WEBVTT

1 00:00:00.000 --> 00:00:03.000 (people chattering)
2 00:00:10.312 --> 00:00:12.012 <v Man>(indistinct) Biostatistics</v>
3 00:00:12.930 --> 00:00:15.107 at the University of Minnesota.
4 00:00:15.107 --> 00:00:18.940 And he's currently attending or an associate professor
5 00:00:18.940 --> 00:00:23.299 at the University of the Texas Dell Medical School.
6 00:00:23.299 --> 00:00:27.160 Dr. Hobbes is a library recognized as an expert
7 00:00:27.160 --> 00:00:31.160 in clinical oncology and research (indistinct).
8 00:00:31.160 --> 00:00:33.743 Among his many accomplishments,
9 00:00:35.309 --> 00:00:39.392 in 2017, Dr. Hobbes (indistinct)
10 00:00:40.400 --> 00:00:44.039 of The National Cancer Institute, clinical trial
11 00:00:44.039 --> 00:00:49.039 (indistinct) a national consensus recommenda-
12 00:00:49.680 --> 00:00:50.513 or (indistinct).
13 00:00:54.840 --> 00:00:59.840 In 2019, Dr. Hobbes (indistinct).
14 00:01:01.000 --> 00:01:13.424 In 2021, Dr. Hobbes (indistinct)
15 00:01:46.510 --> 00:01:47.650 <v Brian Hobbes>Thank you, thank you very much,</v>
16 00:01:47.650 --> 00:01:50.893 and for that long and generous introduction.
17 00:01:53.580 --> 00:01:55.410 I’m excited to give this talk today.
18 00:01:55.410 --> 00:01:57.740 I wish I could visit in-person.
19 00:01:57.740 --> 00:02:00.260 I was fortunate to have that opportunity a few years ago,
20 00:02:00.260 --> 00:02:02.483 so thank you for inviting me back.
21 00:02:04.810 --> 00:02:09.530 Okay, so I got tired of giving talks
22 00:02:09.530 --> 00:02:12.720 that were very technical and very specific
23 00:02:12.720 --> 00:02:14.133 to a specific problem.
24 00:02:15.045 --> 00:02:17.697 Because, if you don’t have an understanding of the problem,
25 00:02:17.697 --> 00:02:19.450 you don’t have an understand that the biomark-

1
or you don’t work in a particular area of methodology,
I think it becomes very, you know, what do you wanna say?
I think people lose interest pretty quickly.
And so I decided to start giving talks
that had to do overview a subject
that I think is very relevant in the field right now.
So recently, I’m an external advisor on a grant
that Genentech has got from the FDA
for developing clinical trials that use real-world data.
I’ve worked with Flatiron in the last few years,
as well as the CancerLinQ.
And I’ve been watching this field,
sort of the discussions in this field
about real-world evidence,
and where does it fit-in?
and specifically in the context of cancer drug development.
So I decided to talk about that today.
So yeah, I think this is what I’m gonna do.
So does real-world evidence
have a role in cancer drug development?
So if you’ll see,
there’s a question mark at the end of the statement.
So I’m gonna talk about this,
and I’m gonna give you the perspective that I have,
which comes from a methodologist
that really wants real-world evidence
to have a role in cancer drug development, because, databases are growing. Data science becomes more relevant if those databases are useful. We all want to write algorithms and do, you know, causal inference on database. We want to unlock those databases with our intelligence, for drug development. Drug development is incredibly expensive. Patients need access to therapies that are gonna save their lives. We have refractory patients enrolling in clinical trials. There’s probably not enough clinical trials. And we do have advances in biology that have manifest themselves in precision therapeutics. So we want all of this to work together. We want this to be true. On the other hand, I have designed hundreds of clinical trials and I continue to, most of my collaborations continue with MD Anderson in this space. I’ve worked with oncologists for over a decade now. I’ve worked with translational researchers in oncology, and I see the issues that are presented. I mean, maybe I should say the challenges, the challenges that we confront
when we think about this space. So I’m gonna talk about this.

And I think if I was the 30-year-old version of myself, there would not be a question mark here. I would be saying we can use real-world evidence, and this is how. But now that I’m 40 years old, there’s a question mark.

I’ve seen several talks at seminars, conferences, where people are presenting very specific cases. Specific cases where they could use real-world evidence and it made sense, or it was the only thing that could be done in that context. So I’ve seen a lot of talks like that.

I’m gonna take this from the other perspective, I’m gonna talk about what’s going on in oncology right now. What are the most important developments happening in oncology?

And then I’m gonna ask the question, can we use real-world evidence to help augment
our trial designs and drug development in general?

So to begin with what is real-world evidence?

So there’s different definitions of this. It tends to be a very broad definition.

That’s, you know, that different people use this.

I’ve taken this diagram from the CancerLinQ, which is a nonprofit organization that works with the American Society of Clinical Oncology.

They are massing and organizing a large database,

I collaborate with Elizabeth Garrett-Mayer at CancerLinQ who’s at ASCO, who’s great.

Who’s have a PhD statistician working on this.

So this diagram, you know,

what we often think about as real-world evidence,

we think about as the electronic medical health records

that are in sort of community hospital systems, right?

We think about data that’s acquired from routine care

or from claims that’s not on patients

that are in a clinical study.

We tend to think about that as real-world evidence.

