Welcome, everybody. We'll go ahead and get started. Good afternoon. My name is Erin Farley, and I am one of JRSA's research associates. For those of you who may be less familiar with JRSA, it stands for Justice Research and Statistics Association. We are a national nonprofit organization dedicated to the use of research and analysis to inform criminal and juvenile justice decision making, and we are comprised of a network of researchers and practitioners which, at the core, include directors and staff from the state's Statistical Analysis Centers.

Before we go any further, I would like to thank our partners at the Bureau of Justice Statistics for helping to make this webinar possible, and I would also like to take a few minutes to cover some logistical items. Today, we are recording this session, and there will be a link to this recording available on our website. Usually, we are able to get a link to the recording the next day. If not, most likely the following day.

Today's webinar is being audio cast via both your speakers on your computer and teleconference. As you entered the webinar, you probably were prompted by a box that you can see here on the slide just asking you to select your audio preference. If you do find that you have any issues during the webinar, you can go and select "audio" from the top menu bar and then select "audio conference." Once that window appears, you can view teleconference call-in information and join the audio conference via your computer. If you do find that you continue to have problems or find that you are experiencing other issues, please contact Jason Trask at jtrask@jrsa.org.

If you do have any questions throughout this presentation, if you would either like to ask a question to the presenter or you would like to communicate with JRSA staff, we encourage you to submit these questions in the chat feature, and this you can find on the right side of your screen. When you do this, please select "host, presenter, and panelists" so that will go to me, Jason, and Kevin Baldwin, who is our speaker today, and, if you are in doubt, you could just pick "all participants," and that will go to everybody.

Then, the next thing to please ask is that you help us count. We do know that some individuals do watch these webinars as a group, and so we do try to track how many people are attending these webinars live. If you are viewing as a group, what we would really appreciate is, if you use the chat window, to type in the name of the person that's registered and the total number of additional people in the room. This, again, just helps us track the number of people who do attend our webinars.

This session is scheduled for an hour and a half. At the end of today's webinar, we will ask that you complete a short survey, and this information that you provide will help us to plan and improve future webinars and meet our reporting requirements.
Erin Farley: It is my pleasure today to welcome you to our webinar again. It's going to be on contemporary issues in risk assessments. This webinar will be presented by Kevin Baldwin. Dr. Baldwin is a clinical psychologist with a dual emphasis on research and forensics. He serves as a senior researcher for Applied Research Services in Atlanta, Georgia providing criminal justice research and policy analysis nationally. He has directed federally funded research projects and has authored over a dozen of articles and peer-reviewed journals and has both designed and evaluated substance abuse treatment programs, so we are very happy to have Kevin here with us today, and I will turn it over to you. Let me just drag the magic ball that gives you all the power, and there you go. Welcome, Kevin.

Kevin Baldwin: Great. Thanks so much, Erin. Welcome, everybody. Thanks so much for joining us today. We're going to talk about current issues in criminal justice risk assessment, and it's a pretty hot topic.

Kevin Baldwin: It's something that's been in the news a lot lately and, in fact, the subject of some rather sensational headlines. You'll see one there on your screen. It says "secret algorithms that predict future criminals get a thumbs up from the Wisconsin Supreme Court." Some of you may remember a movie based on a book called The Minority Report. It came out a few years ago, and there was a pre-crime unit, and the idea was that, essentially, you could tap these individuals to tell you who was going to commit murder. Headlines like this speak to that fear that we're going to be using tools to basically identify those people before they commit a crime, and then we're going to take them out of commission. The source of that article is just below the picture, and you'll see it's not a scientific press. It's actually the popular press.

Kevin Baldwin: The next article. Many of you may have seen this, actually. It was published by ProPublica, which does some really great reporting, and the title of the article was "Machine Bias." The subtitle: "There's software used across the country to predict future criminals, and it's biased against blacks." This article got a lot of press, and, in fact, the feedback and the fallout from this article continues to reverberate about a year and a half after it was published, so it's had a big impact.

Kevin Baldwin: There's also a New York Times bestselling book. I don't think we can say that about too many things of the things we do on a daily basis in our fields, but this book was titled "Weapons of Math Destruction" by Cathy O'Neil. Cathy O'Neil actually is a pretty accomplished mathematician and data scientist. She got PhD I believe from Harvard and taught maths there and elsewhere. The subtitle of this book is particularly interesting to me, and it reads as follows: "How Big Data Increases Inequality and Threatens Democracy." That's a pretty catchy title and a pretty threatening subtitle.

Kevin Baldwin: Risk assessment in particular within criminal justice systems and settings has gotten a lot of press lately. There's been a lot of other articles since these came
out a couple of years ago. I'm sure you've seen them. Let's take a look back though. Let's take about risk assessment is, where it came from. Essentially, we've all been engaged in risk assessment since the dawn of time. Risk assessment is something we do to survive. We wouldn't be here as a species if we didn't develop some pretty good means of assessing and scanning our environment and identifying risks so that we could either avoid them, if at all possible, or manage them. Likely, when we were kids, our parents or our caregivers taught us when we were crossing the road to always look twice both ways. That is a simple algorithm that is designed to manage risk.

Kevin Baldwin: Risk assessment, however, took a big step forward with two independent scientific advances. One is, essentially, an advancement in computing power. Some of you may be familiar with Moore's law. Gordon Moore, one of the co-founders of Intel, said basically that, per unit of time, the processing power doubles and the cost is cut in half, or something to that effect, meaning that there is an advance in computer science. It happens quickly, and what it does is it makes more and more computing power available to more people for less money than ever before. A parallel task is one of big data. Big data, you probably heard that term, refers to the acquisition of vast quantities of information, just huge data sets about us. Now, the government has had all sorts of data about us for some time. There are, however, now websites like Google and Facebook and others that contain a great deal of information about our activities because so much of our lives now are lived online.

Kevin Baldwin: What is risk assessment? We're going to define it pretty specifically today, and that is we're going to say that risk assessment is a process whereby we attempt to estimate the likelihood that an offender will recidivate, meaning go back to a pattern of criminal behavior, after the individual has already been caught doing so. First, they've got to be identified as someone who's committed an offense. Then, what we're interested in is the likelihood that they're going to do so again. When we call about criminal justice risk assessment, that's really what we're wanting to know is: How likely is it that this person before us is going to reengage in a process of criminal behavior? That's our definition, and we're going to stick with it.

Kevin Baldwin: Let's talk about the purposes of risk assessment, why we use risk assessment, and how we use risk assessment in the various aspects of our work. One of the things that we know is that risk assessment pretty much spans the entire spectrum of the adjudication process. When we start at the beginning, even at arrest, there are risk assessment tools that are administered in the field by police officers to determine whether they need even to arrest a person. Some of you may have heard of one of these. It's called the Hawaii proxy. It's four short questions that officers can use in the field to get an assessment, an understanding of the degree of risk for future criminality the person before them poses.
Kevin Baldwin: It's also used in the adjudication process throughout the different phases of adjudication. Excuse me. It's also in sentencing and dispositional purposes relative to pre-trial. In fact, pre-trial is probably one of the most common applications of risk assessment, and we're going to talk a good bit more about that later. Sentencing. That's something that's very controversial. We're going to spend even more time talking about that because it is so controversial. Custody level, institutional placement, supervision level. Imagine what prisons would be like if we didn't have a means of determining how risky someone was to engage in either violence or other undesirable behavior while imprisoned. If we had no information about that, we would stick everybody together. Prisons would be a whole lot more dangerous than they already are. That's some case processing application of risk assessment.

