Okay. It is about 2:03, so we're gonna get started. Good afternoon, everyone. My name is Erin Farley, and I'm one of JRSA's research associates. For those of you who may be less familiar with JRSA, it stands for The Justice Research and Statistics Association. We are a national non-profit organization dedicated to the use of research and analysis to inform criminal and juvenile justice decisionmaking. We are comprised of a network of researchers and practitioners, which at the core include directors and staff from the state Statistical Analysis Center.

So before I welcome our presenter, I want to let attendees know that we will have a skill building webinar in May. It is scheduled right now for May 18th. It is not posted, but it will be available for registration very soon. It will be on ArcGIS, and will be presented by Elizabeth Groff from Temple University, so we're really excited about that, and that's going to be in May.

So today, it is my pleasure to welcome you to our webinar addressing AreaUnder-the-Curve Analysis by Doug Spence. He is the director of the office of research and strategic planning, and is at the West Virginia SAC. So welcome, Doug. Before we go further, I want to thank our partners at the Bureau of Justice Statistics for helping to make this webinar possible. I would also like to cover a few logistical items.

We will be recording today's session for future playback, and the link to the recording will be posted on JRSA's website. We usually post that the following day, so it should be up tomorrow. Today's webinar is being audio cast via both speakers on your computer and teleconference, so we recommend listening to the webinar using your computer speakers or headphones. To access the audio conference, select, "Audio," from the top menu bar, and then select, "Audio conference." Once the audio conference window appears, you can view the teleconference call-in information, or join the audio conference via your computer. If you have any questions for the presenter or you would like to communicate ...

Hold on one second. Let me catch up with myself on PowerPoint. Okay, so there's the information of the call-in information, and then one more. Here we go. If you would like to communicate, if you would like to ask a question, feel free to post those throughout the presentation, and we will be ... On my part, I will be sharing those with Doug as we go through to see if he can answer them at that time. We can also hold questions to the end if anybody would like either/or. Feel free just to chat, add a question by going through the chat feature, and here we would prefer that you either do all participants, which is right there in blue, or host presenter and panelists.
Erin Farley:
This session is scheduled for approximately one hour, maybe. Usually, sometimes we do go over to one, between one and an hour and a half, depending on the questions. We hope that you guys are able to participate, and hold on to the end, because we do have polling that we save until the last five minutes. We would love to get your feedback on what you liked about the session. When we do do that, please take a moment at the end to fill out that poll. If you do have any technical difficulties throughout this webinar or you get disconnected, you can reconnect going through the same link that you received to join, or you can also reach out to Jason Trask, at jtrask@jrsa.org.

Erin Farley:
Okay, and so I think that is it. What I'm going to do ... Oh, one more thing. What we are doing, this is kind of a recent thing we've been adding into our webinars, is that we know that we have people who are viewing as a group. We are really trying to get a close to accurate number on how many people are actually viewing during the live webinar. If you do have a group with you, we would ask that you please go to the chat window and type in the name of the person registered, and the total number of additional people that you have in the room. This, again, just really helps us keep track of how many people are actually attending the webinar, and we really appreciate it. So thank you.

Erin Farley:
All right, and so with that, I'm going to pass it off to Doug. With that, I'm going to drag the little magic ball and give it to Doug. It is ... Whoops, wait. Is it working? Hold on one second. Yes. Here we go. Okay. All right, Doug. It's all yours. Thank you.

Doug Spence:
Thank you, Erin, for the introduction and for the magic ball. As she said, I'm Doug Spence, talking today about Area-Under-the-Curve-analysis, which is an increasingly commonly used technique in criminology and really in a variety of fields that are interested in evaluating the performance of predictions. What I hope you guys will all see today is that this is a technique that's not only pretty easy and straightforward to use, it's something that has a lot of utility and a lot of different applications in criminal justice. Everything from evaluating risk and needs assessments, and doing recidivism analysis, to outcome studies, and all sorts of stuff.

Doug Spence:
Before we get started today, I do want to give you guys a little bit of an overview of what I'll be talking about. We're going to start by talking a little bit about the background and historical context of AUC analysis, 'cause it's something that comes from signal detection theory. We'll talk a little bit about sort of the broader issue of evaluating binary predictors or classifiers. Also talk about the receiver operating characteristic curve, and that is the curve that Area-Under-the-Curve analysis is looking at. Then we'll get into the AUC statistic itself and how to interpret it. Then I'll walk everyone through how to conduct AUC analysis and SPSS. We'll wrap up by talking about ways you can use AUC analysis, and some interesting and useful ways to supplement and enhance analysis results that you might be doing.
Doug Spence: All right, so to start off here with some background and historical context. AUC analysis was first developed in World War II by basically radar operators, so that ROC curve originally stands for radar receiver operating characteristic curve. Basically what they were interested in doing was figuring out a way to maximize the performance of radar. They wanted to basically maximize the amount of correct predictions that they would have. In other words, the true positives, so that they would spot every enemy aircraft that would come in, so they would spot every attack. But at the same time, they also wanted to minimize the number of false positives, because from their perspective, you could imagine, you don't want to miss any attacks, but at the same time, you don't want to be scrambling the jets every time there's a flock of birds, or an especially ominous cloud that comes in.