And so CancerLinQ says Real-World Evidence has a capability,

data tools, processes, organization, underpinning functions
to drive business intelligence.
So that’s kind of, you know, very broad.
They also tell us that there’s other things that should count as real-world evidence beyond the EMR data.
Okay, observational data as well as historical randomized controlled data.
Okay, that makes sense.
Pharmacy data, mortality registries, hospital visits,
lab values, claim databases, social media,
they put on this diagram as well.
So you know maybe, right?
But I think that we’re at a place right now where people are excited about using these sources of information in research,
but somebody really needs to develop a framework
for each of these.
There’s not a single framework that says,
this is how you use all of these in a clinical research program.
If you’re gonna use social media in clinical study
for research purposes, you know,
there needs to be a framework for how to do it,
especially in the context of precision oncology.
But so we have this,
and we have groups that are working on these databases,
they want to make this a realization.
What are the regulators saying?
Well, so real-world data and real-world evidence really got a boost from the 21st Century Cures Act signed into law in 2016.
They advocated for the use of real-world evidence to support new indications for approved drugs.
Of course, the US Government wants the innovations that we have in biology to translate into therapeutics for patients.
And we have a very, you know, forward looking approach when it comes to that,
if the drug is relatively safe and can demonstrate some efficacy it gets to the market, it gets to patients.
So that there’s a guidance document about the use of real-world evidence to support regulatory decision-making, which was initially for devices.
There’s another one for biologics in 2019, there’s actually a website you can go to, which is the framework they discussed in 2018.
If you go to that website, and this was done, they have quotes from Scott Gottlieb here.
You can see that a little more of a definition, real-world data can be used to improve efficiency of clinical trials, even if it’s not used
for product effectiveness.

So the FDA is still saying,

we don’t want to use real-world data as a control arm

to replace a randomized control, for example,

but we could use it to generate hypothesis, right?

What is the expected event rate

for this population that we’re enrolling?

How many events do we expected

to have in a certain timeframe?

How likely is it that we can roll that population.

Trial feasibility and forming prior distributions

in Bayesian models.

So I liked that observation,

but, you know, what is our expectation?

Maybe we’re not starting from nothing.

And then prognostic indicators.

Are there things we should stratify for

or account for an analysis that could be imbalanced,

especially when we don’t randomize?

So this is what regulators are saying,

but they also say the standard

for drug approval remains the same.

And this is an important statement.

The basis of approval remains the same.

Substantial evidence that the drug will have the effect,

and adequate well-controlled clinical investigations.

So. and I was just at a meeting at UNC
with Genentech and FDA and people from the EMA,
and they’re standing firm on this.
While we’re discussing how you could potentially augment a randomized-control
with real-world controls,
there’s no sort of interest in replacing randomized-control right now.
Not unless there’s absolutely no ethical way you could randomize.
So they say that, you know, there’s more flexibility when the disease is rare,
and the patient population lacks a suitable control.
So what about the CancerLinQ?
So these slides are a little dated as of the last year, March of 2020,
but they had at that time over two and a half million patients in their database.
So they have worked on data codes and structuring outcomes, structuring CON-MED data.
They’ve done a lot with this database, and I used it at Cleveland Clinic.
So this is growing as a resource.
Also what happened is that Flatiron, which has over 2 million active patients in their database.
Of course, this is an industry group that’s partially owned by Roche.
They have partnered with Foundation Medicine, and now there’s an intersection of Flatiron patients that also have genetic testing from Foundation Medicine. And they’re calling this the Clinical Genomic Database. At the time that I took this slide, they had over 40,000 patients that had the real-world data matched to the molecular data. I think that’s very interesting, and I think that’s very important. One of the main issues with real-world evidence in the oncology setting is that we don’t have a real-world tumor response. So for those of you that work in oncology, of course, you know that phase one, phase two trials are designed on the basis of endpoints that measure reductions in tumor burden. So for solid tumors, this is done through scans. So patients are scanned at baseline. They’re scanned regularly at follow-up intervals after every visit or every cycle of therapy. Those scans go for an adjudication process, which is done by more than one person where they actually measure how much reduction in tumor burden happens after treatment.
And then we look at that longitudinally, we take the best reduction or the most reduction that we saw, we consider did they have distant migration of disease? So for example, if they had a brain tumor, did they also come in with tumors in their liver? And then we come up, we have a four point ordinal scale, and it tells us whether the patient has a complete response, which means the tumor burden’s gone, right? The lesions are gone, or the blast counts in their blood are gone, if they have leukemia. They had a partial response. That means there was a reduction in their tumor size, and it was a clinically meaningful reduction. They had stable disease, which means that there could have been a reduction, but it wasn’t clinically meaningful, and it didn’t really increase. And progressive disease, the tumor burden is much higher than it was at baseline. So this process is critical for understanding making decisions in phase two trials, as well as now the phase one trials that we have in oncology, which are very large. This forms the basis for many go-decisions.
Whether you continue to develop a drug, did it have a local effect on the tumor burden? It's very expensive to do this. It's very difficult to do this. So now we have to think about how can we get this information from an EMR? Certainly patients may have scans in an EMR that we could use, but there's several issues with that. So if we're going to use scans in a database to assess a patient's tumor burden, number one, those scans don't go for a central review. The process by which the community or the non-trial evaluation of those scans is very different than the clinical trial process. They don't really have an ordinal scale like this that they use. Certainly, I think you could distinguish progressive disease from complete response. I think it'd be very difficult to distinguish partial response from stable disease. So we have groups that are saying they can do this, right? They're going back to the clinical annotations and the writing algorithms that look at the clinical annotations that says, well, if the notes say the lesions are all gone, then they had a complete response, right?
If there was an increase, overall increase, or there was new lesions, they had progressive disease.

So if the annotations are good enough, I guess, you could get to progressive disease versus complete response.

However, there are several issues with this. Everything in oncology is based on the line of therapy.

Patients come in, they get a sequence of treatments. Usually, they progress and go to a second line of therapy. Or they progress again and they go to a third line of therapy.

The expectations for tumor response as both survival and response by that are very different by line of therapy.

So if you’re gonna go into the EMR, you have to now make sure that the scans you’re getting align with the line of therapy that you’re enrolling in your clinical study.

So most clinical studies in oncology require a specific line of therapy.

