[SQUEAKING] [RUSTLING] [CLICKING]

SCOTT KEMP: All right, why don't we get started? So today, we're going to talk about the linear no-threshold model. And some of you, probably most of you, have heard about it.

In the last class on Wednesday, we went through this, the very basics of radiation and dose. And we came up to the point where we talked about these units of sieverts or REM, which are so-called equivalent dose. They have units of energy per unit mass, but they don't really mean that because they've been multiplied by these unitless adjustment factors.

But the problem is that we don't really know what to do with this unit. It doesn't tell us anything about how much cancer that gives us. It just tells us how much energy is deposited, but the amount of energy that's actually deposited by radiation is trivially small compared to, say, taking a bath in warm water, which deposits a lot more energy into your body.

And it has to do with, as we mentioned, how that interaction happens, where the energy is going. So we need some way of converting this otherwise useless unit into something that tells us about health effects. And so that is the so-called dose-response model.

Now there are lots of different kinds of health effects that we could build dose-response models for. So we typically break these down into two different categories-- what are called the deterministic diseases and the stochastic diseases. So deterministic diseases have the property that the intensity of the illness is directly proportional to the amount of radiation that you get. And sometimes there's a threshold effect that might exist in those diseases.

So the classic examples here are like skin burns, cataracts, hematological disorders of various kinds. And it's just basically, if you put enough energy onto the surface of your skin from radiation, more than capillary dilation is able to take away. So what happens is when you put energy on your skin, your blood vessels instantaneously expand to carry the extra heat away. If you can't do that, then you get a burn. And that's going to be worse if the energy goes up in value-- very straightforward.

Stochastic diseases are different. The intensity of the disease is not proportional to the amount of radiation you get. The probability of getting the disease is what is proportional to the amount of radiation you get. And so there are also lots of different things in the stochastic category. But the big one that we always talk about is cancer, genomic mutations.

And some genomic mutations do things other than causing cancer. So for example, you can have mutations that cause all kinds of classical disease by either leading to maybe a single nucleotide variant in some protein that then is miscoded and then doesn't operate correctly in your body, and you get some kind of functional disorder.

But the one that kills us if we are otherwise healthy and then kills us in mid-life is cancer. And so that's the one that we tend to focus on. So we're going to just talk about cancer only. But just to remind you that there are lots of things that can happen. All right.

So how do we construct this dose-response model? So there's some amount of dose, and then there's this black box, which is a blue box, which has some mysterious biological process. And out pops some cancer rate. And we need to understand what's in that blue box.

So one approach is that we say, you know what? We don't bother with what's inside.

We'll just build a statistical model that seems to describe what we observe happening on the outside. You put a certain amount of radiation in, and maybe we get some cancer rate out, some estimate of the cancer rate. So we'll talk a lot about this statistical model. The prevailing statistical model is called the linear no-threshold model. And we'll get into that a little bit more.

The problem with statistical models is a good model requires good statistics. And it's actually rather hard to get good statistics. And I'll show you why in class a little bit later today.

The other way we could try to build a model would be mechanistically. We say, OK, well, there's some Compton scattering, and there's photoelectric effect. And then this is going to do this to the genes, and then this, and et cetera, et cetera. But you can see what the problem is.

We don't understand how the body works. We have some understanding. When you get down to the biochemistry of it, there's a lot we don't really understand. And so this model is not a model that we can actually build.

So we are stuck with the statistical model. But we still have insights from the mechanistic world that we can use to bound how we build these statistical models. And that's what we want to do.

So my goal today is to walk you through how we got to the statistical model we have and go from there. All right. So if we want to build such a model, we need first to decide what our metrics are.

What is our input? What is our output? Is the input going to be dose? Is it going to be the effective dose? What is the output? Is it going to be the cancer rate, et cetera? So here's what we have chosen.

We have these definitions. So the first is what is risk? It's the baseline probability that a person in a defined population has the disease. So you have a baseline risk of cancer, which turns out to be about 20%.

It's about one in five of you will probably get cancer. You have a relative risk, which is to say the exposed, in this case, to radiation, the exposed population divided by the risk of the unexposed, or control group. This is a function of dose.

So here is our first mechanistic choice. Just in writing this definition, it's hidden. But there is a choice about how cancer works in writing relative risk. What this means is if a person is a smoker and they have a higher baseline risk then for a given unit of radiation, their relative risk is the same, but their absolute risk is now higher.

A unit of radiation for a person who has a higher baseline risk of cancer causes a higher probability of getting cancer. Does that make sense? Because it's relative to your baseline risk. And if your baseline risk is higher, then your absolute probability is higher.

And so that is a choice that we are making when we build this model. And it's not one that we actually talk about a lot.

Now strictly speaking, we could have various controls to interact with different kinds of baseline risks and try to control for it. We should ask the question, what is the truth of the matter? Has someone looked into this? And the answer is yes. People look into this. And as best as they can tell, this is about correct.

So for example, a famous one is radon exposure, which is an inhaled form of radiation, and smoking, another inhaled kind of carcinogen. Does of relationship hold?

The answer is for light smokers, it turns out to be submultiplicative. Sorry, light smokers, it's super multiplicative. And for heavy smokers, it's submultiplicative. So it's about multiplicative.

Another classic one is arsenic exposure and radon in miners. And that one is also found to be basically multiplicative. So this definition is basically supported by the evidence, but there are places where the interactions are not strictly multiplicative. This is just a simplification we have to make in order to figure out how to build the model.