Doug Spence: They want to minimize those false positives and maximize those true positives. By focusing on both of those things together, they developed a technique with the ROC curve that proves to be very useful for a lot of other fields. For medicine, where they're interested in basically maximizing the number of correct diagnoses that they have, but not freaking people, too many people out, with a bunch of incorrect diagnoses. Psychology, and then as you guys know today, criminology and criminal justice. This is really sort of the context that it's coming from. It's this dual emphasis on measuring the amount of true positives, but also taking into account the number of false positives that really makes the ROC a useful tool.

Doug Spence: Now, before we get into the ROC curve and the AUC statistic, I do want to talk a little bit about classifiers, because this is sort of where a lot of this language is coming from, is from signal detection theory. It can be something that's a little bit foreign to some of us. Basically when we're talking about classifiers or binary classifiers specifically, what we're talking about is really any type of rule or procedure for classifying cases in terms of their predictive outcome, or for classifying cases into two groups. One group where there's a positive outcome, or a one, and one group where there's a negative outcome, or a zero. We're talking about sort of classifying cases in terms of a dichotomous outcome. In statistics, it's most often the case that classifiers are going to be your predictive probabilities produced by some type of statistical model, usually a logistic regression analysis. That's the context that we're using here. We're basically using ROCs, and the AUC statistics that are calculated from that to measure the performance of a classifier, which in this case is going to be your predictive probabilities that you produce from a logistic regression model.
Doug Spence: Now, one aspect of this, because we're dealing with predictive probabilities, we're dealing with the probabilistic classifier. This is something that is going to have varying degrees of performance depending on where you set what's known as the discrimination threshold. The discrimination threshold is just the level at which you say that a probability has reached a point where you're going to consider it to be a prediction of a positive outcome, or a one. Most intuitively, people would say, "If it's a probability of .51 or more, that's a one. If it's a probability of 49 or below, that's a zero," or something like that. But, wherever you set that, is going to potentially change how accurate your predictions are. That's sort of a complicating factor that goes into evaluating how effective our predictors are.

Doug Spence: Now, what you can see here in terms of the little screen capture I have from SPSS, you can see what this might look like in terms of your data. In that first column, you've got the predicted probabilities that you might save from a logistic regression model, so that pre-one. Those are all just probabilities ranging from zero to one of what the model thinks the likelihood that that case is gonna be a one is. You can have a predicted outcome that you establish based on some sort of discrimination threshold, so that column there is just a one if it's above .5, and a zero if it's below that. Then you have the actual outcome, which is ultimately what you're gonna be assessing your classifier against. How well does it predict whether cases were actually a one or a zero? This is what we're talking about primarily, your predicted outcomes, or your predicted probabilities that you're gonna get from logistic progression analysis.

Doug Spence: When it comes to evaluating your classifiers, there are a couple things to consider. One is that for any given case, you're gonna have essentially four different sort of results off of your classifier. You can get a true positive, which means that your classifier predicted that it was going to be positive, and it actually was positive. You could have a false positive, which means that your classifier predicted that it would be positive, and it wasn't. You could have a true negative, which is that your classifier predicted that it would be negative, and it was actually negative. Or you could have a false negative, where your classifier predicted that it would be negative, and it was actually positive. There are these four different outcomes that you can see over there in that figure to the right. Coincidentally, that figure to the right is called a confusion matrix, which I think is one of the cooler names for figures that are out there. This is basically what's breaking down predicted class versus actual class, and the different outcomes that you can get.
Doug Spence: Now, if you're going to try and evaluate a classifier's performance, there are a lot of different ways to potentially do that. One way would be to calculate the true positive rate, which is sometimes described as the sensitivity. You'll see that in, from you SPSS output. It'll call that, "Sensitivity," instead of saying, "True positive rate," but that's all that is. That's basically the proportion of positives that were predicted correctly to be positives. You could have a false positive rate, that's the proportion of negatives that were incorrectly predicted to be positive. You could have a true negative rate, which is also sometimes described as specificity. That's basically the number of negatives that were correctly predicted. Then you have the accuracy, which is just the number of true positives and true negatives as a share of all the positives and negatives that are predicted accurately. Each of this is individually a good way of assessing the quality of the classifier, but there are some limitations to just using these, because basically, each one of them doesn't consider both true positives and false positives together in conjunction.

Doug Spence: You'd have to calculate both, and you sort of have two numbers there. That's also the case if you wanna calculate these for each different discrimination threshold that you might potentially set for your classifier. That means you have to calculate for every single predicted probability on your data set. Really what we want is a way of simplifying all this and condensing it down, so that we can have one number or one figure that basically takes into account both true positives and false positives, and summarizes all of that information across all the different possible discrimination thresholds. Fortunately, thanks to the efforts of those radar technicians in World War II, we have that with the ROC curve. The ROC curve, really all it does is just plot the true positive rate for a given classifier against its false positive rate varying the discrimination threshold. This takes one classifier, so one set of predictive probabilities from your data set, so the results from one particular model. It plots its true positive rate and false positive rate as a curve, then each point is a different discrimination threshold, or a different probability that exists in your data.