Kevin Baldwin: In terms of correction, just basically talked about that and in terms of assessing the security level, the appropriate level of security based on someone's assessed risk. When folks come out of prison or out of a period of confinement and we're talking about reentry and monitoring in the community, supervision, imagine what it would be like in terms of providing community supervision if you had no accurate way of discerning the risk of the person that you were supervisor. It would certainly be a lot more difficult to allocate your resources accordingly.

Kevin Baldwin: For sex offenders, some of you may be involved in work with sex offenders, either researchers or policymakers. Sex offender research really has driven in large part the field of community-based risk assessment in part because of federal regulations. Many of you might be aware of the Adam Walsh Act and the laws that preceded that in terms of Jessica Lunsford in Florida, the case that basically resulted in residency restrictions; before that, Megan Kanka in New Jersey. These are cases which were horrible, heinous crimes that resulted in federal laws which eventually became what we know as sex offender registration and notification, or SORN, as well as civil commitment proceedings. Those activities, in terms of registries and civil commitment, really rely in large part on the accurate results of risk assessments.

Kevin Baldwin: Finally, community support. It's important when you're working with offenders and when you're educating folks in the community to be able to let folks know that not all offenders pose the same degree of risk and that being able to understand the risks posed by offenders- For instance, with registration and notification, if you've got a sex offender that lives within a certain radius of your home, often times, it's helpful for you to know how risky this person is. Has this person been identified, for instance, as a sexually violent predator? That's something you're going to need to know or want to know, and that something is predicated on a risk assessment result.

Kevin Baldwin: The importance of risk assessment I think really can't be understated in terms of the fact that information on risk and our ability to effectively determine risk really has so much impact on how we deal with offenders and what we do with
them in the community. So many of the decisions we make in community corrections, in incarcerative settings, in adjudication processes, pre-trial decision making, and increasingly sentencing, are really predicated on the accurate assessing of risk of those that we're dealing with.

Kevin Baldwin: Let's take a little bit of a look back, a brief history of the different types of risk assessment that have been proliferated over the years. The first one is, according to Carl Hanson- And that's the slide you have before you. Carl Hanson is a leading figure in sex offender risk assessment but also in risk assessment in general, and he's come up with a few different ways of categorizing different types of risk assessment efforts.

Kevin Baldwin: The first unguided or unstructured clinical judgment. This is something that was in wide use for the majority of the time that we've provided risk assessments in criminal justice settings, and it really depends on the degree to which we trust experts to know what they're doing. Unguided or unstructured clinical judgment involves having someone like a clinical psychiatrist or a clinical psychologist meet with somebody, spend time with them, and then come up with an assessment of risk. It really has no rooting in the research. It's based on someone's judgment and someone's expertise.

Kevin Baldwin: The next stage is guided or structured clinical judgment. That's a step up in that the professional who's doing the risk assessing uses some structure or something to guide that process. Now, they may look at a list of factors. Those factors, however, are not weighted. They're not considered in any appropriate or any consistent manner.

Kevin Baldwin: The next step is research-guided clinical judgment. That's where, essentially, you have a well-quantified list of things that the research has indicated are reliably associated with recidivism, and you use that, and there's really not a whole lot of head scratching and thinking, "This guy's got beady eyes. He's risky." It's based on research.

Kevin Baldwin: The next approach is the pure actuarial approach. Many of you are likely familiar with actuarial risk assessment. We're going to talk a lot about that today. But actuarial approaches are ones in which we use a particular set of factors, and those factors are combined and weighted the same way every time which results in a very consistent approach to risk assessment.

Kevin Baldwin: Finally, Hanson describes the adjusted actuarial approach. That's where you use an actuarial measure, and then you adjust it based on other characteristics or other factors that were not included in the factors that comprised the actuarial measure you're using.

Kevin Baldwin: There's another way to describe this, and that's a generational means of describing these assessments. That was put out by Bonta. Some of you may be well aware of the risk needs responsivity model of Bonta and Andrews. Bonta
and Andrews, they're colleagues, very well known, well respected criminologists from Canada. Bonta, in 1996, came out with a generational model.

Kevin Baldwin: You look at the first generation, unstructured professional opinion, that corresponds to Hanson's first category, basically unguided or unstructured clinical opinion. The second generation is actuarial methods using static predictors. We're going to define that term "static" in a little bit, but, basically, those are predictors or factors that really don't change, and, if they do change, they only go up.

Kevin Baldwin: The third generation are methods that incorporate both static, or relatively unchangeable factors, along with dynamic predictors. Dynamic predictors are those things that can change. Because they can change, they are often the focus of interventions or intervention targets, things such as substance abuse, mental health issues, family dynamics. Those are things that can be addressed in treatment.

Kevin Baldwin: The fourth generation involves assessing static and dynamic predictors but then using that information to come up with a case management plan or a treatment plan. What the fourth generation does that the third generation doesn't is that it uses the information provided about static and dynamic risk predictors to actually come up with a plan to address them.

Kevin Baldwin: Now, as we mentioned earlier, not much you can do about static risk factors. For instance, you can't really change the age at first arrest. What you can change though, if someone has a demonstrated substance use pattern, pretty serious substance use pattern, you can apply evidence-based substance abuse treatment and hopefully ameliorate or reduce that risk. That's a couple of ways of understanding generational and in terms of different methods how we've come to where we are today where we have a proliferation of different methods of risk assessment.

Kevin Baldwin: As you might expect, there's quite a bit of controversy. There's been some pretty contentious arguments about which of these measures is best. Pretty much, it's been decided that unstructured approaches don't really hold a lot of water. You may be familiar with some of the studies that took noted forensic psychiatrists and psychologists, paired them up against, say, schoolteachers, plumbers, other folks who really have no specific expertise, and they found out that there was really no difference in their predictive ability when it came to purely clinical prediction of risk. It's been pretty settled that structured approaches do better than unstructured approaches when we look at them in comparison one to another. There are some citations there going back to the 50s, actually. But each of the structured approaches have their merits and both supporters and detractors.
Kevin Baldwin: Let’s now talk about the types of risk factors. I mentioned these a minute ago, but it’s important, I think, to take a little bit closer look because these are things that comprise the measures that we use today in assessing risk.

Kevin Baldwin: The first are static risk factors. Those really don’t change a lot. That’s why they’re static. They don’t really move, and, if they do move, they increase. One of the static predictors that increases but won’t decrease absent intervention from a governor is the number of convictions. You may have your offense commuted or you may have a pardon issued, but, other than that, that’s a number that’s only going to go up. That’s a static risk factor. They also generally are not amenable to intervention efforts. Not much you can do to change those things. Probably the best example is age of first offense. Without anything other than the ability to travel through time, which I don’t think any of us have mastered yet, you’re not going to be able to change that.

Kevin Baldwin: On the other hand, we can look at dynamic risk factors. Those are also known or referred to variously as criminogenic needs or psychologically meaningful or relevant factors. These are things that can change. They’re dynamic. They do respond, often times, to intervention, and there’s two characteristics or two types of dynamic risk factors: stable and acute. Stable risk factors, dynamic risk factors, are those things that they do change, but they change a little more slowly. Acute dynamic risk factors, those are things like current psychiatric condition or current mood state. Someone might be intensely angry, or someone may be in the midst of a psychotic episode. Those are acute. Stable are things like substance abuse patterns, things like that.