Then we have this definition, which we will tend to use when presenting. We typically take the relative risk, and we subtract 1. Or we take one and subtract, subtract it to get what we call the excess relative risk. So if you have a 20% chance of getting cancer, and then a 21% chance of getting cancer, if you also are exposed to radiation, then the excess relative risk is just that extra 1%. So let's look at some dose responses that are in the literature.

So here's the excess relative risk of a solid cancer. A solid cancer means not leukemia. By the dose to colon-- the colon is, as it turns out, a standard organ. Every organ has a slightly different response, and people tend to use the colon for whatever reasons as the standard. And so what they typically do is they adjust all the exposures to colon-weighted exposure.

And this is for the Japanese nuclear bomb survivors in Hiroshima. It's called the life span survey. And you see the data here in blue. And they've binned it into different neutron-rated doses. And there's a model there.

So this model is a linear no-threshold model in red. The black are the error bars of the model. And let's just look at what it means.

Linear means, well, it's linear. No threshold means that the intercept is at 0, 0. There is no threshold at which radiation produces zero effect. That's what that simply means.

And what do we see? It's pretty clear it doesn't fit the data. The data rolls over at high doses. So there's no actual controversy here because no, if you are, in fact, looking at a population of people who are getting doses of 3 and 4 gray, which is equivalent to basically three or four sieverts, something like that, typically.

Well, this is neutron-weighted, so that wouldn't be exactly equivalent. But anyway, if you're getting people at these high doses, you don't use the linear model. You use one of these nonlinear models, which we do have in the arsenal. Where we use the linear models, only really down here below 1.

And the reason why it's controversial is because when you look at nuclear accident events, this is in fact where most people lined up. Most people get almost no dose from an accident. And they're right down here in the linear regime.

And the question is, what exactly is the shape down here? And it looks like it kind of is linear. But we really would like to know. So that's the story. So let's look at some more data.

This is DNA mutations in thyroid tumors of Chernobyl survivors. The maximum recorded dose here is 1000 milligray, which is one gray. So we are completely in the low-dose regime. And at first, let's just look at what we have here.

The first thing to notice is there's actually considerable background population at 0 dose. This is the distribution of observations. And you see that on average is something like 25 small deletions in this genome, even at no dose.

We can draw a straight line, and we can call this the baseline risk of a mutation. This is like the baseline risk of cancer, if you will. It's just a different measure. And then we can look at what the effect is after that.

The other thing you can observe in this data is there seems to be a lot of people clustered here at 1,000. So clearly, something happened in how they were doing the dose reconstruction. They're basically saying oh, anyone over here, we're just going to put them into this bin, which now we have to be careful of that bin and how that affects the measurements.

So there are some systematic errors, which are a problem. And indeed, if you look carefully, you'll see that there appears to be a discontinuity between the zeroth bin and the first bin, at which there's actually dose in terms of the distribution. So there's something else going on there as well between the two control groups. So lots of issues.

Another issue is that you might just say, oh, I'm going to throw this into a regression to fit that line. Seems reasonable. I would just throw this into Excel, and it would perform a least squares fit. Would that be a good choice here? Why?

AUDIENCE: The data is really wide.

SCOTT KEMP: That in and of itself is not a problem doing least squares. A least squares fit is a maximum likelihood estimate only if the underlying data have a Gaussian distribution. Gaussian distribution has tails that go off to infinity in both directions, but you can't have fewer than zero deletions.

So this is not Gaussian-distributed. And if you now fit something using least squares, you're not going to get the correct fit. You're not going to get the maximum likelihood estimation of the data. So this, turns out, is more of a binomial distribution. And so you would need to use an appropriate regression model.

So all this is to say that this is not a tidy problem. And there's lots of places where people can go wrong in these types of analyzing these types of things. So should we take this fit, which is the one provided to us by the authors, which is a linear model? Well, it looks all right. It does look like there's some low region here and maybe some higher region over here.

Maybe we should take this. Maybe we should do put some lowness, maybe make that green. Maybe that green one is more convincing to us, looks like actually fits the data better. Or maybe we should do that. maybe that actually fits the data better. I don't know.

I mean, the data do definitely seem to go up, and then they definitely seem to go down. And then there's this big outlier point over here. And what do we do? How do we choose these models?

So here's where we're going to have to make a choice. And this is the practice of model selection. And I don't think people really-- people know about overfitting data as a problem. But I don't think people are really given the tools to talk about model selection.

So as a general rule, we want to choose simple models. But there's a good reason behind why we want to choose simple models.

We don't need to choose the simplest model. We still want to describe the data well. But there's a preference for simple models. And that comes from Occam's razor, which is to say, basically, simple things make the most sense.

Let's say I have a picture. And there's a picture of the NW14 building. And remember, there's a road here in front of it, and then there's a sidewalk. And then there are these trees planted on the sidewalk along the front of NW14.

There's a little door here. And then there are some windows in the NW14 building. We're behind the trees? Maybe I should make the windows. I don't really have any different colors.

There we go. All right. How many windows are in this picture? Four. All right. But how do you know there's not like a little tiny window right here? OK, maybe, but it's strange credulity. Maybe there's actually a secret little window right here.

OK, It gets obviously the more complex your model space where you're having to consider all kinds of possibilities, the less likely it is. So that's the kind of intuitive explanation. But we can actually show this mathematically.

