



Data Points

Bringing Data to the Debate on Risk Assessments
November 2024 | Podcast Transcript

Overview

This episode features Spencer Lawson and Gina Vincent, and discusses their research evaluating the disparate impact of risk assessments.

Data Points is an ongoing podcast series produced by Policy Research that discusses research for social change.

Spencer Lawson:

Our jobs as researchers is to serve as a bridge between science and our communities. And if we're making that crossing more difficult because we're using different terms, different conceptualizations and operationalizations, we're not doing our jobs.

Holley Davis:

Welcome to *Data Points*, where we discuss research for social change. *Data Points* is a production of policy research. My name is Holley Davis and I'm the communications director at Policy Research. In this episode, Dr. Gina Vincent and Dr. Spencer Lawson of UMass Chan Medical School will discuss their research evaluating the disparate impact of risk assessments. Gina is a professor and co-director of the Law and Psychiatry program in the implementation Science and Practice Advances Research Center at the UMass Chan Medical School. Dr. Vincent has specialized interests in juvenile justice, particularly with respect to implementation science and policy research, risk assessment for reoffending and violence, behavioral health screening, risk need responsiveness, and the development of screening and assessment tools. Spencer is an assistant professor in the Law and Psychiatry Program in Implementation Science and Practice Advances Research Center in the Department of Psychiatry at UMass Chan Medical School. His career mission is to manage and prevent behavioral health conditions among people across multiple points in the criminal legal system through implementation and deployment-focused research of data-driven risk screening and assessment and risk need responsiveness approaches.

Thank you both so much for joining me today.

Spencer Lawson:

Hello. Thank you for having me.

Gina Vincent:

Hello Holley. Thank you for having me.

Holley Davis:

Can you tell me a little bit about what brings you to this work, why risk in general?

Spencer Lawson:

Yeah. What brings me to this work is deployment-focused research. My passion is being a community-engaged researcher. I thoroughly enjoy collaborating with individuals that are involved and impacted by the system, practitioners that work for the system to really understand processes to provide implementation support and training and to partner with them on studies to unpack some of these social issues around behavioral health in their community so that I can provide them with that actionable knowledge.

Gina Vincent:

And I had an interest from the time I was a teenager to trying to help make the juvenile justice system a better place and make it more fair. And at the time, I was not knowledgeable enough to articulate the way I can articulate this now, but I really was wondering why the juvenile justice system was not more developmentally appropriate, didn't seem to understand youth, was treating them like adults. And I really felt like this needed to be made fairer essentially. And I learned about forensic psychology and this law and psychology intersection field, and became very interested in social justice and that's what brought me here. My background is in forensic psychology and I was trained heavily in risk assessment, and have been applying that to work with juvenile justice agencies and now adult agencies as well for many years. As Spencer said, I also love being able to work with agencies that are really trying to make a difference, impact that through science, take everything that we learn from our research and help them improve practice and it's a wonderful opportunity.

Holley Davis:

That's wonderful. Thank you both so much for sharing that.

Spencer, can you tell us a little bit about the specific study that we're here to talk about today? What was it and what were you hoping to find?

Spencer Lawson:

The research question for this systematic review was focused on whether the use of risk screening and assessment tools across that legal continuum resulted in the unfair application of the tool in legal decision making. The systematic review was focused on race and ethnicity. So whether or not the tool use led to a higher proportion of individuals of color receiving harsher penalties or unequal legal outcomes than white individuals, we were just hoping to find more research in this area on disparate impact in risk screening and assessment than what's a prior systematic review back in 2019 in law and human behavior, what they found so that we can more definitively speak to whether risk screening and assessment tools are increasing or decreasing system disparities.

Holley Davis:

That's great. Thank you so much for contextualizing that for me.

Gina, let's talk a little bit about defining what a risk assessment is. My understanding is that a risk assessment is a tool for criminal legal professionals to assess a person's likelihood for either committing a new crime or failing to appear in court for their proceedings. Is that right?

Gina Vincent:

That is right when we're talking about pretrial tools. Risk assessment has been around for a long time and probably came to be originally in forensics psychology and forensic psychiatry context to identify someone's violence risk. What's the likelihood this person's going to go out and hurt somebody else and what do we do about it? The pretrial risk tool is for identifying the likelihood somebody's going to re-offend, recidivate, and identifying the likelihood they will flee and not appear in court.

Most risk assessment instruments do not consider that piece about failure to appear. Some are designed specifically for violence risk, others are designed for recidivism more generally. There are these short risk tools like a pretrial tool. It can be completed by someone without specialized clinical training. It can be completed largely without an interview and just based on somebody's official records. They're designed specifically just to make those estimates of risk. But then there's more comprehensive risk assessment instruments that are conducted by someone with more in-depth training. It doesn't have to be a clinician. And those instruments are really used to tell us not only what's the likelihood somebody is going to get in trouble with the law again by reoffending or being violent, but they're also going to give us some information about what might be driving that behavior in this individual so that we know what to put in place in order to intervene effectively.

For example, if an individual's strongest risk factor seems to be their drug or alcohol misuse, we can make that leap. That's likely what's causing their offending behavior, and if we just help them to get sober or give them some tools to help with their recovery, we may prevent them from continuing to do that in the future. I think the main point here is that there are many types of risk assessment tools. They have specific purposes. They have specific designs and the types of instruments that were used in the study that Spencer and I did covered the whole gamut.