Kevin Baldwin: Those are important things to keep in mind so that, when we talk about risk factors, we’ve got ones that don’t really change, those are static, and we’ve got ones that do change, and therefore we can often apply an intervention in an attempt to change them.

Kevin Baldwin: Let’s now turn to a discussion of accuracy because, ultimately, the tools we use to assess risk should be as accurate as possible. In terms of accuracy, there’s two really important factors that determine how accurate a risk factor or a risk assessment is.

Kevin Baldwin: First, it's the degree to which the person we're assessing actually matches a known group of offenders. The second is the degree to which the factors we include in a risk assessment are an accurate representation or reflection of the known universe of relevant risk factors. The first one is often the first thing we've got to do. We've got to figure out what group this individual matches, and we'll talk a little bit more about that later. That's another controversial topic, but, essentially, if we've got somebody before us who is so different than any of the groups for whom we have validated risk assessments, it's going to be really hard to do a risk assessment because we can't find an appropriate match.
Kevin Baldwin: The second of these factors, we've got to make sure that those things we're including in a risk assessment are those things that are reliably related to risk. In other words, they reflect the known group of relevant risk factors.

Kevin Baldwin: A couple of other things. When we're talking about actuarial risk assessments, those are ones that are basically mathematical formulas. They use the same factors. They use the same weights. They're combined in the same way every time. The utility of a factor, a specific factor, really depends on its empirical relationship to whatever it is that we're predicting. In other words, there's a mathematical relationship. The stronger the predictor, the greater the weight.

Kevin Baldwin: In order to find that out though, in order to assess that, in order, in fact, to develop an actuarial risk assessment, we've got to know the base rate of the behavior that we're trying to predict. The base rate is the rate at which that behavior occurs in the population of interest. The only way we understand base rates and come to an understanding of what they are is through large meta-analyses of offender recidivism. Those are things that Carl Hanson, someone I mentioned earlier, has really made a name conducting: very large sex offender recidivism studies where he's had populations exceeding 30,000, which is a real big study.

Kevin Baldwin: What we've found though, in all of the years we've been doing this, is that there's no single best predictor of risk. There's no single best factor, and, in fact, there's no single best measure either. Therefore, we're left with a bit of a conundrum. Maybe, someday, there will be. Maybe, someday, we'll find that there's this universal measure of risk assessment, but I doubt it because it really depends on the specific outcome we're looking at. All right.

Kevin Baldwin: Let's talk about some of the common factors that have been identified and their literature because there are some. Earlier, when we were talking about the history of risk assessment, the generational model, we mentioned the work of Andrews and Bonta. They've done a great deal of research, and, in fact, they've identified what are commonly referred to now as the central eight risk factors. These are things that, when Andrews and Bonta looked at the literature, the criminological literature, hundreds and hundreds and hundreds of studies, they came up with this list that pretty much across studies can be found to be reliably related to recidivism.

Kevin Baldwin: Here's that list. It's known as the central eight. As you look at that list, I don't think anything's going to come as a big surprise. Those things are things we're pretty much well aware of in our work as factors that we know are predictive of future criminal behavior. There's the central eight, and then you'll see the first four are referred to as the big four because those are ones that are identified time after time after time in study after study after study as being reliably associated with risk for re-offense.
Kevin Baldwin: You'll see that the first four of those all have the word "antisocial" in them, which is no coincidence. That's the big four. The others- Number five: substance abuse. That's obviously a risk factor. Six, seven, and eight: those are essentially deficits. The first four are things you have. The one, number five, is something you do. Six, seven, and eight are things you lack. That's a list that's very common. You've likely come across that before, and you will again.

Kevin Baldwin: Now that we know some of the factors and obviously not all of them, how do we consider what factors to look at? How do we consider what methods to use in terms of risk assessment? If we're going to do this accurately, if we're going to do it well, we've got to make sure that we know about as many as possible relevant factors that are associated with our outcome of interest.

Kevin Baldwin: For instance, if we're assessing pre-trial risk- That is the risk of someone who is at pre-trial. They've been arrested. They're being detained, typically in a jail. We want to know: Is this person a safe bet to let out during their pre-trial period? The outcome of interest, basically, there's two: Is this person going to fail to appear? Or is this person going to do something that results in a new arrest during the period of pre-trial confinement? Of course, they're not confined. They're out in the open there. We've got to know what our outcome is, and then we've got to know what things predict that particular outcome.

Kevin Baldwin: Another thing to remember- We talked briefly about static and dynamic factors. A lot of the risk assessments, in particular the actuarial assessments that we use, are comprised for the most part of rather static factors. A lot of times, there are dynamic factors that we want to include as well. Well, that's a great thing to do, but one of the ways not to do that typically is by adjusting actuarial instruments after the fact.

Kevin Baldwin: Studies that have actually looked at that had found that making clinical adjustments to an actuarial measure, sometimes referred to as an override, using factors or considering factors outside the actuarial scheme, it can actually hurt the predictive accuracy of the measure. Now, all of the studies that I've been able to find on this particular issue, and that is the adjustment of actuarial measures, all have to do with sexual offenders. There's not a great deal out there on that particular method of adjusting actuarial scores on non-sex offender populations, in other words on a more general population, but there's no reason to believe that it wouldn't work that way with a more general group of offenders.

Erin Farley: Excuse me. Kevin?

Kevin Baldwin: Yep.

Erin Farley: I have a quick question from a attendee. Do these factors apply to all kinds of criminal activity, specifically if you are assessing for domestic violence recidivism, the same for those committing stranger violent crime?
Kevin Baldwin: Yeah. Excellent question. These are very generally associated with criminal recidivism. There actually are specific measures, there’s one in particular, used to assess for domestic violence that’s gained popular use. I’ll look it up during the course at the end, but that’s a great question because, generally speaking, and I put back the slide on the central eight and the big four, generally speaking, these factors are associated with criminality. But, when we talk about predicting really specific outcomes like domestic violence, there are more special, more specified measures that have been developed and then validated on that particular population. Does that answer the question?

Erin Farley: Great. Yes. I think so. Thank you.


Kevin Baldwin: The next thing that we’ve got to ask is: How many measures should I use? There’s a [panabole 00:31:36] of measures out there. There are measures of general recidivism. There are specific measures that can be used for, as I just noted, domestic violent offenders, for sex offenders, for all sorts of different things. What do we do? How many should I use? Should I just use one? Should I use a general one? Should I supplement that? Well, there’s a couple of things to consider.

Kevin Baldwin: I hesitate to bring this up because sometimes it gives people hives, but think about your intro to stat class or if you’ve taken a measures class. You might remember classical test theory. Some of you may be very enamored of classical test theory, but, among other things, classical test theory suggests that, the more items you have in an assessment, the greater its reliability and coverage is going to be. In other words, if you want to increase the reliability of your measure, just add items. If you want to increase the coverage, according to test theory, classical test theory, you add more items that tap different things. Classical test theory, which suggests the more the merrier, the more the items the more measures the better you’re going to be.