So first-line therapy means patients that haven’t been treated previously. Second-line therapy means patients that have progressed on a prior treatment, and now they’re trying a subsequent treatment.

So the expectations are very different for response by that.
So you would have to know that this is the first, if you’re using a first-line therapy study, you would have to know that this is this patient’s first line of therapy and these scans correspond to that. Not only that, you’d have to make sure the scans reasonably aligned with the time-frame by which the clinical trial is actually going to acquire their energy. Beyond that, you’d have to, you know, there are several other issues with that, right? Patients may not be scanned in the community setting. And working with oncologists for a long time, I know that there’s a certain point where if a patient fails a few lines of therapy, they may not wanna risk the patient getting nephrotoxicity from the contrast that are used in the scans. So if a patient doesn’t have a lot of good treatment options or they’re reasonably unhealthy, where there’s concern about kidney or liver issues, they don’t scan the patients in the community. So up till now, I think that the consensus has been, there is no real-world tumor response right now.

We don’t have that. And I think that’s difficult because
we want to use real-world data to sort of augment or supplement the areas where we don’t have a lot of information, right? And that is the early phase studies, right? Once you go to phase three, you’ve kind of established that the drug may be promising and you’re gonna run a seven-year trial. And over that seven years, you’re gonna acquire lots of information, and you’re gonna follow them for survival. This could, with really the narrative about real-world evidence in oncology, has really been we can supplement those early phase decisions.

But to do that, we really have to have a real-world tumor response. And right now we don’t have it. This is a paper from Advanced Therapeutics that was published this year. We have the Flatiron group going back to the major immunotherapy trials that have been implemented in recent years. They’re comparing their algorithm for real-world response rates with the trial confirmed response. So they’re saying, for each patient, what did we say the response was?
Based on our EMR data? What did the trial say the response was? And they’re looking at sort of coordinates between those measures. And they’re doing this by line of therapy. So maybe we’ll get there, but right now the consensus is we’re not there. So we presented this paper at ASCO, which is a big cancer meeting in the US last year, talking about, can we actually replace randomized controls with external real-world controls? And we actually built some tools that Genentech is using that actually calculate, that takes your assumptions about bias, heterogeneity, or other things that you might see in a trial and actually tells you how wrong you can go with a go-decision when you use an external control. And of course, I think maybe everybody knows this, that the reality is that if there’s no bias, it’s useful. If there is bias, things can go really wrong very quickly, depending on the direction of the bias. And that is really unknown. So we tried to think about this in a very systematic way, and I think it’s challenging.
I don’t know that we can do this. So that leads to, you know, what this discussion was at UNC with the FDA, the EMA, and Genentech where we’re talking about now, can we augment randomized control arms with data from real-world sources? We don’t get rid of the randomized control, we keep the randomized control, but we supplement it with some external controls. How could we do that?

And could we even acquire those before the trial gets initiated? Of course, it takes a long time for protocols to be reviewed and other things to happen. Well, this gets interesting to me because while I developed tools to do this a long time ago, which I called Multi-source Adaptive Designs. And this was done many years ago before we talked about real-world evidence. We were talking about historical controls at that time, but of course we can do interesting things with modeling here. We could take real-world controls, we could think about an interim analysis of a randomized trial, where we have randomized treated and randomized controls.
We could do any sort of fancy model that you wanna fit, and we could assess how biased are these historical controls or real-world controls in relation to the control data that we’re seeing in the actual randomized trial. On the basis of this model, we could actually adapt the allocation, right? If we don’t see a lot of bias, so those patients, based on the eligibility of the trial, those patients from the community, they look a lot like the patients in the trial, then you have more information on the control side. You need to rebalance the rest of your allocation so that you can increase power. So this is the only designs where you can actually increase statistical power with a smaller trial. Because what we’re trying to do, is we’re trying to balance the overall information between the treatments, right? If you look at the outcome, adaptive randomized studies, they required larger trials because they’re imbalancing. They’re imbalancing based on outcomes. We’re trying to balance based on bias.
So we worked out this methodology and you know, ASCO and Flatiron are interested in using this. We have a paper that describes an open-source tool that we have. It’s still on MD Anderson’s website that I built when I was at MD Anderson with Nan Chen.

Who is pictured is here. So Nan is now at Gilead. But if you interested in this, it’s here. So I think based on in oncology setting, we need to focus on this area. We need to focus on hybrid controls, not replacing control arms, right? At least for most studies. Of course, in rare diseases or areas of pediatric cancer, or both, you need to do something else, right? And that’s what the FDA is talking about when they talk about flexibility. But I’m talking about from kind of the standard drug development program in oncology right now.

So I’ve talked about the issues, I’ve talked about the databases and sort of what’s going on with real-world data in oncology. There’s another group of sort of players in the space. And I would call them
kind of the real-world evidence zealots.

Dr. Butte from Stanford has, I think represents one of these people. So he is a strong advocate for using databases to replace clinical research.

He has at least three TED Talks, and I was going through them yesterday. He has a very strong feeling that we just need to organize these databases, and we can answer any medical or scientific question and in fact, he even says, the problem is there’s not enough people asking questions.

That’s the real issue right now.

So there’s this other group of people that are you know really hyping up the fact that it’s just a computing problem. That we have the data, we can use algorithms to answer any question we want.

This group of people seems to lack any recognition of the principles of experimental design. They don’t seem to acknowledge them anywhere in the process. And Dr. Butte and his TED talks actually says that we don’t need randomized controls after all, we just need to build databases.
so these are the kind of the players pushing this forward.

So now I’m gonna transition here.

I’m gonna talk about what’s going on in precision oncology.

Okay, so this is how you learned about drug development programs.

You learned that we chose dose in phase one, if the dose was promising and we were able to discover what the MTD was, and we felt like it wasn’t toxic.

And we had a good dose,

we would go to a phase two trial.

