of care.

So the question is really, how can we...

But when we look into the survival differences, we see a significant survival gain by the patient who remain on the therapy for a longer time.

But the question is, can we somehow communicate
that better with this individual event?
So the one on the left hand,
one on the left side is basically the probability plot.

So it is basically the first column.

We looked into the gain of survival bigger than zero month,
one month, two month, six month.

And the plot on the right side is basically median

and then predicted interval.
So as you can see, there is an initial setback.
There is a significant probability

that atezolizumab can actually improve the survival

in the problem.
At least you have,
if you’re looking to this plot,
the improvement of three month or higher,
which is urothelial cancer is pretty good,
you have 30 to 40% probability,
which needs to be considered for these patient.
Because they don’t have,
they only have chemotherapy as a treatment for that.
The most interesting thing comes actually
on the right hand side.
If you look into that, the patient as we talked about,
the benefit actually emerges as the patient went long term
into the therapy.
It starts pretty much, much (faintly speaking) to others,
compared to others. Let’s look into another.
So okay, non-proportional hazard, this may be useful.
But can we still use that for proportional hazard?
Do they have any value? Actually, do they add anything?
’Cause they’re log-rank has a ratio are more popular, right?
Our argument is no, but maybe this measure can help you there too.
So I go back to this lung cancer PA3 example of gemcitabine and gemcitabine plus erlotinib,
which is statistically significant with a very marginal hazard ratio.
But statistically significant and the media advantage was also not very good
by only .3 months of advantage.
The question is really, does the survival effect is meaningful in that way?
Always.
So again, we started to plot these two.
So the one here is basically the plot for the survival gain, X minus Y, once again, control.
Same patient treated in treatment versus same patient treated in control.
What is the survival gain?

And the one on the right is basically the median and the corresponding interval.

As you can see here, especially the patient who discontinued or there are some questionable benefit on that question.

So just having a proportional hazard and giving a hazard ratio may not be giving the full point that we are looking for.

There may be some more to it, which can be further investigated.

When it’s prescribing a patient, maybe there are certain characteristics.

When it’s prescribing a patient, why the patients may be continuing early and some kind of (indistinct) and some kind of special...

So those patient may not be the benefit is as good, as prominent as the patients who are long term treated.

So we also investigate CM.

I just don’t want to show all the plot just not to bore anymore, but messages are very similar.

We also look into CM141.

We also look into that ASPIRE trial as I explained you earlier about multiple myeloma.

and the CheckMate141 is basically squamous cell carcinoma

in head and neck.
And we also looked into the CheckMate057 trial where there’s a significant subgroup that means the one group with PDL-1 expressed has a significant survival. But PDL-1 non-expressed has no survival benefit. So we looked into all these examples into the export data and our codes are also relevant in public if you want to play with that. Basically the message we got, so the one on here is that this table toward the mean survival gain, median survival gain, the predictor interval. And what is the chance that probability that Y, which means your time, if a subject is receiving treatment if a subject is receiving control, what’s the time Y is bigger than X, basically we started to see some interesting feature. It gives us much more quantification of benefit compared to our P value or a hazard ratio in that way. Of course there are uncertainties in all of them as well as the clear cases. We see some uncertain situation, that patient may or may not be benefit in some cases.
So the quick conclusion with the increasing complexity of the drug and the different complex pathways we are talking, it is very important. And different patterns of treatment effect, it’s very important to have a simpler language with the patient. I mean, because we’re all telling what they’re answering their question at least directly. We find predicted individual effects sort of a step towards that, which answers the patient’s question more clearly, kind of a clinically relevant. And also it step forward, and as well as it supports this 21st century cure act of patient centricity. Especially find very useful in where your standard is of proportional hazard fails. I just wanted to quickly thank my collaborators whom I collaborated with and here is outpatient along with the software is available along with the software is available in case you want to try out. And the references that I mentioned including actually the... I just wanted to always mention, if you are reading the famous Lehmann book, it’s actually there as well.
I just find out last night it’s there as well. And not only (faintly speaking) statistics people. And I just want to thank you for your attention. Thank you.

Does anybody have any questions?

Sure.

Is there a notion of why you get that kind of culling effect with the treatment groups you were showing in the survival curves that the, basically the treatment arm looked worse at first.

Yes, I think for a (indistinct), I think there was a certain group biomarker. It’s again, a heterogeneity of treatment effect. Certain biomarkers did, I mean, most of the non-proportional hazard are the same story.

I mean, the certain groups didn’t function well at the beginning until they basically received the treatment follow up as well. They actually worse, they were quite a bit. Okay, so do you think it’s a treatment effect?

not a property of the population that was in it? I think it’s, I mean.

it’s more of a safety of words I guess I believe.
But of course, it’s a road we don’t know all the details inside of it.

But the compound, which is the results which is available in that paper, it seems like there is an effect where it’s really detrimental, the treatment effect.

(faintly speaking)

Any other questions?

All right, anything from Zoom land?

Yeah, I have a question.

So it’s interesting to model non-proportional patterns, but I think maybe another interesting question is to why there was non-proportional pattern (faintly speaking), right?