Kevin Baldwin: There’s another thing though that suggests that perhaps using more than one measure is a good idea, and that’s the idea that criminal behavior is motivated by different factors. People commit crimes for a number of different reasons, and, if the particular type of offense you’re interested in- For instance, in the field of sex offender risk assessment, we know that there are two or three factors in particular that predict the degree to which someone is likely to engage in future sex offending behavior. One is sexual deviance. We know that’s a pretty robust predictor, so things can cap sexual deviance. Another though is general criminality. It’s not enough often times just to have sexually deviant motives. You also, in order to act on those motives, generally have to score in the moderate to high range on a measure of general criminality. There’s one third thing we look at, and that is intimacy deficits. Those are driving forces that we have identified in the literature as being behind sex offender risk for recidivism.
Kevin Baldwin: In terms of deciding the measures that you should use, think about classical test theory, reliability, coverage. Think also about, theoretically, what forces are behind the offending behavior that you're interested in predicting? That'll help you decide essentially whether or not you can stick with one measure or whether you need more than one.

Kevin Baldwin: Let's turn now to communication of risk assessment results. This is something where, frankly, we've not done so well. As far as communicating the results of risk assessments, we've got a lot of work to do, quite frankly, and I'm speaking both as a researcher and as a practitioner, as someone who actually does risk assessments as part of my work on a regular basis. We've tried all sorts of different ways, frankly, and none of them in isolation work all that well.

Kevin Baldwin: One of the things we'll talk about are prescriptive categories, so, in other words, just using words like saying someone is low risk or saying someone is moderate risk or saying someone is high risk. Well, there's a few problems with that. First, your moderate may be equal to someone else's high. Your low may be equal to someone else's moderate. One of the problems is that, when we hear those terms, we often have different frames of reference. We have different ideas as to what that means.

Kevin Baldwin: For instance, when we say "low risk," sometimes, people use the frame of reference of members of the general public. Well, when we're talking about an offender being low risk, we're not comparing him to the general public. We're comparing him to his peers, other offenders. This comes up often time when I am called to testify in court about someone's dangerousness. They'll ask me, "Well, how did he score? And what category does it fall in?" I'll say, "Well, he falls into the low risk category." I can't stop there though. I have to say, "relative to a similar sample of known offenders," because the general proclivity, the natural proclivity, is for someone to use their own frame of reference. When I say that an offender is low risk, he's low risk compared to his peers. He's not low risk compared to members of the general population. That's one problem.

Kevin Baldwin: The other problem is that, often times, from one measure to the next, those words have different meanings. For instance, one measure, low risk may mean, "Oh. This person only has a 15 percent chance of recidivism over the next five years." But, on another measure, low risk may mean he's got a 25 percent chance. They have different probabilities associated with those words. In terms of descriptively communicating risk assessment results, it's really tricky because the people that are consuming our information have different frames of reference, and the measures themselves have different ideas about what it means to be low or high.

Kevin Baldwin: A really interesting study was done by Carl Hanson where a number of researchers compared risk level definitions among five different measures. Sadly, they found that only three percent of the people that they assessed were identified as high risk across all five instruments and only four percent were
identified as low risk by all five measures. The same person can be described by
different categories across different instruments or an individual in the same
category can be described differently across different assessment instruments. I
don't think any of us would say that's a good position to be in. In fact, it's
embarrassing. We've been doing this for a while now, and that's hopefully not
the best we can do.

Kevin Baldwin: Instead of using words, we might consider using numbers. The problem is that
we might be a pretty sophisticated bunch numerically speaking. We may be high
in numeracy or an understanding of statistics, but that's not generally shared by
a whole lot of the consumers of the risk information that we're trying to convey.
For instance, most of the folks in my experience that I'm communicating risk
information to have no idea what an odds ratio is, and it's rather hard to
explain. More people know what a percentile is, but, percentile, that is so close
to percentage that people get those two things confused.

Kevin Baldwin: When we're talking about communicating risk, it's important to know that, first,
when we use descriptive categories, there's different words that mean different
things to different people on different assessments. Well, if we just use
numbers, our numerical estimates, our numerical communication methods
likely are pretty puzzling to the consumers. Frankly, when it comes to risk
assessment, if we can't communicate the results accurately and in a way that is
understandable to those people that need to use the information, why are we
even doing it? It's not going to do what it needs to do. What can we do?

Kevin Baldwin: In fact, Carl Hanson, the person I've referenced a number of times, he along
with the Council of State Governments, CSG- Likely many of you are familiar
with CSG. In 2017, they published essentially a paper that recommends we
together use a common language to communicate risk, and the Bureau of
Justice Assistance has come and supported the conclusions that Hanson and his
colleagues at CSG came out with. What they recommended is that we use a five
level scheme of communicating risk. In other words, with five levels, each of
those levels has a particular descriptive category from very low on up. Each of
those levels is associated with a specific set of characteristics and a specified
range of risk for re-offending. It's an attempt to standardize communication of
risk assessment.

Kevin Baldwin: I think, frankly, this is a very welcome move, a very positive move for those
reasons cited earlier. When you can take five different measures and apply
them to a population and only three percent were categorized as low by all five
and only four percent were categorized as high by all five, we've got an issue.
This is an effort. It's new, obviously. It's less than a year old, but I hope that it
gains some traction, and I hope that it really facilitates communication but, even
more so, understanding of risk assessment because, as I said earlier, a risk
assessment, if we can't communicate it properly and if the consumers of risk
information can't understand it properly, they really can't act on it. All right.
Now, we get to the fun part here where we’re going to talk about current controversies in risk assessment. Earlier, I put up a couple of articles in a book, and that book, in fact, is a New York Times bestseller, and it was nominated for Non-Fiction Book of the Year. We’re talking about pretty widely read stuff. Some of you may have seen stories in the news, articles in the popular press. Magazines like Newsweek have published stories about risk assessment. It’s been in the news a lot and not because of really great things, not for the best reasons. Often times, these articles and books are pointing out specific issues, concerns, or controversies, and I’ve read a lot of that stuff, certainly not all of it, but, the reading that I have done, I’ve grouped these into four general areas of concern.

The first is the development of risk assessments. The second is the use of risk assessments: how we use them, what we do with them. The third, which is really, really important, not that the other two aren’t, but constitutional issues. Risk assessment gets to, in many cases, constitutional issues around our freedoms. The fourth: algorithms and big data and the role that they play in risk assessments. We’re going to go through each of those in turn, and we’re going to start at the top with the development of risk assessments.

In order for a risk assessment to be in use, it, of course, has to be developed. How do we do so? How do we do so responsibly? One of the first things that we’ve got to think about and we’ve got to examine is the data that we use to actually create a risk assessment, and a key question in terms of the controversies that have cropped up lately: Do those data exacerbate disparities? I would, frankly, be quite surprised if anybody on this webinar wasn’t aware of the controversies surrounding bias in our criminal justice system, disparities in terms of the application of justice. We need to make sure that the data that we’re using to develop risk assessments don’t make those things worse. They exist. The last thing we want to do is make them worse. In fact, we’ll talk a little bit more about this later, it’s possible that, through our development of risk assessments, we could actually improve the process, and we could actually reduce disparities. That’s one thing to keep in mind. To do the data, we use exacerbate disparities.