Doug Spence: The way you would interpret this is that basically, the performance of the classifier is gonna get better as you move closer to that top left-hand corner of the figure. If you were in that very top left-hand corner, you'd have a true positive rate of one, which means that you correctly identified every single positive that was in the data. Then you would also have a false positive rate of zero, which would mean that you did not incorrectly identify any of the negatives as positives. You'd be perfectly accurate in that case. In most cases, when we're actually looking at real classifiers is that they're somewhere down there kind of in the middle. It's fluctuating along at different discrimination thresholds. You could use an ROC curve like this to identify the particular sort of threshold that would be the most effective for maximizing the number of true positives and minimizing the number of false positives, or you can also use it to use to just sort of get a visual sense of how well a particular classifier is performing.
Doug Spence: What this will talk about in SPSS, you could actually plot multiple classifiers in the same ROC curve, so you could compare the curves to one another. See if one is sort of bending out more to that top left-hand corner than another and vice versa. Now, to just give you guys a little bit of a, sort of a discussion of how you might interpret one of these. If you're gonna look at this, basically if you looked at this curve and started at the top right where you've got a discrimination threshold of .30. That's basically saying, "What would your true positive rate and false positive rate be if you said that every case in your data that had a probability of above .30 would be considered positive?" What this is telling you is that it would basically allow you to get 100% of all the true positives out there. You wouldn't miss any positives, but if you look down at that X axis, false positive rate would be 0.9, so you'd get 90% of the negatives in there as well.

Doug Spence: If you think of this in a criminal justice context, this would be if you considered above a score of, above a .3 probability of recidivism as being high-risk, and you classified everybody above that as a high-risk offender, then you would get 100% of the recidivists in terms of this data, but you would also get 90% of the non-recidivists as high-risk, as well. You're basically classifying everybody as positive in this case. But what this curve does show you is that if you just increase that threshold in this sort of hypothetical case to .38, your true positive rate only goes down to .8, so you're still getting 80% of the true positives, but your false positive rate drops pretty dramatically to .5. You miss a lot of those potential false positives. This suggests that even a modest increase in the threshold here would probably give you a much better prediction. You can see that visually, it's just sort of moving over towards that top left of the corner.

Doug Spence: That, in general, is what you're gonna get when you get an ROC curve from SPSS or from some another tool. It's gonna visualize this type of performance for you. Okay, so if that's the ROC curve, in terms of the AUC, what we're getting is a single number that summarizes the performance of the entire curve. What the AUC is just the area of this thought that falls under the curve. As it approaches one, you're basically approaching having all of the area, or 100% of the area underneath the curve. What that would look like visually would be a curve that's all the way in that top left-hand corner. As it goes more towards 0.5, that would be a curve that only has half of the area of the figure underneath the curve. That is really all the AUC statistic is. It's just a measure of the proportion of the plot area that's under the curve of the ROC curve. This essentially summarizes the ROC curve across all of those different discrimination thresholds.
Doug Spence: The AUC doesn't really care what's the optimal threshold. It's not gonna be something that you can use in itself to identify that. You'd have to actually look at the curve to do that. What it's doing is just summarizing that classifier or summarizing that predictor over all of the potential discrimination thresholds. Now, in terms of how this can be interpreted, this is a number that's gonna range from zero to one, meaning that zero would be none of the areas under the curve, and one is all of it is under the curve. Mathematically, it's equal to the probability that a given classifier will rank a randomly chosen positive case that's a true positive higher than a randomly chosen negative one, which is kind of a technical way of saying that it's given you sort of a sense of the probability it would classify a randomly selected case correctly. For example, if you had an AUC of .6, it's basically saying there's a 60% change that any random case you select is gonna be classified in the right way. Or in other words, roughly getting about 60% of the cases correctly predicted or correctly classified.

Doug Spence: Now, as far as sort of what are common standards for good or bad AUCs, the most common one you see out there for an effective predictor is an AUC of .7, so this is something I've seen in a number of places of the most commonly used benchmark. That's something to shoot for potentially. One thing that's very important to point out is that even a predictor that was essentially chance, a coin flip or something no better than chance, would have an AUC of .5. That's really the floor of what you should expect to see. That would be a very bad AUC if you had an AUC of 0.5. It's possible to go below that, and it's possibly to be, it's predicted worse than chance, but just as sort of a floor for what you should expect from producing AUCs, .5 is basically saying you're not having any predictive accuracy. .7 is how I'm gonna use this. If you're getting around that, it's a pretty good predictor. The obvious caveat to that is if the standards vary by field and application. Medicine, for example, they might want something like .9, because they're worried about additional treatments.