In oncology, we would look at tumor response.

So again, reduction in tumor burden.

Usually these would be uncontrolled.

They would be about 50-100 patients.

If we saw the drug had local activity on tumor burden,

we would go to a phase three trial.

The phase three trial would randomize to the existing standard of care.

And would see if the treatment prolonged survival.

This is what you learned about,

but oncology has changed very rapidly.

Regulatory policy has changed as well.

So molecular biologists have some victories recently.

They have really, you know, a lot of the biological models

that were discovered a decade ago.
have been translated into therapeutics. So it used to be that we needed one or two well-controlled phase three trials before we got regulatory approval. It turns out that cancer biologists have identified very specific cancer subsets based on genetics and based on immunology. With those cancer subsets, we have seen very promising, very exciting results in phase two trials without controls. The FDA started to allow conditional approvals after phase two on the basis of those biomarker targeted treatments. Now we’re in kind of stage three here. Now we have the awareness that many of the targets, many of the genetic targets, as well as the immune phenotypes that we’re interested in, they actually exist across several different sort of traditionally distinct cancer patients. So patients with pancreatic cancer and lung cancer may be very different from a clinical perspective, but they might share a molecular feature that can be targeted by the same drug. And we’re now in the space of histology-agnostic drug development, where we might be replacing traditional classification criteria.
based on molecular features. So we're basically finding new subtypes of cancers as we go.

These subtypes are very small, and they're becoming smaller as we learn more about cancer biology. But a few of them had had very exceptional results. A few drugs targeting these event, have had very exceptional results, crossing many tumor types. And they have gotten accelerated approval for agnostic drugs and drugs that can be administered without regard to the tissue of origin. This has happened in phase one. So the regulatory landscape has changed, the development landscape has changed. So I got to be a part of this review for Nature of Clinical Oncology, where we talked about these tissue-agnostic drugs.

There's actually four drugs so far that have been approved by the FDA that could be administered based on a marker feature, not on the actual cancer tissue. So now look at the issues with this. There's four drugs, and there's three different biomarkers that have been approved for tissue-agnostic treatment. One of the biomarkers is the NTRK fusion,
which we’ll talk about little later.
It’s exceedingly rare.
You can see that breast cancer,
we’re talking about less than 0.1% of the patients have an NTRK fusion, right?
And CRC it’s about 1% of patients.
There’s a few tumors where it’s more common,
but this becomes very challenging.
It becomes very challenging to design a study
where we can actually study patients with NTRK fusions.
And then who are you gonna get in your study?
You’re going to get a mixture of many different tissues
that were traditionally thought to be separate cancers.
So with this transition to tissue-agnostic drug development,
there’s a statistical question that we have to answer,
and that is who can be averaged?
Which tissue types could be averaged statistically,
when we assess the effectiveness of a biomarker targets
and a therapeutic?
And that’s the question of statistical exchangeability.
So we have developed patient models
that actually characterize
what subsets of tumors actually respond in a similar way
to a targeted therapy.
And this gives us statistical criteria for understanding what is agnostic and what is not.

And I got the, you know, this is the first time, I got to collaborate with Dr. Kane on actually building out tools for this.

So I can do the methods, but the tools or something else.

So Michael got these incredible tools. And we have an open source package for fitting these models.

Just to give you sort of motivation here. This is an actual trial that was evaluating a drug called Bendroflumeth in BRAF tumors, patients that have BRAF mutations.

So there is BRAF mutations can occur in many different tumors. They initially developed this drug in Melanoma, but then they saw BRAF tumors, BRAF mutations exist in these other cancers. Histiocytosis, thyroid cancer, cholangiocarcinoma, for example. So they ran up, what’s known as a Basket Trial, where they allowed these different tumor types in the same trial. So we show in this nature of these clinical oncology paper, how these exchangeability models work.

We call them multi-source exchangeability models.
Where we start with an assumption that these tumors are gonna act in the same way, right?
So the drug target combination is going to be kind of equally efficacious among all the tumors.
So they’re exchangeable statistically.
We can average them.
As we start to get data from the trial, we can now start to assess the heterogeneity that we see across these tumors.
And we can ask the question, is it really agnostic to the tumor type?
Now, when it comes to vendor afatinib, we had three tumor types that did really well in this trial,
histiocytosis, thyroid, and non-small cell lung cancer.
Colorectal did not do well.
So even though colorectal cancer patients had BRAF mutations,
they did not respond to vendor afatinib.
These tumors did respond.
We don’t know about cholangiocarcinoma.
There wasn’t enough information in that trial to really tell us.
So they’re kind of in the center here.
You know, this is just to give you a flavor of what’s going on in oncology right now,
As we start to go towards precision medicine,
that means that we have features across traditionally
very different cancers. And we have to understand whether it’s actually the feature that’s driving, what we see in the response. Okay, so this is an issue that I don’t think is that well understood outside of our sort of biostatistical and statistical communities. And that is how, just the extent to which prognostic heterogeneity plays a role in the precision oncology space, or any space where you’re doing biomarker driven therapeutics. So what I’m showing you here is the cancer immunity cycle by Chen and Mellman. This diagram sort of revolutionized how we think about how the immune system identifies and counteracts malignant cells. So cancer cells release antigens. They have to be detected by the immune system. If the immune system detects antigens, means your immune system is actually aware that you have cancer. They have to produce natural killer cells. They have to infiltrate the tumor. They have to recognize which cells are malignant cells,
and then they have to kill the malignant cells. This process is very complicated and there are biomarkers that can tell us about what’s happening with the patient. What’s happening with the patient’s innate immune response to cancer. So the biomarkers that have been most developed recently are the PD-L1 biomarkers, which is this last step. So if a patient is expressing a lot of program death like in one, it means that the malignant cells are actually hiding from the T cells. So the patients might be producing lymphocytes. They might be getting to the tumor, but they can’t attach. They can’t identify which cells are malignant cells, malignant cells are hiding. So there’re very interesting things that happen when you get to a biological perspective. The immune phenotypes based on these biomarkers. If we look at T-cell infiltration versus PD-L1 expression. Patients that are producing T cells and that have low PD-L1 expression. So that means T-cells are being produced, they’re coming to the tumor
and then they’re effective when they get to the tumor.