So maybe (faintly speaking). So say for example if we owe something like a random voice or like classification, then we’ll be able to see each subgroup benefits from the treatment to see each subgroup benefits from the treatment (faintly speaking)?

(indistinct) comment something else, like why this (faintly speaking)?

(indistinct) comment something else, like why this (faintly speaking)?

(indistinct) the modeling, the causes or figuring out why there’s non-proportional patterns.

Sure, I think what you just said.

that was basically the method, some of five star, some of the method that more people develop.

(indistinct) and their group did.
They basically looked into a elastic net to find out the sets were basically you have, which is heterogeneous. They divided the group, this heterogeneous clusters, basically, into that. And then tried to interpret treatment of problems. I mean, the major problem is sometimes those groupings are very hard to interpret. Because it’s so much data driven, right? And secondly, specify such a method as an analysis and this is a great method to exploration. I fully agree. But if you think about a drug and kind of a reporting of a drug, that could be a very risky method to do. But definitely, they are thinking down on that avenue. The only reason I think the five star was a very interesting idea, the only problem really came in is the estimation of treatment effect at the end of the day. Because now, you have a selection, right? Now you have to have the selection probability incorporated into the treatment effect. Some clusters are so small when you put the adjustment to the selection probability,
it’s not very intuitive to non-status station anymore.

But it’s been done.

I mean, there are an example,

I think in (indistinct) medicine,

if you search by five star, you can see that.

I think that the major got hit by the interpretation

of the treatment effect.

But you know what they did?

They actually fit a parametric...

First, they did three things.

They looked into each set,

because those population are homogenous.

So they fit the Cox regression model there.

And also they looked into a parametric regression model.

They looked into each set,

because those population are homogenous.

So they fit the Cox regression model there.

And also they looked into a parametric regression model.

But the only problem is,

as soon as you adjust for your selection probabilities,

if you have a huge effect, right?

The selection probability somehow do a tool on that,

which is clinicians don’t find very intuitive, that case.

Because at the end of our regular...

I mean, how do you put such a thing on a drug level?

That is a problem.

But I think such a thing should be done for our...

If we have already face data, we should explore this.

I really think that should be the case.
Okay, so we don’t have any questions.

Let’s thanks again.

Yeah, I have one, just one quick question.

When you’re predicting your why, why not augment that data with publicly available data based on features like comorbidity, age, some of the known predictors in terms of survival rates?

Absolutely, absolutely.

Sorry, I skipped that.

But that’s a great question.

Actually, if we see that, I just wanted to go...

We actually said if you have such a thing, we just convert this into a regression. So you can actually, instead of having just a unstructured model here, we can actually plug in all the rigorous (indistinct).

But of course then, you don’t have (indistinct) in your family anymore. It’ll be conditionally extensive.

But that’s a very easy extension of this.

Yes, absolutely.

Absolutely.

We actually did that for, sorry.

We actually did that for this example where there’s heterogeneous effect by the biomarker.
You basically fill in the regression, and that’s how we operate.

Sorry, I interrupted.

Somebody was...

Great, thank you.

Oh, I have another question.

So when you tell the patients about their predictive treatment effects, if that affects I’d say how patients attach to the assigned treatment, that some way actually affects how you estimate your, for example, the first step for the estimates.

And does that actually affect the way how you...

Because I think estimation and also the following steps are actually the different, for example, the purpose of clinical trial will be estimation.

But the purpose of, like say telling patients the individual treatment effects, a predicted individual treatment effects will be like a different purpose.

It's not really the estimation for the clinical trial.

Those two can become comfort when you, for example, using a patient algorithm to update your, like say marginal distribution.
And then the next step is telling like say patients about prediction. And for new patients coming in, you do the marginal distribution estimation again. And does that actually pose a little bit of a problem when prediction actually change people’s mind about their, like say how they attach to the assigned treatment?

Yeah, I think that’s a very valid question. I think that’s the main reason we needed this algorithm, right? And we did not stop just marginally. But if you look into the new data that’s coming in, of course we need to be careful because that’s not coming from a clinical trial data.

And this one is a more simplified because the third step was calculated from clinical trial data, which is a randomized study. But if you add more observational data to it, if I understand how they’re touched it just need to be more careful about using two different quality of data in that way.

Does that answer your question?
990 00:50:10.590 --> 00:50:11.963 <v Attendee>Sure, thank you.</v>
991 00:50:15.240 --> 00:50:19.031 <v Presenter>Okay, so we are running out of time.</v>
992 00:50:19.031 --> 00:50:22.308 so let's thanks Dr. Roychoudhury again.
993 00:50:22.308 --> 00:50:23.558 Wonderful talk.
994 00:50:24.447 --> 00:50:27.030 <v ->[Dr. Roychoudhury] Thank you.</v>
995 00:50:27.912 --> 00:50:30.713 <v Presenter>Please make sure you sign</v>
996 00:50:30.713 --> 00:50:32.551 the admit sheet.
997 00:50:32.551 --> 00:50:36.634 (attendees chattering continues)
998 00:50:52.896 --> 00:50:56.979 (attendees chattering continues)