Doug Spence: But most common in what I've seen is about .7. Now, one other thing to note, just if you look at this figure here on the right. It's demonstrating some, basically two different ROC curves. What you can see from this figure is that, what I was saying earlier is that the AUC statistic doesn't take into account any particular individual discrimination threshold. What you've got there is in that one curve, B, that sort of bends out father than the curve behind it, A. That's one that's gonna obviously have a larger AUC statistic, at least a little bit larger AUC statistic than curve A. B is gonna have a higher AUC than ROC curve A. But ROC curve A, at that one particular discrimination threshold that you can see there, it's that one point where curve A meets curve B, has the same predictive accuracy, or the same ratio of true positives to false positives as B. But B as whole is better at predicting at all the different discrimination thresholds than A is.
Doug Spence: Somewhat of a technical point, but it is an important thing to point out, that just because you have an ROC curve that has a higher AUC, it doesn't necessarily mean that it's always the better predictor. It's just the better predictor across all of those different discrimination thresholds taken collectively. That's the basics of what the ROC curve is. It's really just looking at true positives and false positives and discrimination thresholds. AUC is just the area under the curve. That being said, here are basically the steps for how you would conduct an AUC analysis after a logistic regression in SPSS, which I think would be the most common way that you would wanna do this. These are all here for reference. Basically what this boils down to is that you're gonna conduct a logistic regression. You're going to save the predictive probabilities from that logistic regression, then you're gonna open up the, "Analyze-ROC Curve." Then you in SPSS and plug-in the predictive probabilities and your dependent variable.

Doug Spence: In terms of what this could look like in SPSS, if you look here, if you go at the top pull down menu, click, "Analyze and logistic regression," and it will bring up this logistic regression menu that you seen on your left. I plugged in, in terms of the dependent variable there one of the recidivism measures that I had from that study on recidivism. I plugged in, just for example, one covariate, and I will see a minus score, which is a risk assessment score. That's setting up a logistic regression. Then, this is the important step, is that you'll wanna click the, "Save," button there on the right. That's gonna bring up a secondary menu, and it's gonna allow you to save particular aspects or features from your logistic regression. Now, what you're gonna wanna check there is the predictive probabilities. For those of you who have done this before, what it will do is just basically produce a new variable in your data set that's gonna be comprised of the predictive probabilities for each case that the dependent variable, in this case where they were booked into a regional jail, is one for that case.

Doug Spence: That's gonna be the predictive probabilities produced by that model. You run a new model and click, "Save," again, it's gonna give you a second set of predictive probabilities that's specific to that model and so on and so forth. It's gonna create basically a new variable in your data set that is the predictive probability. That is gonna be the classifier that you're gonna evaluate with your ROC analysis. Once you've done that, then you're gonna wanna go back to that pull down menu, click, "Analyze." Hopefully if you've got a good new version of SPSS, it'll have the ROC curve option there for you. You go down there, click that, and then you will have the ability to plug-in from any of the variables in your data set what you want your test variable to be. The test variable is going to be your classifier. I put in the predictive probability that was generated from that regression. This is a thing you'll, it's important to keep straight if you're doing this with multiple models which predictive probability you're looking at, so you may wanna label them and do all that.
Then in the state variable box, you're gonna put your dependent variable, and the value of the state variable that you're trying to predict, which in this case, it's dichotomous. Zero and one. You're gonna wanna type in a one. If for some reason you had dichotomous variable that was one and two, and you're trying to predict two as the outcome, you're gonna wanna type in two and so forth. You need to know what your variable is that you're using there, but you're gonna type in the value that you're trying to predict. Below that, you've got some options for what you can ask SPSS to display for you. It may, it won't automatically display the ROC curve, it'll just give you that AUC number. The AUC number's fine, but if you wanna see the ROC curve, which I usually do, you can just check that box. Then you can also ask it to give you a diagonal reference line, which is just going to give you a line that, sort of that bare minimum of what an AUC of .5 would look like.

It's just gonna be a line that runs from the bottom left of your ROC curve plot to the top right. A diagonal line across there. It's sometimes usual to let you sort of see where your curve is in relation to that sort of bare minimum threshold there. There are also options to get confidence intervals for the AUC statistic where you actually get it to tell you the coordinate points of the ROC curve, that those aren't as commonly used. Now, in terms of the outcome that you'll get, you'll get something that looks like this. On the left, you've got your ROC curve plot. This is from that model that I ran there. As you can see, it's kind of a sad ROC curve. It's not the most powerful predictor, but it is substantially different from that line there, that bare threshold. So predicting better than chance. You can see from the area under the curve box there on the right, which is your output for bearing under the curve statistics, the main thing you're gonna be looking at is this area box here on the far left.

That's gonna give you your AUC statistic. That's where that is. As you can see there, that's an AUC of .261. It's not the best AUC, it's a little short of that .7 threshold, but it's above .5, so it's predicting outcomes about 62% of cases correctly. I've also asked for some additional stuff here, the confidence interval and all of that, but it's usually not something that is as commonly used or presented, although it is sometimes. That is the output from SPSS, and that's basically what you can expect to get. In terms of sort of why you would wanna do this, there are a few things to keep in mind. One is that the AUC provides a really good summary of the predictive accuracy from limited dependent variable models, or models that are looking at dichotomous dependent variables. That's one limitation or issue that anybody who's done a logistic regression knows is that it's not quite like a linear regression, because you can't interpret the coefficients directly.
Doug Spence: You don't have a real R-squared. There are some pseudo R-squared that SPSS will generate for you, but they do have some issues potentially. You don't have some of those things are as nice and useful to have as you would get in our sort of standard OLS linear regression context. What this does is it provides you a good, easy sort of post-estimation thing you can do to get something like an Rsquared for you logistic regression model. The important thing to point out about this is it's not affected by base rate. That means the sort of underlying rate of ones compared to zeros in your data set, or positives compared to negatives. This is an issue because a lot of other sort of similar measures, pseudo R-squared and other things like that, that are trying to get a predictive accuracy for dichotomous dependent variables are affected by base rates, which means they aren't as reliable when you have a low base rate.