These patients have a different immune profile, where patients are not producing T cells. So it’s like their immune system isn’t aware that they have cancer.

And then even if they did produce T cells, they’re not effective once they get to the tumor.

So there’s various things happening in this phase. And so now I go back to Professor Butte and sort of what he’s saying, there’s several articles that he’s written that say things like this, precision medicine makes doctors nervous. The reason that makes doctors nervous is because they have to admit that what they were doing was not precise.

So we see these things and we see these kinds of narratives coming from the group that’s really pushing that we just need to analyze these databases. So he’s talking about retroactive crowdsourcing, right?

A high school kid can do it. So if you’ve listened to his talks, he’s always saying, a high school kid can do that.

A high school kid could do this.
I think a high school kid could apply a T test to a dataset.

I don’t disagree with that.

But I have a 14 year old at home and he has trouble making his bed.

So I think that there’s a narrative out there that doesn’t recognize things like this.

So when I was at MD Anderson, we spent a lot of time thinking about these immune phenotypes.

And I actually developed radiomics models, that characterized patterns that we saw in images in the tumor that reflected these immune phenotypes.

And the reason we were doing that, is because these biomarkers were incredibly unreliable.

What I’m showing you here is a scatter plot, that this came from the Garcia student’s lab at MD Anderson, probably the best immune pathologists in the field right now.

These are patients with non-small cell lung cancer.

They all got treated with definitive surgery.

So there was no chemotherapy.

They came in, they could be treated with surgery.

So we don’t have sort of a confounding factor here with these patients.

We got their tissue microarray staining, and this was both malignant cells and immune cells.
are PD-L1 positivity at biopsy.

So the patients are coming in, they’re getting a biopsy.

The biopsy is taking a needle, sticking it in a few different locations.

We use that tissue and we try to assess how much PD-L1 expression do they have and their lung cancer?

Then they go in, they had surgery.

We took their whole excise tumor.

And we went back and we did whole section staining.

This is a scatterplot we got.

So each point is the same patient.

This patient at biopsy, just this isn’t the worst one, but this patient at biopsy was over 50%.

After surgery, they’re only at 15%.

This patient is much worse.

So what’s going on here?

Either the immune system is constantly changing and these biomarkers are not reproducible,

in the sense that your state is changing, or when we stick that needle in we take just a few points,

we get a very different answer than when we do surgery.

Of course, we have to use biopsy if we’re gonna make a treatment selection.

So this is problematic.
So when I think about, you know, we just need databases, we don’t have to understand the science and we can answer all these fundamental questions, I don’t think it’s true. You know, you have, There’s issues like this with every biomarker. The biomarkers have to be reproducible. We have to understand them in a rigorous manner, if you’re going to use scanning data. So, you know, so we’ve published this paper in scientific reports. It has been cited I think almost a hundred times in a few years. Where we actually developed a radiomics model for understanding the immune pathology. Now, why did we do that? We did that because we didn’t think these biopsy assessments were reliable. So we thought that maybe the scans were more reliable. Maybe we could take the scans and we can understand the patterns in the scans. And you can see that patients with different immune phenotypes, but in terms of T-cell infiltration and PD-L1, they had very different expectations for survival.
So this is not a treatment effect. This is just simply the impact of the fact that the patients have different immune systems. And those immune systems have differential effectiveness in fighting the tumor. So patients that have T-cells and low PD-L1 positivity, they're doing well. The opposite is true for patients that have high PD-L1 and low T cells. So we developed a radiomics model, which take the scans and actually assess these patterns. Of course, there’s complications with that. If you’re to scan any data in oncology, you’re probably having contrast. You need to understand what the protocol for contrast was for that scan. Because you need to take the image when the contrast is in the tumor. So of course you can’t just go blindly and grab a bunch of images from a database. So, I’ve talked a little bit about what’s happening on precision oncology. Where we’re developing biomarkers, we want to use to guide treatment, but it’s very complicated. And I don’t think doctors are scared because they’re not precise, they’re scared because we need to understand.
that these biomarkers and make sure they’re reliable and reproducible.

And that knowledge is important.

Not only that, but because of all this complexity, drug development on oncology has changed a lot.

And we no longer have this, phase one to phase two.

This is what early phase drug trials look like now,

especially for the big companies that have a lot of money to invest.

They’re taking multiple dose levels from dose expansion,

they’re running massive dose expansion cohorts.

Those dose expansion cohorts, usually span multiple tumor types.

And they might randomize across dose level,

because we don’t have these very clear monotonic relationships between dose and toxicity anymore.

And selecting a dose isn’t as simple as it used to be when we did cytotoxic drug development.

So these non cytotoxic targeted therapies, it’s hard to select a dose.

These dose expansion cohorts can be hundreds of patients.

They may not even stop for a phase two trial.

