Dr. Roychoudhury, he had 15 years of extensive experience. He served as the co-chair workshop. And he’s serving as co-chair for DIA, FDA, biostatistics. Dr. Roychoudhury was exacted to be a panel of the American Association. The country international society in 2019. So let’s welcome Dr. Roychoudhury. Now I’m yours. Thank you. Thanks a lot Dr. 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  of clinical trial data,
30 00:01:57.007  especially bringing the different perspective
31 00:01:59.024  as some of you know and some of you don’t.
32 00:02:01.617  FDA actually specifically started
33 00:02:04.927  a patient oriented drug development program,
34 00:02:07.890  which kind of how to make the data more
35 00:02:11.550  understandable
36 00:02:13.950  to the patient, more reachable to the patient.
37 00:02:18.360  Before I begin, just I wanted to mention
38 00:02:21.031  the standard disclaimer, this my own view,
39 00:02:24.515  and not necessarily reflect the view of the Pfizer.
40 00:02:29.700  So often, in critical trial setting,
41 00:02:32.670  we call that analysis over time to event data.
42 00:02:36.180  Because we look into the data that up to time,
43 00:02:40.710  so we analyze the time up to a certain event
44 00:02:44.070  or sensor the data looked at.
45 00:02:46.710  And the standard way of analyzing such data,
46 00:02:51.172  these are the, more or less, a standard tool.
47 00:02:54.390  I’m sure you all have done or are going to do
48 00:02:56.577  looking into Kaplan marker
49 00:03:01.320  which looks into the survival function,
50 00:03:04.161  the log-rank test, which basically tests the difference
51 00:03:07.083  between the two curve, that if the treatment is
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,

which actually kind of increases its popularity.

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 it 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 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.
The one on the right is basically on refractory multiple myeloma disease, which is a very fatal disease.

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 all 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, and nivolumab is one of the immunotherapies of the first (indistinct) cluster for their client. When they looked into the actual treatment effect, it 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 and once it gets into the 18 month, it started to be pretty significant. 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, you can quickly see they’re not constant over time. They’re actually emerging or varying there. I mean, there was been a working group,
there has been a number of workshops here, and even FBA and other regulators got interested.

There has been a lot of discussions started in 2016 when nivolumab, the first class of 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 datas, 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 of the compound?
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,
it 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. 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 into the first part when (indistinct) not separated. 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 into the core value.

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, 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 settings 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 coming into the 12 month zone, but RMST was still showing, okay,

there is no confirmation of such an effect.

So there has been an 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 an 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, \( Y - X \). What is \( Y \)?

So \( Y \) is a survival time at the treatment arm. I’m just making simpler, instead of progression survival,

\[
Y = \text{survival time at treatment} \\
X = \text{survival time in control}
\]

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 of Y and X.
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 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, so 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. 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 (indisitinct) 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 Beysesian 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 was 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 in 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.
591 00:29:20.069 -> 00:29:23.807 and also of statistics, very simply written paper.
592 00:29:23.807 -> 00:29:27.810 And what they were saying that they were looking
593 00:29:27.810 -> 00:29:32.760 into distance between two normal curves there.
594 00:29:32.760 -> 00:29:36.090 And the property they introduced was scale shifting,
595 00:29:36.090 -> 00:29:39.843 a shift function, which basically makes that,
596 00:29:40.992 -> 00:29:43.320 so your X and Y, what you just mentioned,
597 00:29:43.320 -> 00:29:45.450 we said X plus delta X,
598 00:29:45.450 -> 00:29:48.917 which is one can interpret as gain also,
599 00:29:48.917 -> 00:29:52.233 and Y kind of have a same distribution.
600 00:29:53.970 -> 00:29:55.782 That was the main property.
601 00:29:55.782 -> 00:30:00.210 But it's very, basically, that means you project this curve
602 00:30:00.210 -> 00:30:01.457 to other curve.
603 00:30:01.457 -> 00:30:03.543 There is a path to project all.
604 00:30:04.980 -> 00:30:06.713 And the solution of delta X is,
605 00:30:06.713 -> 00:30:10.050 I mean, what can easily cost start that it’s not unique.
606 00:30:10.050 -> 00:30:13.770 But they actually, this is how they constructed it
607 00:30:13.770 -> 00:30:16.950 using the shift function.
608 00:30:16.950 -> 00:30:18.750 But what did that means for our case?
609 00:30:18.750 -> 00:30:20.310 Why we need that, right?
610 00:30:20.310 -> 00:30:22.890 So this means that, so if we,
611 00:30:22.890 -> 00:30:26.410 if I know marginally, the (indistinct) tells of X,
612 00:30:26.410 -> 00:30:29.940 I can project the same content into the Y distribution,
613 00:30:29.940 -> 00:30:32.340 that basically we’ll coordinate ourselves.
614 00:30:32.340 -> 00:30:36.900 So that means if we build, if the ordering of X remain,
615 00:30:36.900 -> 00:30:38.430 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...

I mean, like I said, sometimes simple papers give you very nice ideas as I find.

So 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.

So basically what we did,

we simulated from this posterior distribution and then 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

and we can make the distribution.

But of course, this is questionable, right?

I mean, we are saying the order remains,

which is not always true.

Suppose this is a very classic example

from the nivolumab trial.

So it's attacking the PD1 inhibitor.
So people with PD1 expressed, it’s supposed to work better.

So if you pick two subjects, one is PD1 expressed and PD1 expressed not, if in control, PD1 expressed one actually works worse than the PD1 non-expressed treatment that can reverse.