So we can go back and derive Bayes' rule. So let's say we have the probability of A and B-- A and B. Well, what is that? Well, it's the probability of A times the probability of B if A happens given A. This is straightforward.

And then we can just flip the variables and write the reverse. Everyone is happy with that? And does everyone agree that the probability of A and B is the same thing as the probability of B and A?

Yes, so then we can just create these. Probability of A times the probability of B given A is equal to the probability of B times the probability of A given B. And then I can just divide both sides by this. And that's Bayes' law. So that's a very useful way to derive Bayes' equation if you don't really remember.

This actually embodies Occam's razor. And let me just show you why. So let's make a substitution.

Let's call B here a hypothesis, and let's call A the data. If I write over here, people are going to have to turn their heads too much? No? OK. All right. So we can write something like this.

The probability of hypothesis one, given some data, is equal to the probability of the data given hypothesis one. Or let's actually call this model one. It's a better model one. Model one, given the data, times the probability of the model divided by the probability of the data.

All right. So this is the probability that the model is true given the observed data, which is what we would like to know. And so we could also have a second model. We say, OK, well, we can have the probability of model 2 given the observed data, which are unchanging.

All right. And we can just divide these two equations. So we get the probability of this minus this. These terms cancel.

And so we have this times this. Let me just rewrite this here, make it a little bit cleaner. This is equal to the probability of the data given model 1 times the probability of model 1 divided by the probability of the data model 2 times probability of model 2. All right.

So maybe it will make more sense if I just do that. So what is this term here? So this is your prior. This is the probability that you believe the model without the data.

So is my prior belief in the probability of model 1 versus my prior belief in model 2. And if I'm agnostic between models, then this is basically unity. This goes to 1.

All right. So now we simply have-- this looks a bit sloppy. Let me just rewrite over here. Probability of model 1 given the data probability of model 2.

So now we have this, which is what is the ratio of that we believe in view of the data, model 1 over model 2. And it's just the ratio of the likelihoods. This is what's called the likelihood-- the probability that I see the data given the model.

So here's the insight. If model one is the simple model, it has a small number of possible states. If model 2 is a complex model, it has many more possible states. So therefore, the probability of seeing the given data for a complex model is always going to be smaller.

So for model 2 complex relative to 1, this ratio here is going to be greater than 1 because this will be less than this. Does that make sense? And so that means we will prefer by some amount model 1 over model 2.

I can just take model 2. I can multiply it by both sides, put it over here, and it has 1 times model 2, which means something greater than 1 times model 2 is the probability of model 1. So that just tells us that we have a preference for models which have more simple descriptions, which is essentially this, just mathematically.

So it takes a lot of data or some kind of prior reason to support choosing a different model. And so let's look at how we decide.

So one of the things that comes up often in radiation is should there be a threshold. And so one question might be, is this model actually really more complicated than this model? And the answer is yes.

So in the no-threshold model, our model is something like y equals alpha x. And in the threshold model, our model is piecewise, and it looks like y equals 0 if x less than or equal to some threshold. Otherwise, x greater than threshold is beta x minus some intercept.

So we have now extra parameters in the threshold model. And the threshold model thus requires more data to defend it than the no-threshold model. All right.

So how is model selection done? So the problem is that when we look at this going back to the Bayes rule here, when we look at the probability of the data which. Well I deleted. And this, I was able to delete it here and not worry about it.

But when you want to calculate it directly, you need to do this bit of math here. And the problem is that this bit of math is often impossible to do or very difficult to do. The way you traditionally do it numerically is through a Markov chain Monte Carlo. But analytically, it's a really difficult thing to do. So instead, what we tend to do is come up with some kind of approximation method.

So the best way to do this if you have a large data set is called cross-validation. So what you do is you have a lot of samples in your data. Here's a whole bunch of people in my data set. And I train my model on this population. And then I use these people who didn't go into the trained model to test how well the model works to see, OK, what is the prediction error for each of these observations?

And then I do that again, but I choose a different subset. And I go through all the possible subsets. And I basically measure how badly the model predicts the outliers. That only works if you have a lot of data. And you can actually train the model on a small subset.

If you don't have a lot of data, then you're kind of in trouble. The other way to do this is that you can use a scoring function. And there are two good scoring functions. One is called the Bayesian information Criterion, and the other one is the Akaike Information Criterion.

If you have reason to believe that the true model that actually contains the real physics is somehow in the set of models that you're evaluating, which is almost never the case, then BIC can be shown to be the correct way to approach it. Otherwise, the Akaike Information Criterion is usually a better way to do it. Here it is-- the number of samples times the log of the residual sums of squares, plus the number of free parameters.

Basically, if you run if you run any data through NumPy or something, it has a way of just calling for this value. And you can get out the value. And basically, what you do is you get this number out for every model, and you just choose the model that has the lowest AIC. That is the model that contains the most possible information.

This is a very simple result which comes from a long information-theoretic derivation, and you can go look it up if you're interested. But it's basically simply a way to penalize overfitting-- that is, at least has a foundation in information theory.

So this is how you would do model selection. And you should know this because you're going to have to do model selection if you're a researcher, especially if you're a grad student. And people frequently say, oh, here's some experimental data, and it looks like this. And you fit something, and you maybe don't give it a second thought. And really, you should be trying multiple models that possibly make sense and scoring them either using k-fold cross-validation or using AIC before just choosing one for your thesis.