They may go straight to phase two and expand on the expansion.
Or they may skip phase two altogether because they’ve already acquired so much information in their phase one trial. So this is what we see happening now. Of course, the keynote trial evaluated in Pembrolizumab had eight expansion cohorts. There was over a thousand patients in this first in human phase one trial. This trial is what motivated that NCI Clinical Trial Design Task Force, that I got to be a part of, because this was a massive departure from what we saw typically in oncology and how IRBs would review these studies. More recently, Genentech drug (indistinct) had a phase one trial with nine expansion cohort. Looking at the dose, expansions alone, the bladder cancer cohort had 97 patients, and they randomized the three dose levels. So this is a new world. 97 patients already in their dose expansion. So this is where Master Protocols come in. So we have innovations in design that are sort of targeting this. The other thing that’s happened in oncology is that phase three continues to be poor. So phase three trials continue.
to have a poor track record relative to other areas of medicine. You can see lots of articles that described this. Of course, Gan et al did a review of 235 published randomized controlled trials. Regulatory approval was, you know, less than 38%. And what’s happening? While the investigators are not very good about making the assumptions for that phase three trial, we see a lot of phase three trials in oncology that have unrealistic expectations. Okay, so now I talked about precision oncology. I’m gonna go into some case studies that I think are interesting. And I want you ask the question, how could you have used real-world evidence to change what happens here? So this is coming at it from, these are the high profile trials that we have been running in the last few years in oncology. We want to know, how could we have used real-world evidence in these settings? So I’m gonna talk about Atezolizumab. Atezolizumab is another PD-1 inhibitor. So it was developed for many different areas.
Again, we're talking about tissue-agnostic here.

So it's targeting a feature of the immune system, that feature of the immune system can exist across many different tumor types.

They evaluated nine in their phase one trial. So after the phase one trial, they ran a bunch of trials and different types of cancers and different lines of therapy.

One of them was second-line bladder cancer. So these are patients with bladder cancer that have progressed on a prior therapy. So they already progressed on chemotherapy, now they're getting this immunotherapy.

So they ran this study and the biomarker they're targeting is they're calling IC2/3. That is immune cell staining of PD-L1. And those immune cells have 5% or more expression.

So 5% of the immune cells that they stained had, at least 5% had Programmed Death Ligand 1.

That's their target.

So, but they enrolled in this phase two trial, they enrolled all comers. It wasn't restricted to the target. They enrolled all comers.

So the IC2/3 population is their target.
That’s where the mechanism is supposed to work. So among a hundred patients with that target they got a 26% response rate. You can see patients that don’t have the target, there was 11 and there was eight. And if you look back at their paper, they told us that they expected 10%. So they said the null hypothesis was 10% for this population. We got 26%. This is very exciting, right? This is the survival curves that they present from their phase two trial. Again, this is uncontrolled. There’s no chemotherapy arm here. This is just the treated arm, Atezolizumab by biomarker status. And when you look at this, you see this blue Kaplan-Meier curve, that’s above everybody else. That Kaplan-Meier curve is the target feature. That’s the IC2/3 population. So they’re responding, their tumors are shrinking and they’re living longer. It looks like this is very promising, right? On the basis of that, they got accelerated approval. And that was given in 2016.
And the reason was increased levels of PD-L1 expression on immune cells are associated with increased response. Let’s go to the phase three trial. So as a part of the conditional approval with accelerated approval, they have to run a randomized phase three trial and sort of replicate this result. So they designed this trial, IMvigor211, multi-center open-label phase three trial. They compared to three chemotherapies, which were standard chemotherapies used at the time. So there was a physician’s choice. If the patient was randomized to chemotherapy, the physician would choose which among these three chemotherapies. Which among these three chemotherapies. So what happened? We had this blockbuster results in phase two, but there was no difference in overall survival in phase three. Not only was there not a difference in overall survival, the objective response rates were similar. So the tumor responses were similar. Moreover they enrolled 931 patients and only 234 actually had the target. 24% of the trial was used for the primary analysis. When we look at the data, what happened?
23% of the IC2/3 population responded. So that’s close to 26%. It looks like that was replicated.

When you look at the intention to treat populations, that’s everybody here, regardless of biomarker, it’s 13 and 13. So it was also lower without the target. But what’s happening with chemotherapy with the target? It’s 22%, right? So chemotherapy is doing great with this biomarker.

So this biomarker profile is doing just as well as the targeted therapy, when the patients get the standard of care. Here’s the survival curve. Okay, proportional hazards is probably violated. There is a heavy tail here for the Atezo group. Maybe there’s, it looks like there’s some long-term stable disease, people that are benefiting. But overall, this is not significant.

And on the basis of this, actually this year, this drug was withdrawn from accelerated approval. So it got the accelerated approval, which was for very exciting drugs that need an accelerated pathway for regulatory. And then this phase three,
on the basis of this phase three trial, they had to withdraw from that. So the question is, how do we use real-world evidence to change this? At the end, there were flaws in this design. They didn’t understand the biomarker. They didn’t understand the biomarker profile on the basis of the standard of care. So when I first got involved in sort of, well, over this past year, I’ve been thinking how could we have used real-world evidence? Here’s the case where, you know, there’s, it’s kind of a failure of the system here that we had this drug withdrawn from accelerated approval. And it’s not the only one, by the way. Is there something in the historical data or the real-world data that we could have used that we had this drug withdrawn from accelerated approval? And it’s not the only one, by the way. Is there something in the historical data or the real-world data that we could have used that could have informed us to design a better trial, or could have told us something about the fact that this biomarker may be prognostic? Now it gets complicated because actually it’s not prognostic for surgery. Patients that have surgery that have IC2/3 status, they’re going to die sooner than patients that have IC1, IC0,
So this marker seems to be a predictive marker for both chemotherapy and for Atezo. But we didn’t know. So my postdoc and I went back and we did a meta-analysis. We went and we extracted all of the trials that had enrolled second-line bladder cancer patients in a prospective study. When we did this literature search, there were like 200 papers on second-line bladder cancer. Most of them were retrospective reviews and case studies, the overwhelming majority. There were only 11 that were actual prospective studies. We’re looking at the overall response rate.
from these 11 trials, we’re doing a standard meta-analysis. You can see that Genentech said 10% was their null, right? And really the case for real-world evidence is you can do a better job specifying your null hypothesis. Your null hypothesis can be specified better because you know what to expect for control. So based on our meta-analysis of the objective response, 10% is really good estimate. And 10% is like the hierarchal mean of this meta-analysis. So now we go to the chemotherapy arms that we saw in the Atezo trial. We see the IC0/1 population is right at 10%. But look at this, this IC2/3 population. Again, this is with chemotherapy. They’re statistically significantly better than the hierarchal mean that we estimated from meta-analysis. So what does this mean? This means that this profile has not been studied before. These trials are mixtures of different immune phenotypes. So we don’t know which mean phenotype they’re studying. They have a different distribution. Maybe this one has more IC2/3 population because it’s pulled over.
But the reality is the information in these historical studies doesn’t tell us about immune staining. So this is a biomarker that wasn’t studied before. And certainly that’s gonna be the case in the community databases. Because there’s only a few institutions that really can have the infrastructure to quickly stain these patients. So we extracted the Kaplan-Meier curves from these historical studies. And we did a meta-analysis of these Kaplan-Meier curves. When we put the Kaplan-Meier curves together with the survival data, we see this purple line is the chemotherapy arm with this targeted biomarker from the phase three study, that was implemented by Genentech. So responses is better, survival is significantly better than what our expectation was based on historical evidence. And we actually went back and did simulation studies.
where we fit piece-wise exponential and Weibull curves,