Doug Spence: In the criminal justice context, our base rates are often pretty low. If we're looking at recidivism, it may be the case that a minority of people recidivate, most people don't. If we're looking at who's high-risk or who successfully completes something, they often have situations where the base rate can be pretty low or inconsistent. The good thing about the AUC is it's not affected by that at all. Any time you see an AUC statistic, it's given you an accurate measure of how many cases are predicted correctly by the classifier that you're looking at. Some other big advantages of the AUC is it can be interpreted directly. When you see an AUC of .7, that's a random case that has a 70% chance of being correctly predicted. In other words, roughly about 70% of cases are correctly predicted. That's regardless of the context, regardless of how the dependent variable is measured, what the base rate is, anything like that. It always means what it means.

Doug Spence: That's the nice advantage over something like a law of likelihood, which can be used to measure sort of the goodness of fit for a logistic regression model, but there's gonna be a number you can't interpret directly. It's gonna be kind of meaningless unless you do something like a likelihood ratio test to compare two different models to one another. Another advantage of the AUC is that you can compare across models. It doesn't matter if you're looking at models with different dependent variables or dependent variables that are measured in different units on analysis. AUC is always gonna mean the same thing anytime you do it. That's a nice feature to have. Now, in terms of what this opens up in terms of analytical possibilities are a number of different things. One is that I find it useful for assessing model specifications of logistic regression models, because it's gonna play that role kind of like an R square would.

Doug Spence: If you're estimating a model, and you're wondering if you should include an additional variable or not, you're not quite sure about that variable, you can run a model without it, calculate the AUC, put that model in, calculate the AUC again. If the AUC stays about the same, you know that model didn't really enhance your predictive accuracy very much. Conversely, if the AUC jumps
dramatically, you know, okay, that's an important variable that you're gonna wanna include. It's nice for that purpose. You can sort of tweak your models a little bit, run them a different way and see the AUC, and see how it's changing that. Another good use for the AUC is to compare the performance of different predictors for a given outcome. For example, if you have a couple of different risk assessment tools, and you're wanting to see how well they predict recidivism, you could run models with each of those and see how they predict, and compare those AUCs directly.

Doug Spence: The one that has the highest AUC is the best predictor. That's a useful tool to have. Then you can also look at it the other way and use one predictor with different outcomes or population. You could see does that risk assessment tool predict different measures of recidivism in the same way? Or does it predict things like misconduct in prison or other sorts of outcomes you might be interested in predicting? You can also run the model with subpopulations from your sample, and see, okay, is it just as predictive for males as for females? Or is it just as predictive for different types of offenders with different types of needs or offense types, or something like that. It's gonna open up a lot of different possibilities for comparing models to one another with logistic regression. That's nice, because without the AUC, you really, all you're stuck with is just sort of comparing odds ratios for variables. Those things are more prone to fluctuate, and a little bit more difficult to say if you're just saying, "Oh, this variable increases the risk of recidivism by 5% in this model and by 8% in another model."

Doug Spence: It's not as intuitive as saying, "This variable accounts for 60% of cases in one model and 70% in another case," or something like that. I guess here to sort of wrap up and show you guys here som examples of where I've used this in the past. This first table here is from a report that we did looking at regression or recidivism by Day Report Center clients in West Virgin. The table's sort of broken up there 'cause this was originally a much bigger table with more variables between, but we can see here you've got the usual logistic regression stuff. We've got coefficients, and odds ratios, and standard errors, and all that. Then at the bottom, we've got the n for the sample size. We've got a pseudo Rsquared, and then the AUC statistic. What you can see here is that you've got a comparable sort of measure of predictive validity for the model as a whole across three different dependent variables. You've got arrests with an AUC of .68. Jail bookings with an AUC of .68, and incarcerations with an AUC of .82.
Doug Spence: What that allowed us to say was that what we've got here when we consider this model as a whole, we're predicting pretty close to that .7 threshold. We've got a pretty good prospective model of predicting recidivism when all these variables are included together. Our model predicts, 'cause it's the same model in each of these, our model predicts arrests at about the same rate that it predicts bookings. They're both performing fairly equally well, but it predicts incarcerations significantly better. That's something that was intuitive to us, because one problem we have in West Virginia is we don't have very good convictions data. Our arrests and bookings measure, not all of those people actually committed crimes or were convicted, but incarcerations is much closer to what would be the ideal with convictions. We know that pretty much everybody who gets incarcerated was convicted of a new offense. We would expect that AUC to be higher in that case. That's what we see here.