Don't go on R-squared. R-squared only tells you how well it matches the data points. It doesn't tell you how much information is in the model. This is elementary, but if I have these data points like this and I have a model that does this, it fits the fits the data perfectly, but probably has almost no predictive value.

OK. So this is where linear no threshold comes from. If you actually go through this process rigorously and you start testing a bunch of models, as I have been doing lately, you will find that the LNT model, this is a k-fold cross-validation result for a 20-fold cross-validation.

This is for the lifespan survey data. These are the people in Hiroshima who were exposed to radiation from the atom bomb, which is the largest single data set for low-dose exposure. I think I have some information about how many people are in a subsequent slide. Only below 1 sievert-- so we are in the potentially linear regime.

And these are different models-- linear model, linear-quadratic model, a threshold model where we allow the threshold to exist at the optimal possible location that give the best possible fit to the data. Quadratic, this is a Gaussian. These are non-parametric models, Gaussian process, Linear Gaussian process other.

And you see that basically the linear model just constantly wins. And this is frustrating because I would like to-- I would like a model that explains why this happens. But that shape, you'll see it over and over. But statistically, the only thing we can really defend is the linear model. All right. Now I have to go through all this again.

So there are some other pitfalls that you should be aware of when looking at radiation dose-response models that sometimes people fall into. And here's an example of one of them. So this is the excess relative risk for cancer for exposures per sievert of radiation to different organs. And so remember we talked about colon-weighted dose. You can see here that it's not uniform depending on the tissue.

All right. So let's see. So 0 excess risk means that that's your baseline risk for no radiation. And here, we see that for uterus exposure, the probability of cancer is actually lower. So should we go around irradiating women's uteruses?

AUDIENCE: Just for fun?

SCOTT KEMP: No, for their benefit. Because we will reduce the risk of uterine cancer. Should we go and do that? This data is correct. I mean, the real data.

AUDIENCE: How much data is it?

SCOTT KEMP: Well, that's a good question. There's the error bar.

AUDIENCE: The gist is it doesn't cause cancer, but then maybe there's other effects that we're not accounting for.

SCOTT KEMP: There's other effects. So you're being precautionary. But on the basis of cancer, you don't have an argument against it.

AUDIENCE: Cancer for the mother?

SCOTT KEMP: For the uterus.

AUDIENCE: Yeah.

AUDIENCE: You just have to look at how cancer affects the quality of life, not just the incidence of some.

SCOTT KEMP: Well, if you don't get it, that's better than nothing. This is called the look-again effect. Have you heard of the look-again effect. All right. So if these error bars are 95% confidence intervals, which means 95% of the observations are estimated to fall within the error bar, which means one out of every 20 is going to fall out of the error. The estimation will be wrong by the amount of the error bar.

So there's about 20 things up here. And one of them is wrong. And that's what it is.

But if you look at all of these things and then just picked this result out and wrote a paper, then you would have a problem. So in fact, this is easily shown if we also plot the mortality from radiation risk. And you see, it was just a statistical fluke.

The mortality shows that, in fact, radiation exposure does increase your mortality of uterine cancer, even though it appears to reduce the probability of uterine cancer. And here you see here the effect has shown up here for oral and pancreas in the opposite direction. It's just this effect.

So there was a famous XKCD cartoon about this. Jelly beans cause acne. Scientists investigate. And there's this claim that they found this link. And then here is this thing. It appears to be a certain color.

And so they go through all the colors. Found no link between purple and brown and pink and blue and teal and salmon color, red color, turquoise, magenta, yellow. The p-value is always too large to defend any link. And then, lo and behold, they get to green.

And the p-value appears to justify that there is acne caused by jelly beans. And then you have this publication.

If you look at the data and you start ignoring things, you actually have to adjust how you calculate the p-value, because every time you pull a sample, every time you run a test, you are actually adjusting. You are making choices and filtering the data, and therefore you have to actually change how you do that calculation. This is frequently forgotten about.

And people do this, and they run their science experiment on their instrument until they get the result they want. Now finally, the instrument worked right. And then they publish. It's like, well, actually, you have to take into account all of those unsuccessful runs, and maybe it's just noise.

So this happens all the time. And one of the ways this manifests is that people get these results out of small little laboratory experiments about radiation, and they show that some radiation is good for you. And then they say, Lo, and we have discovered hormesis, this idea that a little bit of radiation is good for you, and it's not real. So you've been warned.

All right. This is kind of like the grand plot from the lifespan survey. This is from *ICRP Publication 99,* which is actually assigned reading. I hope you looked at it.

And here is up to two sieverts of dose. And you can see the linear regime. And then you see in the low-dose regime, there does appear to be this like super linear region.

You could say, well, maybe that's just it does appear consistently between data points. So that could be. It looks like it's not just noise.

But I have tried as I could to find a non-linear model to describe this. You can't do it. It does not. Nothing is statistically justifiable.

So I mentioned this data set earlier. It's the single largest data set. It tracked 80,000 individuals who were exposed to radiation because of the atom bomb. A lot of people think, oh, well, that's a lot of radiation.

Well, it turns out that 64,000 of those 80,000 were exposed to less than 100 millisieverts. So most of these people are really in the low-dose regime. To put that into perspective, 100 million millisieverts is about the same amount of dose a radiation worker in the United States would be allowed to receive over a 20-year career. So it's not that much radiation.