The second thing that we’ve got to think about are the factors that we choose. Do those factors impinge on constitutional issues. Are we introducing things, demographic and other variables, that would be prohibited otherwise were they not part of a risk assessment?

Third, do our risk assessments actually represent a black box? Are they opaque? Are they open to examination? Are they open to being tested by other people? Or, for trade reasons or otherwise, are they a black box? Nobody knows what happens. Numbers go in, and a decision comes out the other side. What are the implications if they are that way?
Kevin Baldwin: Use of risk assessments. This is the second of the three areas. First, development. Now, use. One of the key questions around risk assessments is when we're using them or how we're using them. In other words, at what point in the adjudication process are we conducting the risk assessment and for what purposes? One of the most popular forms and, frankly, one of the least controversial and relatively straightforward risk assessment venues is in pre-trial risk assessment. There's two things we're really interested in: likelihood of failure to appear and the risk to the public which is operationally defined as a new arrest while on pre-trial release. Pretty simple ...

Erin Farley: Kevin?

Kevin Baldwin: Yes.

Erin Farley: Hi. Sorry. I don't mean to interrupt again.


Erin Farley: Great. I think that this question just popped up, and I think it might be related to the last slide. It's from, again, one of the attendees. Offense history is a known predictor of future criminal activity, yet it is also a disparate factor. How do you resolve this conundrum?

Kevin Baldwin: Oh, man. That's an excellent question. If I could actually give a complete answer to that, I think they would have already awarded me the Nobel Prize. But here's one of the ways you can deal with that, and we'll talk a little bit about this in a subsequent slide, but essentially there's different ways of measuring prior criminal behavior. One of the things you can use is number of arrests. Another you can use is number of convictions. A third is the number of times you've been to prison. Well, we all know that those represent different ends of the spectrum of the adjudication process. We also know, for instance - And, in the book Weapons of Math Destruction, the author points out the difference between reported crimes and found crimes.

Kevin Baldwin: Well, one of the things we are aware of is that people gave greater arrests or more arrests often times, and those arrests may be based on where they live and maybe based on the color of their skin, things like that. When we're talking about using a predictor of future criminal behavior and deciding how to incorporate their past criminal history, a lot of folks would argue that arrests is not a really good measure because there's, built into that, some biases. We would then move up into the adjudication spectrum a little bit further and look at convictions, and that will help address, but not completely, that issue of differential arrest and policing policies. Often times, and, in fact, in measures that we use, we try and shy away from arrests, and instead we look at convictions.
Kevin Baldwin: That's one way to do it. There are others, but it's really important to know, for instance, if you're considering using arrest as a measure, it's really important to know the population that you're predicting in. If you're predicting in a very heterogeneous population, then it may be that, when you use arrest as a measure of prior criminality, it's going to over-predict among African-Americans and under-predict among whites because African-Americans often have a higher rate of arrest because of policing, neighborhood issues, all sorts of other things.

Kevin Baldwin: Very complicated, but a very great question. Unfortunately, there's no real clear answer, but those are some guidelines, and we'll discuss those a little bit more later.

Erin Farley: Thank you.

Kevin Baldwin: Do you think that gets to it, Erin?

Erin Farley: Absolutely.


Kevin Baldwin: Pre-trial, pretty straightforward. Sentencing, not so much. In fact, when we talk about using risk assessments in sentencing, we're getting into some pretty controversial issues and some pretty complicated questions. One of the reasons is because, at pre-trial, the thing we're predicting is pretty straightforward.

Kevin Baldwin: Well, when we're talking about sentencing, we're talking about two crucial decisions. What type of sentence? Are we going to sentence this person to probation? Community release? Are we going to sentence them to prison? Are they going to get a combination of both? Are they going to be sentenced to death? Those are really ways. The second though, in addition to the type of sentence, is the length of a sentence. How long is this person going to serve? Those are informed by things outside of risk or other than risk, things that need to be considered such as individual retribution, potential for rehabilitation, the deterrence effect of the sentence, and the degree which the person needs to be incapacitated. Those things really go beyond just risk information. They have to deal with someone's theory of punishment.

Kevin Baldwin: When a judge or when a jury is weighing those questions, when we're talking about something at a pre-trial, it's pretty straightforward. But, when we get into sentencing, it's really fraught with potential complications because we're talking about things that are considered at sentencing that go beyond what could just be addressed via an understanding of what risk factors are.

Kevin Baldwin: Some of you may have heard of evidence-based sentencing or the Model Penal Code. The Model Penal Code has been around for a long time. It was last revised I think in 1981. Now, I'm not an attorney, so maybe some attorneys out there know the answer to this, but, essentially, there are some recommendations to
incorporate in the Model Penal Code or the next version of the Model Penal Code evidence-based sentencing. Evidence-based sentencing often relies on risk assessment.

Kevin Baldwin: As of last year, five states actually require the use, they don’t just recommend but they require the use, of risk assessment in sentencing. They do it in different ways, but those states are there. Ohio and Pennsylvania together seem to have taken a comparatively thorough approach. What they’ve done is they’ve consulted and received guidance from experts in academics in how exactly to use risk assessment in the sentencing process. The other states, they’ve left it up to the judiciary to sort out the issues. This is something that is a growing edge. It’s out there on the leading edge of application of risk assessment.

Kevin Baldwin: Some of you may know that Eric Holder, the previous attorney general, strongly urged that we take a very, very cautious approach to the use of algorithms and risk assessments for sentencing purposes. There’s a quote there, and he specifically notes the potential, when we’re using risk assessments in sentencing, to have a disparate and adverse effect. There are some other quotes from him that get to that, basically talking about how a defendant’s demographic characteristics, their features such as where they’re from, their socioeconomic background, when incorporated into risk assessment can not just perpetuate but make worse the unjust disparities that already categorize so much of our system. Eric Holder and the justice department under the previous administration really cautioned against the use of risk assessments in sentencing.

Kevin Baldwin: Let’s turn to the third area, the third controversial area, and that is constitutional issues. Again, I’m not an attorney, so apologies to anyone who is out there and for my butchering of the fourteenth amendment, but essentially there’s two issues that come up. The first is right to due process. The way that relates to risk assessment focuses on two things, basically.

Kevin Baldwin: When we apply a risk assessment, what we’re doing is we’re looking at an individual. We’re comparing him or her to a group, and then we’re taking the observed recidivism of that group and applying it to an individual. Many would argue that that violates due process because we’re using group norms to come to a decision about an individual’s propensity to engage in future criminal behavior. Another thing about right to due process that some risk assessments violate are those risk assessments that, for trade purposes, are a black box. We don’t know what goes into them. An example of this is the COMPAS, and that was a big knock that the COMPAS took as a result of the ProPublica investigation, that there was no way you could go in and see exactly what those factors were because to do so would reveal trade secrets. That’s another way that risk assessment is implicated in due process.

Kevin Baldwin: Now, the other fourteenth amendment issue is the equal protection clause, and probably the most vocal critic that addresses risk assessment from an equal
protection clause is Sonia Starr. She's a law professor at University of Michigan. Basically, what she argues is this: In that, by using risk assessments, and in particular actuarial risk assessments, in sentencing, what we're doing is we're allowing factors to be weighed that wouldn't otherwise be allowed in terms of sentencing, factors such as age, gender, socioeconomic status. Those things can't be considered when we're sentencing, but, when we use a risk assessment that might include those factors, it's allowing them in, and she uses some pretty strong language as you can see in that quote, basically otherwise condemned discrimination which has been sanitized by scientific language. I really recommend you look up some of her publications that are readily available on the internet and they get to that issue. That's the third issue is constitutional issues.