Doug Spence: You're gonna get a sense of how you can use this as sort of a, kind of something the equivalent to an R-squared, not only in a regression context, but for logistic regression. Now, this next table is where we did something a little bit different, and a little bit more creative with the AUC. What we did here was basically we ran a bunch of different models. These are all AUCs from separate models. That's a total of 27 in all. 27 different models. It was just a bit of a tedious day in the office that day running all these models, but we've got 27 different AUCs here. In the first row, you've got the AUCs from a model that includes only the total risk and need score from the risk assessment pool that we were interested in, which is the Level of Service/Case Management Inventory, LS/CMI. What we can see here is that with just the total risk and needs score from the LS/CMI, we have AUCs of about .62 for arrests and bookings, which is pretty close to what we had with the full model, which is about .68.

Doug Spence: What that allowed us to say was that, "Look. If all you had for an offender was just this one number from the LS/CMI, we can predict almost as well as a full model that has a whole gamut of other variables, whether it be demographics, offense type, experiences in the Day Report Center program, length of stay, successful program completion. All of this other sort of stuff, all of that collectively only gives you an additional 6% or 7% of cases that you can predict. All of that additional information only helps you get another 7% of cases. You get most of those with just the LS/CMI number. That was a powerful thing to be able to say, that you even know this AUC isn't all that great. Taken in itself, it does suggest that the LS/CMI is a pretty good predictor and that all this other information doesn't add that much to it. Now, what we did below that was we took the eight subsections of this tool, and we did analyses of each of them individually.
Doug Spence: Each of these AUCs in this case represents the sort of predictive power of each of these subsections individually. This allows us to compare and see which subsections appear to be the best predictors of recidivism, and which were the most effective. Also, identify potential areas of weakness and compare across all these different recidivism measures. A few things that jumped out in this is that we could see that criminal history was the most effective AUCs around .6 and up, up to .67 for incarceration, but it also let us see that quite a few of these are pretty low. The most surprising and the lowest was for pro-criminal attitudes, which is basically performing at about the same effectiveness as chance. It's at about .5. It was pretty much about as bad as you could get for an AUC for this particular subsection. That was a big deal because, for those of who are familiar with the LS/CMI, the pro-criminal attitude is tapping into a lot of those cognitive behavioral things that we really think drive offending behavior.

Doug Spence: That is really the target of a lot of our interventions. The fact that this wasn't seen to be a good predictor was a big potential issue. It's also a part of the risk assessment tool that is one of the more difficult to assess, because it's sort of less objective and requires the assessor to really try to get at what they think the attitudes and beliefs of the offender are. This was something in conjunction with some other results that we had from a similar study of inmates led us to recommend an additional inventory that would be used to try and help assess criminal attitudes and thinking in order to help supplement and improve these results. We're working on some trainings that will help improve performance in this area, as well. This sort of helped us hone in on particular parts of this risk assessment tool using AUC analysis and see where these were stacking up. Now, final thing I wanna talk about is this another example from an article that did something very similar.

Doug Spence: Basically, this is what inspired us to do that previous analysis. We weren't entirely that creative on our own. We sort of saw a good example of this thing done in another previously published study. What they're doing here is they're looking at three different risk assessments and three different measures of recidivism, and calculating AUCs, but what I they're also doing, which I think is kind of neat is that they're breaking up the sample. They kept the total sample there on the left, and then also breaking up the sample by gender. It allows them to compare the predictive accuracy of each of these assessment tools against each other, but also for different subpopulations in their sample. It's a nice use of this technique. You can see here, it does make this pretty intuitive. We can see, okay, I look at women for felonies. All of these assessment tools are predicting less than they do for men. Or if look at this recidivism risk three, I can see that it predicts any offense roughly the same for women as for men, and for the total sample as the whole and so forth.
Qualitative Research Methods for Policy, Practice, and Research
April 20, 2017

Doug Spence: It allows you to make those comparisons in a relatively nuanced way without having to throw in a whole bunch of different analysis and test statistics and all that sort of thing. That should hopefully give you all a pretty good understanding of sort of what's going on with the ROC curve and with the AUC statistic that comes from that. Hopefully, which concedes the message that in ROC curves and AUC statistics, they provide a pretty good, effective means for measuring the accuracy of predictions, and that there are a lot of potential applications for this. I know we focused here primarily on risk assessments and recidivism, but you could do this with any dichotomous outcome that you're interested in. With any non-dichotomous outcome, if you could find a way to sort of dichotomize it in a useful and sound way, you can apply these same techniques. Well, that's also something that I think is a very nice compliment to logistic regression analysis.

Doug Spence: Whenever I do a logistic regression, I'm pretty much always gonna do an AUC, just to get a sense of outcomes, sort of the predictive effectiveness of the model as a whole, and to have that sort of R-squared-like thing to have in a logistic regression context. But before I do sort of end today, I do wanna add one note of caution or thing to think about going forward. That's that really this question of, "What constitutes a good AUC value?" Is something that's still an open, empirical question, right? It's something that it varies by field, and that while there is this norm of .7 out there floating around that I've seen in different numbers and different publications. It's something I think is still sort of yet to be determined. I would caution anyone from sort of taking this technique and looking at your notes. Saying, "Okay, we've got .7. We're good." Maybe .7 isn't really good enough, because you could still have .8 or .9, and really the question of what constitutes an effective prediction really depends on what you think your goal is for predictive accuracy.