So we do have 64,000 observations in here. And it still doesn't support anything other than the linear no threshold. One of the things I want to talk about, though, is just to look at the slopes here.

Right now, the ICRP, the International Committee on Radiation Protection, which is one of several bodies that recommend safety standards for radiation, they need to publish a number for the slope of what this dose response is-- how much cancer you're going to get for a certain dose.

So the ICRP recommends 0.53. But if you read the slope of that chart, it's actually 0.64. And if you really limit yourself to the low-dose regime, you would actually read one, about one per sievert, excess relative risk of one per sievert.

So should we conclude that the ICRP is being conservative, and that the linear no-threshold model is conservative, as a lot of people say? No, in fact, the opposite-- the data show that it is not conservative. They are being generous in allowing more radiation out there than is defendable and as opposed to the idea that LNT is somehow a conservative model. It is also not supported because of the statistics. The threshold, the threshold model is not really supported.

So let's say, we really liked the idea of a threshold because most people in the nuclear community would love that to be true. Someone here was doing a thesis on this. And so we could ask, OK, what would it take to actually establish that a threshold does exist at some point? And so we can do that calculation. So let's just do it really quickly here.

We set up a little thought experiment. And this is by the way, borrowed from *ICRP Publication 99*-- section 2.42 if you want to go read their version of it. And what we're going to do is ask how many observations or basically, what does it take to observe a threshold at one millisievert?

Now, one millisievert is not all that low, but I think it's like just 50% of natural background. So that's not really super low dose. People who talk about threshold models often talk about even lower thresholds.

But it seems like a reasonable place to put your test. If you put it really, really low, it gets really, really hard to see because you have very, very small effect. So we're going to be a little generous, and we'll put it we'll put it at one millisievert-- half of background.

All right. Now we need to make some assumptions about how the world works in order to run this model. That first assumption is we're going to assume that the baseline lifetime lethal cancer risk is 10%. Now, earlier, I said it was 20%, which is true. But what we're going to do is assume it's 10% in this little thought experiment. Why?

Well, if the background rate is lower, it's easier to see the effect of the radiation. So it's going to bias the result of our model to make it appear to be easy to see the threshold, easier than reality. Does that make sense? All right.

We need another assumption. We need an assumption about what of excess risk of cancer is per unit radiation exposure. So let's adopt the excess relative risk of one per sievert.

Now remember that the ICRP recommended 0.53. But we are going to take one, which is to assume that the effective radiation is stronger than the consensus view, which makes it easier to see the effects of a little bit of dose of radiation and makes it artificially easy to estimate the existence of a threshold. So both of these are biased in the same direction. It's making the experiment easier than it really is.

And three, we're going to assume perfect experimental conditions. You say we can have any population. We'll know the dose that everyone gets perfectly. There's not going to be any uncertainties there. And we'll know all the control parameters exactly. We don't need a fancy control group.

We don't have confounders from people smoking and other stuff noising up the data. We have exactly the right number of identical individuals, and we're going to all expose them to the exact unambiguous amount of dose, which, again, is making the experiment artificially easy. So we have made all of these generous assumptions that will cause us to underestimate the difficulty of actually looking for a threshold. All right.

So now we have some hypotheses. And we're going to set the null hypothesis to the threshold. We'll say the null hypothesis is that at one millisievert, the dose is zero. The effect is zero. At 1 millisievert dose, the effect is zero. That's our null. That's the threshold model.

And then we'll have this alternative hypothesis that says that the relative risk at one millisievert will be as predicted by the linear model, one per sievert. So those are our two hypotheses.

OK. So let's see what we would expect will assume to make the math easy incorrectly, that these are Gaussians which enlarge n is a reasonable approximation. And so here we have the null hypothesis, which is the threshold model. We expect that the population sample will have a mean at the baseline cancer risk of 10%.

Because at one millisievert, there's no effect from radiation. So you just get the baseline risk. And then there will be some statistical variation around that expected mean.

If hypothesis one is correct, then we expect the background rate plus the excess risk, which is the background rate-- remember because it's relative-- times 1 per sievert times 1 millisievert. So it's going to be 0.1001. So that's where we expect the mean of the distribution to be. And again, we'll have some variance around that.

So the question is, how big of a sample do we need so that the uncertainty is squished sufficiently small that we can tell the difference between these two distributions. That's the test of what we need.

So we need to choose two things. We need to choose this critical value, which will be our cutoff. And we need to choose the sample size, which will determine the width of this distribution. So the cutoff is basically where we decide to accept or reject the null hypothesis. If we see a cancer rate above this level, we'll reject the null.

Otherwise, we're going to assume that the null hypothesis, which is in this case is the threshold model is correct. So we are being very generous because what we've done is we've used standard 95% confidence here and 80% power, which is to say that we will accidentally reject the null only 5% of the time. And by setting 80% power for the other test, it's OK if we reject the LNT theory at least 20% of the time that it was correct.

We don't really care about getting the LNT part wrong, but we want to be very careful not to accidentally reject the threshold model as being wrong. So we're being very generous to the threshold model.

Normally, you would do this in the reverse. You would have a higher standard for the threshold model because it's a more complicated model. But here we're going to be generous and assume the threshold model is the notional model. So now we need to basically write what I've drawn here in some equations with these type I and type II error. And so this is just the z-test, as you recall.