Because PD1 is the line, the target of the truck.

On such a case, our solution is basically divide the groups into the homogenous biomarker class. And then, the (indistinct) still works.

And that can be easily done adding it as a regression into the model.

So on such a case, our solution is basically divide the groups into the homogenous biomarker class. And then, the (indistinct) still works.

And that can be easily done adding it as a regression into the model.

So then finally, we summarize this with a survival gain and out the loss of tests, I mean basically using the posterior, sorry, the predictive distribution of Y minus X. And then also, we summarized the median and the 95% prediction intervals. And also to us, the Kaplan-Meier plot is still important because that’s where the whole story begin.

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
668 00:33:20.280 –> 00:33:23.220 after all doing fancy things,
669 00:33:23.220 –> 00:33:26.850 make things complicated, making lot of mathematical names.
670 00:33:26.850 –> 00:33:28.407 Did you improve anything?
671 00:33:28.407 –> 00:33:32.160 So did you gain more insight into this?
672 00:33:32.160 –> 00:33:33.540 So let’s go back to this trial
673 00:33:33.540 –> 00:33:36.615 where begin the urothelial cancer trial of atezolizumab,
674 00:33:36.615 –> 00:33:37.890 which is basically the low-rank test says
675 00:33:40.890 –> 00:33:42.140 it was basically...
676 00:33:43.338 –> 00:33:45.780 The survival curves are not separated.
677 00:33:45.780 –> 00:33:49.923 Basically significance that doesn’t reach.
678 00:33:49.923 –> 00:33:53.880 The Cox has a ratio, upper bound is about one,
679 00:33:53.880 –> 00:33:55.770 even the median, because it’s very interesting.
680 00:33:55.770 –> 00:33:58.890 Because the curve actually is separated after median.
681 00:33:58.890 –> 00:34:00.821 So even somebody looking into the median,
682 00:34:00.821 –> 00:34:04.620 the treatment is worse than that for the standard of care.
683 00:34:04.620 –> 00:34:07.920 So the question is really, how can we...
684 00:34:07.920 –> 00:34:10.950 But when we look into the survival differences,
685 00:34:10.950 –> 00:34:13.680 we see a significant survival gain by the patient
686 00:34:13.680 –> 00:34:16.283 who remain on the therapy for a longer time.
687 00:34:16.283 –> 00:34:21.283 But the question is, can we somehow communicate
688 00:34:23.700 –> 00:34:25.893 that better with this individual event?
689 00:34:29.190 –> 00:34:32.406 So the one on the left hand,
690 00:34:32.406 –> 00:34:37.140 one on the left side is basically the probability plot.
691 00:34:37.140 –> 00:34:39.930 So it is basically the first column.
692 00:34:39.930 –> 00:34:44.910 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 start 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 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. 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.
769 00:38:33.146 –> 00:38:35.970 with the (indistinct).
770 00:38:35.970 –> 00:38:38.010 So we looked into all these examples
771 00:38:38.010 –> 00:38:41.790 into the export data and our codes are also relevant
772 00:38:41.790 –> 00:38:44.190 in public if you want to play with that.
773 00:38:44.190 –> 00:38:46.020 Basically the message we got,
774 00:38:46.020 –> 00:38:49.320 so the one on here is that this table
775 00:38:49.320 –> 00:38:52.470 toward the mean survival gain,
776 00:38:52.470 –> 00:38:54.990 median survival gain, the predictor interval.
777 00:38:54.990 –> 00:38:57.450 And what is the chance that probability
778 00:38:57.450 –> 00:39:01.547 that Y, which means your time,
779 00:39:01.547 –> 00:39:05.380 if a subject is receiving treatment
780 00:39:05.380 –> 00:39:08.040 and X if a subject is receiving control,
781 00:39:08.040 –> 00:39:11.046 what’s the time Y is bigger than X,
782 00:39:11.046 –> 00:39:15.360 is basically we started to see some interesting feature.
783 00:39:15.360 –> 00:39:19.200 It gives us much more quantification of benefit compared
784 00:39:19.200 –> 00:39:21.907 to our P value or a hazard ratio in that way.
785 00:39:21.907 –> 00:39:24.960 Of course there are uncertainties in all of them
786 00:39:24.960 –> 00:39:26.460 as well as the clear cases
787 00:39:26.460 –> 00:39:28.806 where we saw statistical significance.
788 00:39:28.806 –> 00:39:32.580 We see some uncertain situation, that patient may
789 00:39:32.580 –> 00:39:36.303 or may not be benefit in some cases.
790 00:39:37.253 –> 00:39:41.640 So the quick conclusion with the increasing complexity
791 00:39:41.640 –> 00:39:45.300 of the drug and the different complex pathways
792 00:39:45.300 –> 00:39:47.349 we are talking, it is very important.
793 00:39:47.349 –> 00:39:50.970 And different patterns of treatment effect,
794 00:39:50.970 –> 00:39:54.900 it’s very important to have a simpler language
795 00:39:54.900 –> 00:39:55.733 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 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, (indistinct) then we’ll be able to see each subgroup benefits from the treatment or like why this (faintly speaking)? (indistinct) comment something else, like the modeling, the causes or figuring out why there’s non-proportional patterns. (indistinct) 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.

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.

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 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 yeah. 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. 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 and we continue to update that, it just need to be more careful about using two different quality of data in that way. Does that answer your question? Sure, thank you.

Okay, so we are running out of time, so let’s thanks Dr. Roychoudhury again. Wonderful talk. Thank you. Please make sure you sign the admit sheet.