Doug Spence: This is just a measure. It shouldn't be treated in the same way as some other statistical test where you're just trying to get above a certain threshold to say that something's significant or say that something matters. Here, we really are getting a concrete measure of how many cases you're predicting. Having 30% of cases or 40% of cases that aren't predicted might not be something that we can be satisfied with. We wanna maybe push that AUC higher, or conversely, if you're working with something that's really hard to predict, maybe something that's just better than .5 is good enough. I'd say a word of caution. Sort of being open about how you're gonna interpret that AUC is in order. Okay, well, that's all I have for today. Ready for questions if anybody has those.

Erin Farley: Thank you, Doug. Let me just scan down to see if anyone ... Okay, so there is a question from [Rachel 00:39:56]. "Is there a minimum n that is required to run ROC?"
Doug Spence: I'm not aware of a minimal n. In general, I think with most statistical tests, I get a little uncomfortable if the n goes down below 100. I don't know if you had 50 or 30. It really depends on how confident you are running the logistic regression, 'cause that's ultimately what's gonna provide sort of the input for it. But I would sort of treat it like any other statistical analysis that you would do.

Erin Farley: We have another question. "I am often asked how much better .7 is than .5. Is it fair to say that .7 is 40% better than guessing?" And in parentheses, it says, ".2/.5."

Doug Spence: Yeah. I mean, I don't know if you would say 40% better than guessing, because I think what makes the AUC useful is that it is pretty directly interpretable. The technically correct interpretation would be that you can predict, your odds of successfully predicting the outcome for a randomly selected case are 20% more, because it's 70% as compared to 50%. You've increased the odds that you will correctly predict cases. I think it's okay to say that you're predicting an additional 20% of the cases in your data set. That's the frame of reference I would use. But I can see what you're saying, that it's sort of increasing the amounts predicted by about half. Then it gets a little bit confusing, 'cause you're talking then about a proportion of what's already a percentage. I would kind of just use a base percentage, but I think probably the safest would be to say it's an effective predictor at that point. It's significantly better than chance.

Erin Farley: Okay. [Lonnie 00:41:53] says that she doesn't want to ask a question. She says, "Thanks." Then one person, this is kind of not related, but asked about the slides in recording. I was hoping to get everybody, send people out the PowerPoint this morning. If that didn't go out, then we will have everything available tomorrow on our website, on JRSA's website. If you go to our page on webinars, you will find it there sometime tomorrow. We have another question. "I'm not clear about how ROC is different from R-squared. Aren't they both talking about how good the model is?"

Doug Spence: Oh, there is a difference, because the R-squared, if you think about a linear regression context, it's really about the proportion of variation and why that you're accounting for with a regression line. It's looking at the distance between each of your data points and the regression line. The ROC curve is not so much about variation in values for the dependent variable as it is sort of the ratio of true positives to false positives. I guess maybe the simplest way to think about it for me is just to say that with the R-squared, it's about how close your sort of predicted value is. What you predict Y to be with your linear regression model to the true value of Y. The true value of Y might be 25, and you predict 23. You're off by two. You're dealing with a dichotomous, nonlinear situation when you're working with logistic regression. What you're really seeing with the ROC line is that you're correctly predicting if you've got an AUC of .7, you're correctly predicting 70% of cases. You're correctly getting one, seven percent of the time.
Doug Spence: You're right on the money a certain amount of time, or you're completely off. You're either right or wrong with a dichotomous dependent variable. You could potentially have a situation with R-squared where your R-squared could be a lot just because your predictive values get closer to your actual values, but you maybe don't actually predict more cases correctly. But with the AUC, it's about the number of cases rather than the value for the dependent variable. Don't know if that's clear or it makes sense, but ...

Erin Farley: Oh, wait. Okay. That person says, "Thank you." Then we have another question. "The field of predictive analytics seems to be exploding with a dizzying array of techniques and performance measures. Logistic regression and the AUC feel understandable and comfortable, but is this approach still considered acceptable in the criminal justice field, and for how long?"

Doug Spence: I would think it is. I think it's here to stay, I would think, because it is an effective measure and it is one that has these advantages. It's not a new technique, but I think it is relatively new in criminal justice as far as within the past decade or two. I would think that it's here to be around. I'm not personally aware of something that's superseding it as of yet.

Erin Farley: Well, I can say when I was in graduate school, I was just talking about this with a colleague. They did not teach AUC analysis, but just within the last year, I feel like I'm seeing it in so many different places within criminal justice. It makes me feel like it's being utilized so much more now, because just 10 years ago, I had never even heard of it. I think that this is definitely, researchers are utilizing this technique. I feel like we're gonna keep seeing this and seeing it more in the future. Okay. We have more questions. [Steve 00:45:54]. "Many of us work with three or four level classification instruments. Of what use is AUC with such instruments?"