Here's the definition. Z is the number of standard deviations away from the mean. X-bar is your sample mean. Mu is what you expect. Sigma over root n is your standard deviation of your sample.

Just some criteria-- this technically requires a normal distribution. We're using the central limit theory to basically say that as long as the sample size is large, the number of people that we're going to test is large. Over 30-ish, this will hold. We'll go back and check this. When we do the math, if we get a number less than 30 individuals, then we'll say this was a bad choice. And we'll see what it is.

So let's just put in the math. For type I error, you can look this up. If you want a 95% confidence of accepting the model, or only a 5% chance of wrongly rejecting it is 1.5 sigma out.

So there's your equation. There's the equation for type II error on these other distribution -0.84. And we have two equations and two unknowns. And so we do the math, and we solve it. And we see the threshold is 1.0007. Anything above this, we'll reject the linear model at 56 million people.

And actually, to be more precise, 56 million people is the minimum size. If your observed incidence of cancer rate deviates from this, you will need actually potentially a larger population.

So here's the issue. Everything we've done favors, underestimates the difficulty of seeing the threshold. And the question is, can you do an experiment with 56 identically humans and give them all one millisievert of radiation in order to see the threshold? That is an experiment you probably cannot practically do. That's the point.

We'll never have this experiment. And this is really optimistic. If the threshold does exist, it is extremely difficult to show that it exists. So since we can't prove or disprove the existence of a threshold with any reasonable experiment, what happens?

People just choose to believe what they want to believe, but they're not supposed to. They're supposed to follow Occam's razor and choose the model that has the best predictive power given what we know. So that's the story between threshold models and no-threshold models, is that the no-threshold is not defensible because it is a higher complexity model, but because you can't prove the threshold people wrong, there will be people who will insist on using threshold no matter what.

All right. Now, could we actually do better than this? And the answer is yes. If we take a mechanistic perspective, we can learn something about how radiation biology works and inform us as to whether we think there might be a threshold or not.

So this is basically saying, now we're going to allow this term here to be something other than unity. We're going to inject a prior into the model based on some additional information. So ready to learn how cancer happens? All right.

So you have some amount of radiation coming in. And one of the things it does is it hits water because you are mostly water, as usually what happens. And so what does it do? It ionizes the water, kicks off an electron, and it produces hydronium. And eventually, you get OH hydronium and OH dot.

And this is a free radical. And it goes and attacks one of the sugars on your DNA and causes a problem. So this is usually what happens. And this is happening all the time in your body, not just from radiation. Also, chemicals cause free radicals and do this, and your body has to deal with it.

Another thing that you can have is that you could have the radiation, either the photon or an electron, directly interact with an atom in the chain in your DNA, the non-water atom, and cause that molecule to break apart. And this is where things get interesting.

So recall that it's not really like the total energy deposit. It's how and where it's deposited. So I showed you this on Wednesday.

And what I've done now is I've superimposed in approximately to scale the size of a cell. And you can see here that for low LET radiation, you can have lots of interactions on the scale of the nucleus. And so it is possible to have radiation.

A single radioactive particle, a single electron or single photon enter the body and undergo photoelectric effect and Compton scattering and cause all of these subsequent changes multiple times within a cell. And this may generate multiple free radicals from water, these OH dot ions, or it may directly attack the DNA. And so the probability of getting multiple breaks to the genome goes up because of the nature of radiation, because of this low LET radiation.

So this is looking at different kinds of breaks that might happen. The top one is what we get when we have that OH dot showing up just at some random place, and it breaks one side of the DNA. And then we have this process called non-homologous end joining. And it's like a little thing that comes along and detects the break and glues it back together.

And it turns out we have one of these happen every half-second per cell in our body. It's happening all the time. All the cells breaking, breaking, fixing, breaking, fixing, breaking, fixing all the time. It's amazing.

But this kind of break where we have a double-strand break is pretty rare. That just two happen to happen in close proximity to each other. It's something like once every-- turns out this is like every two hours. So it still happens. And you still have non-homologous end joining coming and gluing this together.

But sometimes you get something like this. And what is the end? This thing floats away, and the enzyme comes along and goes, uhh. and then it cuts these things off and it glues it together and just takes out a few base pairs. And this happens all the time in your genome. And sometimes it's benign, and sometimes it's not. And when it's not, you get cancer.

So here's the key takeaway. The repair process is what's giving you the cancer. If you simply break the genome and you don't repair it, then generally what happens is the cell becomes senescent and doesn't continue to operate and just dies generally. But if you go and you introduce these genomic errors from this repair process. That's when you get cancer.

So sometimes you hear people who have seen the idea of hormesis, that a little bit of radiation is good for you, and they are looking for ways to justify that. Remember, most likely that is actually the look again effect that they've picked up on initially. And they say, well what would describe this? And they'll say, oh, it turns out that if you give a cell radiation, we can observe that the repair process is upregulated.

Of course, it is. You've done damage to the genome. It's going to have more repair to do. But then they think oh, but maybe that repair process is preventing background cancers. It's like no. That repair process is causing the cancers. And so this is a key insight here.

The other insight I want you to take away this is possible from a single photon-- comes in, it comes through. You get Compton scatters. You get an electron. You get another photon, another Compton. Two electrons. Boom, boom, boom.