Kevin Baldwin: A couple more slides there or another slide about some specifics around these constitutional issues, but let's talk about algorithms and big data because this is something that gets a lot of press lately. It seems, a few years ago, we were really enamored of algorithms and enthusiastic about the promise of big data. I read a book as did one of my colleagues called Algorithms for Life. It basically taught you how you could use different algorithms to improve your state of life. Big data, we also read a few books here at where I work about, oh, how big data is going to revolutionize things. It's going to make things better and easier for us, make us smarter.

Kevin Baldwin: Well, since then, we've learned a lot, and we've got a lot of concerns now about the degree to which these algorithms control our lives. In the news right now are stories about Russian interference in the elections and big data and how Facebook and other social media platforms were manipulated to essentially mislead people or get them all hot and bothered about certain things. Some of you may remember the movie 2001: A Space Odyssey. For those younger members of our audience, it did not come out in 2001. I think it came out in 1968 or '69. Another movie called iRobot. I think, often times, we have a fear of artificial intelligence that we've conflated with a fear of big data and algorithms, and we think, "Goodness. Robots are going to take over, and we're all going to die." Whereas, once, we were enamored of these things, now we're skeptical and I think with good reason.

Kevin Baldwin: Edward Snowden and Wikileaks really opened our eyes to the amount of information that our government keeps on us. When we learned all of the different information, how it's kept, how it's recorded, also we've learned a lot about Google, Amazon, and Facebook, their intrusion into more and more areas of our lives, how much they know about us. A couple of books on big data tell some real interesting cautionary tales about just how much is known about us out there. At the same time, we seem to readily accept the role played by algorithms in other areas of our lives, in particular the insurance industry, our insurance rates, and in healthcare, our healthcare. Both of these fields, more and more of the decisions around those two areas are made through algorithms informed by big data.
Kevin Baldwin: These concerns, though, are valid, and one of the issues concerns the lack of transparency. We mentioned this a couple of times. When we're talking about risk assessments, when a risk assessment uses information in a way that is opaque, when there's no transparency, the question is: Well, how do I confront my accuser? How do I confront a black box, a formula which never will be revealed to me? How do I know what factors it even considered? How do I know if it's fair? That lack of transparency, referred to as opacity, is a really big concern when it comes to algorithms and big data, when we don't know what factors are included, how they're combined, or the weighting that's given to them, and, when we can't find that out, it really does have some implications in terms of our ability to confront our accuser.

Kevin Baldwin: Another issue with algorithms and big data: bias and the lack of reliability or validity. Going back to the ProPublica study of COMPAS, essentially the ProPublica article argued that the COMPAS was inaccurate and that it didn't do a good job of assessing risk and that it was also racially biased even though it didn't explicitly include race in its algorithm. Now, the publisher of COMPAS, formerly known as Northpointe, they've changed their name to Equivant- I wonder why. I guess they got a lot of bad press. I don't know. But Northpointe has vigorously disputed the findings described the ProPublica study, and they continue to. An article was published just a couple of weeks ago that again looked at that issue, the issue of fairness. They're making some good points.

Kevin Baldwin: It's a very controversial issue, and it's not one that's easily answerable. But we certainly, when we're talking about risk assessments, we want our risk assessments to be reliable. We want them to valid. We don't want them to perpetuate bias.

Kevin Baldwin: Another thing is that, when we use these risk assessments, we've talked about this a little bit before, it's not enough to just have things be the status quo. Some folks, in fact a growing number of folks, think that, as we apply these risk assessments, we may actually be making things worse because what we're doing is we're using things that are already baked into the pie, making decisions on that. It's somewhat tautological in their argument that, for instance, if we use number of arrests to predict future likelihood of getting into trouble and we know, for instance, that there is a bias in that data, if that's already baked into the cake, it's just going to make things worse. It's going to exacerbate existing racial and ethnic disparities, and, frankly, that's the last thing we need to do.

Kevin Baldwin: I've mentioned already Weapons of Math Destruction, Cathy O'Neil. This is a really great finding, a great point that she makes, the difference between reported and found crimes and how those reflect and are reflected in policing patterns and then risk. For instance, found crimes are more likely to occur in areas where police observation is more frequent, such as poor and minority neighborhoods. That will then increase the likelihood of arrest and therefore contribute to or even exacerbate bias in the system.
Another thing we've got to think about, and this harkens back again to the ProPublica article, are the types of errors that risk assessments produce. No risk assessment is perfect. They're going to be accurate hopefully more often than not. If that's not the case, don't use it. But different types of errors are going to be found. For instance, when you're talking about risk assessment, you can over-predict or you can under-prediction. Over-prediction is saying basically that this person is going to recidivate, and they don't. Under-predicting is saying they're not going to recidivate and that they do.

We can't, as developers of these risk assessments, decide the ratio that will be acceptable of those two different types of errors without lots of consultation with policy and decision makers because those decisions are context-dependent. We've got to figure out and work closely with our colleagues who are policymakers and decision makers in terms of what errors they are likely and willing to accept, and we've got to continually monitor them to make sure they're free of bias, and that was one of the main arguments that was made in the ProPublica study and that the COMPAS over-predicted risk among African-Americans and under-predicted risk among white defendants.

That gets to the concept of fairness. Again, I'm not a lawyer, but, when we talk about the legal system, how the legal system defines fairness, for instance, is very different than how my teenage kids define fairness. All right. Fairness is equality in the process or parity in the process, not the outcome itself. If I tell my kids, "Hey. Yeah. The outcome isn't fair, but just know, be confident, that I arrived at the decision the same way with you and your brother." That doesn't fly. But, from a legal standpoint, it's not the outcome; it's the process. Legal fairness emphasizes and prioritizes the equality of the process. You've got to use the same process. That process has to be characterized by equality, not necessarily the outcome.

Those are controversies, big controversies. Many of you are likely familiar with them. I'm hoping that you'll go and you'll read up on some of those things that we've mentioned.

What do we do though? Can we go back? Can we say, "Hey. This is all too complicated. This issues are just too weighty. Let's just go back to having an expert mental health clinician sit down with a person then tell us how risky he is maybe based on his skull features like phrenology or whether he has beady eyes or whether he seems shifty. Maybe we need to do that." I don't think so.

The research is clear we can't do that. If we're going to do a risk assessment in a way that is valid and reliable and meaningful, we've got to move beyond clinical risk assessment. We've got to use actuarial risk assessments where possible, but we've got to do so in a way that is responsible. I think, the way things are going, more and more, we're going to move away from paper and pencil measures, and more and more risk assessments are going to rely on big data. They're going to combine them in an automated, actuarial way.
Kevin Baldwin: How do we then practice going forward? Well, if we have objective, research-based tools, if we're using risk assessments responsibly, we might be able to actually help reduce the bias that historically has been part of our criminal justice system. We know so much more now than we did 50 years ago about what causes crime, about the factors that lead to crime.

Kevin Baldwin: We can't pretend bias doesn't exist. It does. It's been a part of our system for decades. That's before the advent of actuarial tools, so that bias exists. The last thing we want to do though is institutionalize that bias through furthering it with risk assessments.