Doug Spence: Let's see. That's with multiple outcomes, I guess, is the, sort of that question there.

Erin Farley: I think. Yeah, that's what I think. I mean, it's limited, right? 'Cause it can really only be utilized to interpret the predictive probabilities from a logistic regression.

Doug Spence: Right. You would have to maybe get creative and sort of break it down into steps or sort of figure out a way to dichotomize it. If you had something where you could have sort of a process where somebody's being classified as one thing, and then another thing, and then another thing. It's contingent on what they were classified already. You could look at each step individually and figure out some way to collapse them into groups that are dichotomous. But yeah, this isn't gonna be a technique that you can use with multiple different outcomes.

Erin Farley: I'm sorry. Okay, Steve, you wrote ... I misinterpreted. "Not multiple outcomes." He's saying, "Low, moderate, high, very high."
Okay, yeah. Well, I think one thing to think about too is that if you're thinking about sort of the risk assessment tool, that's the predictor or that's the classifier. It's really the dependent variable that has to be dichotomous. As long as you've got a dichotomous recidivism measure, recidivated yes or no or something like that, then you could still use AUC analysis. It's just that, the logistic regression, when you save your predictive probabilities, is it's gonna give you basically an interval probability that might not be perfectly tied to the threshold. You could do an AUC analysis just on the tool itself or the thresholds. You can have a variable that's 1, 2, 5, from very low to very high or something. You could do an AUC analysis on that, it's just it would give you a curve that would basically just have five steps. It'd be a crude looking ROC curve, 'cause it would just be your first discrimination threshold is a score of one or very low. Your second discrimination threshold is a score of two or low. Your third discrimination threshold is a score of two or low. Your third discrimination threshold is a score of three and so forth.

You could still do it, it's just not gonna be as sort of curvy as what you would get with predictive probabilities.

[Kelly 00:48:25] has a question. "What are your thoughts on using the AUC statistics for validating pretrial risk assessments, particularly given that you don't have a true false positive rate for those that are high-risk and detained, but would not have reoffended if released?" I can repeat that if ...

Yeah, maybe one repetition would be good, actually.

"What are your thoughts on using the AUC statistics for validating pretrial risk assessments, particularly given that you don't have a true 'false positive' rate for those who are high-risk and detained, but would not have reoffended if released?"

Yeah, it's saying there is you've got people you've identified as high-risk, but then you remove them from the possibility of actually becoming a true positive because they've been pretrial [crosstalk 00:49:24]. That's a tricky one, because obviously they can't become an actual false positive in that sense. I would think that the only way you could evaluate that would be to look at everybody who didn't get that. What you've got is essentially a skewed template, because your individuals that are identified as the highest risk are removed from the sample. But that would be sort of the same case if you were doing a recidivism study and the highest risk people weren't released. They're just not gonna be able to be observed. That's just gonna be a limitation to that study, but as far as a method, I think AUC would still be valid for assessing your false positive rate, or your ratio of your true positive, false positive rate for that population that's still out there and at risk of recidivating.

It's just that the limitation would be that you know you don't have the true population, 'cause you've got this measure in place [crosstalk 00:50:20].
Erin Farley: No more messages coming in right now. Quick question. I was wondering, that article that you referenced. That 2009 article from Brennan, Dietrich, and I think it's Ehret? I was wondering what journal that was from? People might be interested in seeing if they can find it. Do you remember the [crosstalk 00:50:48] article?

Doug Spence: Yeah, I've actually got a [crosstalk 00:50:50]-

Erin Farley: Or the, I'm sorry. The journal.

Doug Spence: Yeah, here.

Erin Farley: Oh, okay. Great.

Doug Spence: From Criminal Justice and Behavior.

Erin Farley: Great. Thank you. While I hope I still have some people, I'm gonna jump a little bit and go and open the polling so that people can do that before they leave. Let's see. Give me one second. It's not working. Let's see. Okay. Oops, hold on one second, guys. There's usually a, "Run poll," thing that we can't seem to find. Oh, I know why. I don't have the magic ball, guys. Okay, hold on one second. See, when you come to a JRSA webinar, we always like to have some fun here. If you, Doug, if you can drag the magic ball to my name that would be great.

Doug Spence: [crosstalk 00:52:05]-

Erin Farley: I am now, all right. Here we go. Opening poll. All right, guys. There you go. Everybody who's still on, if I haven't lost anybody, please take a moment to complete our poll. You're guaranteed for entertainment when Erin Farley leads a webinar, that's for sure. Yeah. Okay, so let me just go back and check and see if we have any more questions before we finish. Okay. Looks like we don't have any more questions. Doug, I want to thank you so much for this presentation. I think it was really interesting. We will have this posted again on our website, hopefully tomorrow. If there are any questions, you can follow up with us or Doug. I think we probably have your email address on our website as associated with the West Virginia SAC, so I know that people can find you through that. I think that's it. Look forward to our registration opening for next month's webinar on ArcGIS. Thank you so much for attending, guys. Have a good afternoon.

Doug Spence: You too.

Erin Farley: All right. Bye.

Doug Spence: Bye.