This is technically possible. Therefore, with some very, very, very small probability, a single photon could give you cancer-- very very, very small, not normal. Normally, it requires multiple hits to the genome and multiple places, et cetera. But if there's a very small non-zero probability, a threshold cannot exist. So this is the mechanistic argument for why we don't have a threshold model. Yes.

AUDIENCE: What energy photons are required to do a damage like that? Is there a threshold?

SCOTT KEMP: You need to deposit at least basically an EV in the interaction to ionize something. And typically, these photons have hundreds of hundreds of keV of energy-- so plenty of energy. Yeah.

OK. So we do live in a radioactive environment. Our bodies do have these repair mechanisms. There are error-proof repair mechanisms.

There's something called, sorry, homologous recombination, which is a way of error-checking the genome. But in human cells, it only exists during cell division. In the normal growth phase of the cell, we only have L1 copy of the genome, so we have nothing to fact-check it against.

Certain bacteria, like Deinococcus radiodurans, has multiple copies of its genome. And every time it fixes it, it checks repair against the other copies. And so it makes it extremely robust against large doses of radiation. But human cells, only once-- only copy, except during cell division. And so we have no way of enduring those radioactive assaults.

So yes, we are evolved to live in the radioactive environment. We have repair processes that keep the cells going, keep them alive. But it doesn't guarantee us that the cells don't eventually get cancer. It only guarantees that on average, most people will live long enough to procreate. And that's all evolution has given us.

So little bits of radiation do matter very likely. And we shouldn't be seduced by any of the ideas that radioactive environment, and we're accustomed to this, cetera, cetera. That's it. That's why LNT is right, or very likely. Yep.

AUDIENCE: Here's a question. So if the cluster double-strand breaks, if the repair mechanism happens to repair in a way that's malignant, is there a threshold for a certain amount of those dangerous repairs have to happen? Because I'm assuming if probabilistically, there's been cancerous repairs.

SCOTT KEMP: Yep. Yeah. So they call them promoters or potentiators. So I believe-- and this is going from memory, so you should double-check this-- but I believe the general consensus is that you need on average about six or so of these things to get the cell really going into the cancerous mode.

AUDIENCE: In a specific cell?

SCOTT KEMP: In a specific genome. Yeah. That doesn't mean you actually need six. If the damage happens in the right region of the genome, you in fact can cause problems right away. You can also, if the damage happens in the right region of the genome, you can also just instantly kill the cell, which is not a problem.

Cell die is fine. But on average, it does need to happen more than once. But remember, it is happening all the time in your body. So you are accumulating these potentiated cells over your lifetime, which is why you typically get cancer when you're old.

And this is also why the relative risk model makes sense. The idea that a smoker has a higher probability of getting cancer from radiation makes sense, because they've already trashed their genome from cigarette smoking. So they've accumulated these potentiated genomes. So an additional damage from radiation is more likely to actually cause the cancer.

So yeah, that is what's likely going on in most cells. Yeah.

AUDIENCE: That's a good point. It's like you have these--

SCOTT KEMP: Just depends on where it is. It's a stochastic process. It's not a threshold phenomenon. It's stochastic. Yeah.

AUDIENCE: What is the effect of how long you get the radiation dose over [INAUDIBLE]?

SCOTT KEMP: The dose rate. Yeah, so there is evidence of dose-rate effects. If you want to read about that, you can go to our ICRP, which is in the reading.

Generally, what the dose rate effects show is like if you increase the dose rate, then it tips over, and then it's to say that the additional units of dose at a high rate don't have as big of an effect. And that kind of makes sense. You have a lot of damages occurring, and they're all happening before the repair process happens kind of a thing.

So that is the extent of my understanding of it. And you can go and look more. Yeah. Yep.

AUDIENCE: Any of the nonlinear models end up kind of converging on setting those nonlinear weights to 0? I guess, do you see those models converging on a linear model still?

SCOTT KEMP: The nonlinear models are defensible once you allow the high-dose regime like this. Then they do converge for very high doses. Yeah. For what we would still permit in the lower-dose regime. I think we now have enough data on leukemia to support a nonlinear model on leukemia, but they're basically all linear in the extreme low-- and so in the limit of ultra-low doses. They all kind go like that. Yep.

AUDIENCE: [INAUDIBLE] why would it be? You said earlier that people in industry would want this to be true. Why is there such a small effect? Why would it be--

SCOTT KEMP: Oh. Well, maybe we can have an answer from the audience.

AUDIENCE: I mean, I'm not really trying to-- I haven't really looked into threshold models or not. Just what would the cost difference be? And I'm going to tell you. It's going to be small.

The purpose of the research is to say that for nuclear power, it does not matter. Just focus on why the containment is so damn expensive and all these things. That's essentially what I do. I'm not looking into why threshold. Why not threshold? Obviously, I've done my [INAUDIBLE] review. But it's [INAUDIBLE]

SCOTT KEMP: But what's motivating your study? So you believe that it's likely to be or have a small impact? Bongiorno believes likely to have a small impact on the total cost of reactors. I agree with that. But there are people out there who don't believe that, which is why you're doing the study in the first place.

AUDIENCE: Yeah, exactly.

SCOTT KEMP: Yeah, and so, just to drive the point home, in May of this year, the White House issued Executive Order 14300, which instructed the Nuclear Regulatory Commission to quote, "specifically consider adopting a radiation limit," which is, again, the politicization of our regulation with nonscientific views. There are people out there pushing hard for this because they believe, without evidence, that this is going to make it cheap to build nuclear power. So yeah.