till all of these Kaplan-Meier curves that we extracted from the web digitize the tool.

When we actually simulated, was it the probability of success

And we looked at that for the PDL-1 population,

as well as the ITT population.

We only give this trial 20% chance of success based on the extent to which chemo is interacting

If you like, if you wanna account for the heavy tail

that we see in the model, and do piece of exponential,

it goes up to 24%.

So another case of a phase three trial

there was, had unrealistic expectations, right?

And it’s a case

where we didn’t understand the biomarker profile.

That biomarker profile had not been characterized

in the historical evidence.

It’s not only Atezolizumab, this happened to Durvalumab.

It happened in bladder cancer for Durvalumab again as well.

Also a PD-1 inhibitor from AstraZeneca.
1161 00:50:30.664 --> 00:50:33.747 So Precision Oncology is hard, right?
1162 00:50:34.610 --> 00:50:35.443 It’s hard.
1163 00:50:35.443 --> 00:50:39.210 It’s not, what I presented here,
1164 00:50:39.210 --> 00:50:43.750 was not really about the lack of having information,
1165 00:50:43.750 --> 00:50:47.040 it was a lack of having the biomarker target characterized
1166 00:50:47.040 --> 00:50:48.683 in prior research studies.
1167 00:50:49.710 --> 00:50:50.950 And without the understanding
1168 00:50:50.950 --> 00:50:55.950 that profile could be predictive for the standard of care,
1169 00:50:55.210 --> 00:50:59.950 we have these drugs withdrawing from accelerated approval.
1170 00:51:00.030 --> 00:51:01.310 There’s other issues when we look
1171 00:51:01.310 --> 00:51:03.210 at tissue-agnostic development.
1172 00:51:03.210 --> 00:51:06.140 So I worked with Bayer and MD Anderson last year
1173 00:51:06.140 --> 00:51:09.360 to investigate and NTRK fusions.
1174 00:51:09.360 --> 00:51:12.260 This is that rare biomarker profile
1175 00:51:12.260 --> 00:51:16.810 that has led to two drugs getting tissue-agnostic approval,
1176 00:51:16.810 --> 00:51:18.850 larotrectinib and entrectnib
1177 00:51:20.410 --> 00:51:24.330 So Bayer bought larotrectinib from LAKSO
1178 00:51:25.520 --> 00:51:28.173 and Roche bought entrectnib from Igniter.
1179 00:51:29.120 --> 00:51:31.330 So they wanted to understand,
1180 00:51:31.330 --> 00:51:35.690 and this was used actually in the Canadian approval process.
1181 00:51:35.690 --> 00:51:37.540 The Canadian approval process is different.
1182 00:51:37.540 --> 00:51:39.160 You have a higher level of threshold
1183 00:51:39.160 --> 00:51:41.660 that you have to characterize
1184 00:51:41.660 --> 00:51:44.750 for biomarker targeted therapies.
1185 00:51:44.750 --> 00:51:46.600 And they wanted to know specifically,
What is the evidence that NTRK is a prognostic marker?

How do we know the drugs working and when it may just be the profile is favorable?

And that’s kind of exactly what happened with Ateza.

So we thought that maybe we could interrogate this by matching.

So we had 77 patients from MD Anderson where MD Anderson did the staining, we knew they had NTRK fusions and we followed them.

And some of these patients were on clinical trials.

So we thought, okay, real-world evidence, right?

We could match these patients to TCGA data.

And we could use TCGA data kind of as a control.

And we could compare them.

We can kind of get a sense of what the expectation was based on TCGA data.

TCGA doesn’t have NTRK fusion as one of the mutations.

But they have these indications that were enrolled.

So among these 77 patients,

we have like 14 different tumor types.

So we did this study,

we went to TCGA,
here are the different tumor types that we had in this trial. So we’re talking about breast cancer adenocarcinoma, cholangiocarcinoma, GBM. What I’m showing you here is we extracted the TCGA data