Kevin Baldwin: Given all this controversy, given that this is a really fraught issue, that it's not easy, I'm going to humbly offer some suggested guidelines that we can use, that we can forward keeping and adhering to these guidelines as much as we possibly can so that we can use these tools, we can develop these tools, we can communicate about these tools, make decisions about these tools in a way that is respectful and in a way that doesn't make things worse. Here are some guidelines to consider.

Kevin Baldwin: First, it gets to the transparency issue. Don't use opaque tools. Be open and transparent regarding the factors that you use, how those factors are chosen, and how they're weighted. You can't just apply the black box methodology and tell people, "Well, that's just the math, and you can't see it because it's private." We've got to be able to not only publish and put out there what those algorithms are but we've got to get outside folks to look at them. We've got to have our work examined by other people, and we've got to be open to looking at other factors and other ways based on the input that we receive.

Kevin Baldwin: Another thing we've got to do, our second guideline, is to ensure that risk assessments are being used correctly for their intended purposes. For instance ... Oh. Yeah. Go ahead. Is there a question? Nope? Okay. We want to make sure that we're using them for the appropriate population. We can't use a measure that was normed in Hawaii and just take it and put it in Iowa and say that it's valid. We've got to make sure that it's normed on local populations.

Kevin Baldwin: Another thing goes back to what we talked about just a few minutes ago, and that is that we've got to address different error types, and we've got to reduce as much as possible the error and increase the degree to which we're making good decisions. We've also got to make sure that those error rates don't differ based on demographic factors such as race, ethnicity, income, those sorts of things.

Kevin Baldwin: The way we do that gets to guideline number 4, and that is to continually test these risk assessment tools to make sure they're performing equally well across the different groups to which we're applying them. If we find that they're not, we're going to have to revise the measure. We might even have to use two
different measures, but they've got to demonstrate equal levels of reliability and validity across the subgroups to whom we’re applying them.

Kevin Baldwin: Number five: We've got to revalidate the measures. Conditions change. People change. Laws change. At a minimum, I would recommend, as a guideline, that you revalidate your measures at least once every five years.

Kevin Baldwin: You've got to also stay current on training and developments. If this stuff is hitting the popular press, if there are articles about this in Newsweek, if there are, as a result of the attention being paid to this, lots of developments in the science of risk assessment, it really behooves us to be as current as possible on training so that we’re aware of those developments in the field.

Kevin Baldwin: Also, number seven, understand that gender and class are linked with so many different issues and so many other issues. You can’t just separate them. They impact so many things, and it’s critical to keep those in mind when we’re using and developing and communicating about a risk assessment.

Kevin Baldwin: Another guideline comes from Kehl. It's a great article. It's called "Technological Due Process." It goes back to the issues of big data and algorithms and core values of transparency, accuracy, accountability, participation, and fairness. We’ve talked about each of those in turn. They also [know posit 01:09:49] that it's necessary to have an audit trail, meaning that you can go back and look at the different steps to ensure these things existed in the risk assessment process that you have applied.

Kevin Baldwin: Policymakers need to be involved. They just can’t stand back and say, "Okay. When you've finished, when you've got a risk assessment, just let me know." They've got to be involved and aware of all steps in the process.

Kevin Baldwin: We’ve also got to demonstrate, number 10, a very strong commitment to fairness, and that's fairness in the legal sense, that the outcomes aren't always going to be equivalent. If they were, we wouldn't need a risk assessment but that the process has to demonstrate fairness.

Kevin Baldwin: We also, and this gets back to some of the things that Cathy O'Neil argues in her book, is that, if there are certain factors that are very troubling, we might want to exclude them, and the one I mentioned earlier is number of arrests. Number of arrests may be very empirically-related to our outcome of interest, but it may also bake more bias into the cake, and therefore we might want to consider something else like number of convictions or number of times in prison, something like that. We also want to ensure that there are strong procedural safeguards in place to make sure that the scores are used properly and their adverse impact and inadvertent impact is minimized as much as possible.

Kevin Baldwin: Communication. We spent some time on time. Risk assessment results are no good if they're not communicated in a way that the consumer of that
information can use that as actionable information. The results of risk assessment should not be communicated in a way that is difficult to understand or ambiguous. The work of Carl Hanson and the Council for State Governments to reduce as much as possible those factors and to standardize the language around how we communicate risk assessment I think is really laudable, and I am pleased to see them doing that.

Kevin Baldwin: The future of risk assessment. The genie is out of the bottle. We have risk assessments. They are proliferating. We are more and more relying on big data and algorithms. Because of that, we've got to make sure that, as we develop these, as we use them, as we make policy around them, we've got to do whatever we can to make sure they're used ethically and that they do not violate constitutional standards. I think it really behooves us as a field to commit to those values of transparency, accountability, and fairness at every step along the way.

Kevin Baldwin: It's not going to be perfect though. The only way we'll ever arrive at a perfect risk assessment is if we're somehow able to collect all of the information on all of the relevant factors and know exactly how to combine those factors mathematically in a way that provides 100 percent perfect level of prediction. I don't know about you. I don't think that ever is going to happen. I'm an optimistic guy, but I don't think that's going to happen. As a result, we've got to accept there's going to be a certain amount of error in our processes.

Kevin Baldwin: Now, we accept error in many other things. Good example: open heart surgery. Open heart surgery is actually fatal. People die on the operating table in about one to three percent of the cases. I think we could all agree that's not an optimal outcome, certainly not the intended outcome. But just because it's not perfect doesn't mean we throw it out and we don't do open heart surgery. Likewise, when we talk about risk assessment, if we're overselling risk assessment by saying it's more accurate than it is, we're doing ourselves more harm than good. It's not perfect. It's not likely to ever be perfect, so we've got to help people understand that just because it's not perfect doesn't mean it's not good and it's not better than the alternative, which is clinical acumen or relying on judgment.

Kevin Baldwin: In the end, really, we should be committed to using the best tools we have, choosing them appropriately, applying them appropriately in a reliable manner that doesn't go beyond the data and certainly doesn't exacerbate disparities that we know already exist. Finally, we've got to communicate them effectively using as standardized a means as possible and always commit to improving our tools and our methods.

Kevin Baldwin: That's it. Any questions?

Erin Farley: Kevin, thank you so much. A quick question on my part. There were three key-at least I caught three key references that you made in terms of other types of
reading. One was the Cathy O'Neil book and then the Kehl et al. 2017 article, and I actually found that online.

Kevin Baldwin: Oh, great. Thank you.

Erin Farley: I also found the- Let me pull that up. Oh. I found the Hanson et al. that just came out in January, the five level risk and needs system. I was wondering- There's a couple things we could go, and there was an attendee who asked about this as well. We could provide a recommended or associated reading when we post this on our website tomorrow. Would that be okay with you?

Kevin Baldwin: Oh, yeah. That's a great idea.

Erin Farley: Okay. Is there anything else besides those three that I might be missing?

Kevin Baldwin: I think it would be helpful for folks to know that they could look at a book. There's a book actually called Big Data. There's a few books written a few years ago that will, I think, be scary but very informative if they have not looked at those things. Other than that, if they're not familiar with the risk needs responsivity model of Bonta and Andrews, look that up. That's really readily available on the internet and, in particular, understanding the risk principle and that informs risk assessment.