AUDIENCE: A more general question. So when you're not familiar with the subject-- so for example, this is going back to the look-again effect. Yeah, so using the example of the jelly beans or whatever. Say I was familiar. How do you tune your radar to be aware of those types of things, besides knowing the subject and being like, that's very contrary to all the existing evidence?

SCOTT KEMP: How do you tune your radar? I'm not sure. Are you asking the question, if you had only looked at the green jelly bean scenario?

AUDIENCE: I think I mean, as a reader who looks at various articles--

SCOTT KEMP: Oh, how do they're not doing the green jelly bean scenario? Yeah. Yeah. You don't often. Yeah.

And if you go read the medical literature, you will find all kinds of articles where you're like? Why did you look at this? Why did you study this? And I think a lot of that is oh, look. And then they write a paper about it.

So that's why you, you want to avoid what I call single-study syndrome, where you just find one study that supports something. If that study has not been replicated, you shouldn't do it. And so we try to teach people not to do this.

One of the things is something called hypothesizing after the fact, which is to say, oh, I saw this thing. I wonder why it happened. Oh, maybe this is the reason. This is my hypothesis. Here's the data, publish. You're not supposed to do that.

What you're doing is you're exacerbating the look-again effect. You're supposed to have the hypothesis do the test of your hypothesis. And then if it fails, it fails. But unfortunately, this is a problem with how we promote people. It's a social problem because we say you need publications to show your impact as a scientist.

And therefore people are incentivized to have bad behavior and to publish everything they can when it ought not be published. And then journals also don't want to publish null results because they're not exciting. And so there are lots of experiments that are done that are never reported in the literature. And all kinds of people are coming up with the same hypothesis, doing the same experiment, getting the same null result, and not knowing that the repeating people's experiments because we don't publish null results. It's a real big problem. Yeah. Yeah.

AUDIENCE: Follow-up question of mine-- since an important part of the scientific process is repeatability. This is also coming from a lack of knowledge. Is the system set up so that it is hard to get funding to just prove or disprove someone else's stuff, because someone would? We want you to make breakthroughs.

SCOTT KEMP: Something novel. Yeah, sure. Absolutely, it's really hard. It would be really hard to justify a replication study unless the initial result was, for some reason, very controversial or very important. Someone say, oh, it's already been done.

AUDIENCE: So does that make it such that single studies set the tone of the body of literature, when in reality, it should be, a collection of studies?

SCOTT KEMP: Yes, they absolutely do. And I mean, in the best of all worlds, the first study changes the direction of research, but research goes and checks. It is the case that sometimes we go down the wrong path, sometimes for a very long time. So I mean, maybe the classic example, which is not really relevant to how modern research is done and published, but would be Ptolemaic epicycles to explain how the sun and the planets orbit the Earth. And it took hundreds of years before that view was refuted because of the firmness.

There are other phenomena which I can't bring to mind, which have existed for various periods of time. Does anyone remember the-- during the COVID pandemic, there was a big push for-- what is the name of the anti-malarial drug?

AUDIENCE: Hydroxychloroquine?

SCOTT KEMP: Hydroxychloroquine. Do you remember hydroxychloroquine? So that didn't just come from some wacky person. I was about to Google it, but you got it. That actually came from a scientific study published in *Nature,* top-quality journal. And what they did is they found-- people who want to go and don't want to learn about rate biology, you can go.

So what they did is they used Vero cells, which are derived from liver. And it's a standard test cell to see if different drugs would prevent COVID infection. So they put the cell inside a solution of hydroxychloroquine. They infect the supernatant with little virus of COVID. And lo and behold, it reduced the probability of infection. So everyone assumed that this would work.

The thing is that it turns out that viruses can infect cells in a couple of different ways. So one of the classic ways is the virus lands on the top of the cell, like here. And the cell detects something is wrong, and it endocytosis, creates a vesicle, brings the virus inside.

And then what hydroxychloroquine does is it acidifies. This is called an endosome. And it acidifies this endosome and destroys the virus.

And so it prevents the viral, which is being endocytosed is the term, from infecting the cell. So it does indeed prevent infection via this mechanism. But this mechanism is the dominant mechanism in liver cells, but not the dominant mechanism in your respiratory tract, which happens to express a protein here called ACE2. And most of the viruses land on ACE2 and get in this way.

And so hydroxychloroquine does nothing to prevent infection via ACE2. So it was good science, but they forgot that this cell type did not express this protein, which is expressed in your airway epithelial. And so the result was wrong.

But you will not believe how hard it was to get hydroxychloroquine for a large number of months because of this one paper, which had never been validated. And that wasn't even the look-again effect. It was just that they used the wrong cell type. So it was good science, but it wasn't complete. So yeah.

AUDIENCE: Continuing unless someone else has a question.

AUDIENCE: I was just going to say, I feel like there'd be a need for a scientific magazine called *Null, Not Void,* which is just null results.

SCOTT KEMP: I think there is a journal of null results. But I don't think most people want to bother taking the time to write up their work. So yeah. Anyway, this is a problem.

We'll talk more about this in the final class. No other questions about radiation biology? All right, good. So on Wednesday, we will calculate how many people die from nuclear reactor accidents. And that will tell us something about whether nuclear power is safe or not. And we'll use the linear no-threshold model, which I hope you will believe now that I've defended it.