We matched on stage. We matched on sort of performance status. We matched on gender or sex, I should say, we matched on all these factors that are relevant for understanding whether patient’s expectation is for survival. And look at the tumor driven heterogeneity. Thyroid cancers is way up here. The patients with thyroid cancer that are matched to these patients at MD Anderson, they’re living a really long time. Down here we have patients with GBM, Glioblastoma. This is pancreas. So even though these patients share a common biomarker profile, they have tissue types that are very different, and have very different expectations for survival. Putting this all together to try to understand whether NTRK was prognostic or not, was almost impossible to do. So, you know, conceptually, we have the idea,
1236 00:54:04.720 --> 00:54:06.140 we have The Cancer Genome Atlas,  
1237 00:54:06.140 --> 00:54:06.973 we should be using it.  
1238 00:54:06.973 --> 00:54:08.530 We can use it to do these things.  
1239 00:54:08.530 --> 00:54:10.920 But when it comes down to actually doing  
1240 00:54:10.920 --> 00:54:12.240 it’s a real challenge,  
1241 00:54:12.240 --> 00:54:14.843 and it may not provide the information that  
1242 00:54:15.870 --> 00:54:19.640 Okay, so I have two conclusions, very simple  
1243 00:54:19.640 --> 00:54:23.500 Okay, so real-world evidence in precision  
1244 00:54:23.500 --> 00:54:24.740 oncology,  
1245 00:54:24.740 --> 00:54:25.860 how do we use it?  
1246 00:54:25.860 --> 00:54:27.350 The reason we wanna use it,  
1247 00:54:27.350 --> 00:54:30.253 again is because it’s very expensive to do,  
1248 00:54:31.510 --> 00:54:33.850 to run trials in oncology.  
1249 00:54:33.850 --> 00:54:35.490 We have these biomarker profiles.  
1250 00:54:35.490 --> 00:54:37.360 Patients have to be stained repeatedly.  
1251 00:54:37.360 --> 00:54:39.350 They have to get imaging,  
1252 00:54:39.350 --> 00:54:42.520 it’s it’s burdensome for the patient, and it’s  
1253 00:54:42.520 --> 00:54:46.620 So we wanna make better decisions in early  
1254 00:54:46.620 --> 00:54:48.980 phase because we have all these failures in phase  
1255 00:54:48.980 --> 00:54:50.020 We want to do a better job  
1256 00:54:50.020 --> 00:54:52.490 of designing our phase three trials as well.  
1257 00:54:52.490 --> 00:54:53.660 So what we really wanna know  
1258 00:54:53.660 --> 00:54:56.030 is can we use real-world evidence to do a  
1259 00:54:56.030 --> 00:54:57.850 better job  
1260 00:54:57.850 --> 00:54:59.203 Where our expectation is.
In that case, in that way we can run these uncontrolled trials in early phase and, you know, save all the patients to be treated on the potentially promising therapies. Because we’ll have a better idea of what our expectation is and whether this is really promising or not. That is the promise that you hear. So it’s really about, can we define the null? We really couldn’t. We tried to. Like the case is second-line bladder cancer, the TCGA data didn’t characterize NTRK, but NTRK is so rare that didn’t bother us. Those patients are a mixture. It’s just that, that profile was not steady before. So we couldn’t do it there. When we went to the NTRK studies, the TCGA data didn’t characterize NTRK, but NTRK is so rare that didn’t bother us. So those patients are a mixture of different other mutations. We matched them based on the clinical prognostic characteristics, but the tumors are so different.
It’s really hard to understand it from the TCGA data.

In fact, this draws in the question, can you really say, if a patient has GBM and they also have thyroid cancer, but they share a mutation, can we really say something that mutation is the target?

Can you really treat those cancers as one cancer type?

Which is what the tissue-agnostic model says you can.

Right, so the biology is that important.

In some cases it has been, like immunotherapy, the immune phenotypes really seem to transcend these cancer tissues,

but for other genetic markers,

it doesn’t seem to be the case.

So I guess my conclusion is retrospectively, we really can’t.

We can’t use it right now.

It doesn’t seem like we can

because we are now in the precision oncology setting.

And yes, of course,

if you’re in rare disease setting

or you’re in a non something else, that’s unique.

You may have to, and do the best you can.

But for the trials I showed you here,

I don’t see a solution here
based on retrospective real-world evidence. I think you could do it prospectively. But I think if you’re gonna do it prospectively, there has to be a commitment that right when you start the phase one trial, you need to start staining patients and following them for survival. We don’t have a real-world tumor response right now, we can’t use that. But you need to have a sort of prospective cohort study that enrolls patients from the community. You need to pay for them to get their assays. You need to understand that the assays may change. And then I think later on when you’re coming to a decision about phase three or phase two, you go back to that prospective cohort. And you look for patterns based on the relationships between the biomarker and the standard of care. So I think you can do it prospectively. That’s not what people wanna do though, they want to use these retrospective databases. So yeah, I guess that’s the end of my talk. I didn’t leave very much time for questions, but I’m happy to take a few if there’s any.
1340 00:58:24.079 --> 00:58:26.496 (indistinct)
1341 00:59:26.340 --> 00:59:28.760 So it’s somebody asking a question?
1342 00:59:28.760 --> 00:59:30.760 I can’t really hear.
1343 00:59:30.760 --> 00:59:33.423 I’m sorry, I can’t hear it at all, actually.
1344 00:59:35.900 --> 00:59:38.900 (people chattering)
1345 00:59:45.810 --> 00:59:48.140 <v Man>Professor Hobbes, can you hear
us?</v>
1346 00:59:48.140 --> 00:59:49.780 <v Brian Hobbes>Yeah, I can kind of hear
you,</v>
1347 00:59:49.780 --> 00:59:51.543 but there’s a lot of noise.
1348 00:59:52.520 --> 00:59:54.826 <v Man>We also have people trying to get
in the room</v>
1349 00:59:54.826 --> 00:59:56.646 so (indistinct)
1350 00:59:56.646 --> 00:59:59.813 (students chattering)
1351 01:00:07.500 --> 01:00:09.502 Thank you so much, Professor Hobbes.
1352 01:00:09.502 --> 01:00:10.827 (students clapping)
1353 01:00:10.827 --> 01:00:12.533 <v ->All right, thank you very much.</v>
1354 01:00:19.030 --> 01:00:20.780 <v Man>Have a great time, thank you.</v>
1355 01:00:20.780 --> 01:00:23.780 (people chattering)