Kevin Baldwin: That's a great place to start, and, if folks have any other questions, my contact information is ... I should've- Sorry. I didn't put that up. There we go. Feel free to email me. My email address is on the screen. I would be more than happy to answer any questions we're unable to get to today or provide you with other references. Risk assessment is something I feel very strongly about. I find it very interesting, so, if you have any questions at all, you can go there. You can also look at our website, which you'll see on the screen now, ars-corp.com, and there's our phone number. If you've got specific questions in particular about risk assessment, and, in particular, even among that, if you've got questions about sex offender risk assessment, that's my particular area of expertise, more than happy to help you in any way I can.

Erin Farley: Great. Thank you. Yeah. A couple more things have popped up. Let's see. One question is there are many papers online on how organizations develop their risk instruments. Are you aware of any resource for best practices for instrument development including testing for racial bias?

Kevin Baldwin: That's a great question. In fact, there's very little, and that's why I put together the guidelines I did that I presented in our talk today is because there's a few things. The Kehl article is the closest I've come to, so, if you post a link to that, that would be really helpful. But, in general, I think we've been so busy responding to some of these constitutional issues and questions about risk assessments that we're not really yet at a point where we've come up at a comprehensive effort to develop these things, incorporating some of those
criticisms. The Kehl article and then the guidelines that I suggest I think is a good starting point.

Erin Farley: Great. Thank you.

Kevin Baldwin: Sure.

Erin Farley: Yes. Just as a reminder, we will have all of this up tomorrow, I believe sometime tomorrow, on our website. If you go to our JRSA website and you click on I think it might be resources and then one of them- Yes, resources along the bar across the top, and then, if you click on that, it says "JRSA Webinars," and it should be up there sometime tomorrow.

Erin Farley: We do have a poll. Let's see if we can run it. There we are. If you all, who are still on, if you can take just a few minutes to complete this poll, we would again greatly appreciate it, and I'll keep track over the next couple of minutes and see if there are more questions. Let's see. Here we go.

Erin Farley: Are there standards for validating, evaluating, or auditing pre-trial risk assessments?

Kevin Baldwin: Yeah. That's a great question. There are a couple of articles about developing pre-trial risk assessments that you can find on the web. There's a BJA document, basically a research document, that looks at the development of pre-trial risk assessment.

Kevin Baldwin: Probably, however, the thing to do is to just look at the broader risk assessment field, and that is, when you're developing risk assessment, of course, you want it to be reliable, and there's different ways to assess reliability. One of those is test/retest reliability. If you administer it over time to the same individual, how varied are your scores going to be all other things staying constant? Another is reliability when you've got two different assessors, inter-rater reliability.

Kevin Baldwin: Then, there's some internal reliability things. It's important just to test the general psychometrics of your measure. It's also, and with a risk assessment, critically important that you look at validity, and predictive validity is the primary means by which we assess validity of risk assessments. When we're talking about predictive validity, we're talking about: How well does the measure predict the outcome of interest?

Kevin Baldwin: There's a couple of statistics that we look at. There's this thing called the receiver operating characteristic, or ROC, and the relevant statistic is the area under the curve, or the AUC. Receiver operating characteristic stuff comes out of basically World War II when we were trying to distinguish noise and get a signal in a batch of noise. How well can we pick something out of a noisy environment? The area under the curve describes our ability overall to predict the outcome of interest. Typically, we're looking for an area under the curve of
Kevin Baldwin: The other are the error rates that we talked about earlier. The specific terms for those are sensitivity and specificity, the degree to which we make errors. In our particular application, we want to make sure those errors don't differ based on demographic characteristics. The way to do that is simply to look at your data, separate it out by ethnicity, race, the other characteristics that describe your populations you're assessing, and then make sure that those error rates don't differ.

Kevin Baldwin: Those are some general guidelines.

Erin Farley: Great. Thank you. Two things real quick. One: I don't know, Kevin, if you're familiar with the Pre-Trial Justice Institute.

Kevin Baldwin: Yes.

Erin Farley: JRSA, we did a project. We were involved in the Smart Pre-Trial Project with them, and so I would recommend anybody who's listening who - Just in case you're not familiar, you may want to go on their website and look at their information about pre-trial risk assessment. They actually- Because our project closed a little while ago, so it looks- I don't want to support something that I haven't dug into, but working with them was really great. They have really good people, so I would recommend that as a good resource to just check it for people.

Kevin Baldwin: Great. Yeah. In the broader pre-trial field, there's an excellent volume called Measuring What Matters: Data Collection and Measurement in the Pre-Trial Justice Field. If you just go to your favorite search engine and you type in "Measuring What Matters: Pre-Trial Justice," it's a PDF. It's a great document that talks about all of the different things that you need to measure, how you can keep track of how well your pre-trial system is doing, and it talks about pre-trial risk assessment in there as well.

Erin Farley: Great. Then, real quickly, we did have one more follow-up. If you had a chance, if you could mention some of the problems with the ProPublica article. Someone mentioned that there were some significant amount of problems and offers a misleading perspective of risk assessment tools.

Kevin Baldwin: Yeah. Yeah. That's true not only of ProPublica but of some of the other critiques. For instance, one of the things that some of these recent critiques have said is that the COMPAS uses 131 factors to determine risk. Well, they really don't. There's 131 questions or thereabouts on the COMPAS assessment. Not every one of those goes into the risk prediction. That's one thing is they mischaracterize the COMPAS. ProPublica did the same thing.
Kevin Baldwin: The other thing, and this gets to a very fine argument about the meaning of fairness, the ProPublica article used a particular definition of fairness, and they really got all over the COMPAS, at that point Northpointe, and said, "Hey. This is unfair." Well, Northpointe then did an analysis of the data, and they said, "Well, using this understanding of fairness, our measure is completely fair." An article came out just in the last couple of months that said, "Yeah. You can't maximize both definitions of fairness. The tool is going to do well in one and not do well in the other."

Kevin Baldwin: Certainly, there are some methodological issues. If you go to the Equivant website, formerly Northpointe, you can download their responses not only to the ProPublica article but also to an article that just came out I think it was on January 21. An article came out which cast a lot of concern about COMPAS. Northpointe/doing business as Equivant had a response I think the very next day addressing some of the criticisms, so they're very good about putting those things out, and I recommend you go to their website to read some of that.

Erin Farley: Great. We can actually try to probably provide a link to that as well. Then, one more question, do you know maybe- could speak to some of the criticisms from the Flores, Bechtel, and Lowenkamp article?

Kevin Baldwin: I can't. I know those authors, but I'm not familiar with the particular citation that they're referencing. Yeah.

Erin Farley: Webinar report part two.

Kevin Baldwin: There you go.

Erin Farley: Well, I think we are just about- We're wrapping up. We're almost out of time. Again, you had your email up there, so, if there's anybody who still has a few more questions, please contact Kevin. Feel free to email us as well, and we will pass on the question. We will have this with some additional links tomorrow. Thank you so much, Kevin, for doing this. I really appreciate it, and thank you, everyone, for attending.

Kevin Baldwin: Definitely. Thanks very much, Erin, and thanks, everybody. I hope it was profitable for you.

Erin Farley: Great. Thank you. Take care.


Erin Farley: Bye.