Holley Davis:

That is such helpful context and that really broadens my understanding of risk assessments overall. Thank you. Gina, let's go a little bit deeper then into this study. How do race and ethnicity impact risk assessments in the scoring, the application and the decisions?

Gina Vincent:

That's a terrific question, Holley, and it's a loaded question because there is a lot of concern that race is influencing scoring of risk assessment instruments, especially these pre-trial ones. If it's a good,

well-validated instrument, race is not going to impact the scoring of it at all. In general, on these risk instruments, they have a number of items. These items go together in some kind of summation or algorithm and that total score should then be used to determine what's that person's risk level. That final score should not be influenced by an individual's race.

However, some of the items in these tools may reflect systemic inequities in our country that can carry forward into inequities and risk screening tools. And the items that we're most concerned about, not shockingly, are any items that are tied to one's official criminal record or their official juvenile record. They could have racial differences in them. Are anything pertaining to the individual's neighborhood, things like community disorganization, are there high crime rates there? Is there a lot of policing there and is there a lot of drugs or gang activity?

So if you've got risk tools that are heavily reliant on one's official records of criminal history or have anything pertaining to one's neighborhood or geography, you're going to have some racial differences on that tool most likely, unless, again, it's been well-validated and there's other items to balance it out. A lot of risk instruments have been tested at this point to see if they're leading to racial bias. And in general, the research is showing that most risk tools do not have racial bias.

Spencer Lawson:

I just want to chime in. The phrase that always sticks out to me when we're looking at some of these items that folks are calling into question is, are we baking in some of these larger systemic inequities into those items? And for me that really resonates and it's easy for me to comprehend that side of the argument when it is just framed like that. So I just wanted to chime in and provide a different way to phrase the conversation.

Holley Davis:

That's very helpful. Thanks so much. What would balance the risk assessments?

Gina Vincent:

What I mean by disparities in this case is just significant mean differences between racial groups. We don't see big differences in those dynamic risk factors, which are things like their attitude towards the legal system or their attitude condoning violence or condoning crime. We don't see big differences when it comes to education background or work background. We don't see huge differences when it comes to some types of personality characteristics like do they have difficulties focusing? Are they impulsive? Do they have difficulties with remorse or empathy? We don't see big racial differences on those kinds of dynamic risk factors, and yet those things still predict offending.

Where we do see big racial differences on some of those dynamic risk factors tends to be around drugs and alcohol use. The white individuals who get arrested or end up in the system tend to have much higher rates of drug and alcohol use than the people of color. I am a fan of instruments that have

some measure of prior legal history, but also should be balanced with some of those other changeable dynamic risk factors that help us determine where to intervene and that have less racial differences at the main level.

Holley Davis:

Spencer, when I was reviewing your paper, I came across some terms that I was unfamiliar with and was hoping you could define for me. So test bias, predictive impact, and disparate impact. Could we start with test bias? What does that mean?

Spencer Lawson:

Not only is there a laundry list of terms, scholars may be using them interchangeably. They may be having differing definitions associated with these terms or they're using antiquated indicators. When you start having all of these different terms and these variations across studies, I would say that really just muddies the literature on fairness and bias in rescreening and assessment. It introduces this idea of theoretical indeterminacy, plus it also presents practical challenges to practitioners who are wanting to advance reform efforts in this space.

So in this risk context, test bias can also be called predictive bias or there's differential prediction. In my view, I would lump those all three in the same bucket. And since the systematic review was focused on race and ethnicity, once again going to use race as an example to explain test bias. The tool is functioning differently across groups. When I think of test bias or predictive bias, it's how does the tool function? And really, scores should statistically relate to the outcome it was designed to detect, the same rate regardless of group membership. Race really should not be impacting scoring at all. So there's that term, test bias.

The second term that I want to focus on is disparate impact. The word that I like to think about when talking about disparate impact is application. It's the application of the tool in legal decision making. So this would mean that one group is treated more harshly by the system than another group because they score higher on a risk instrument on average. There's a phenomenal paper in the Journal of Criminal Justice and Behavior led by Sam Zottola that looks into this idea of fairness within risk screening and assessment and intervenes on some of the debates that we're having right now with what is the best way, what's not the best way. So I would highly recommend that paper.

Holley Davis:

I'm hearing we need some more common definitions that are more widely applied in the field.

Spencer Lawson:

I welcome a conversation around having a structured vocabulary taxonomy in this space because I do think that's a struggle when our jobs as researchers is to serve as a bridge between science and our

communities. And if we're making that crossing more difficult because we're using different terms, different conceptualizations and operationalizations, we're not doing our jobs.

Holley Davis:

So Gina, what does it mean to be fair when using risk assessments? I know fairness is a subjective term.

Gina Vincent:

Fairness is tricky, right? A lot of scholars, legal scholars, researchers, people who work in this space have tried to define fairness. The short answer is, with respect to a risk instrument, fairness can be defined as a tool that is functioning in the same way regardless of people's demographic characteristics. So in other words, a female should be scoring in the same way relative to her risk level for reoffending as a male would. Given a score that is equivalent to a male score, they should be having the same likelihood of reoffending. The same is true of race. Regardless of your racial group, the tool should be operating the same for you that it is for any other group. So higher scores mean higher risk of reoffending. That's one definition of fairness and we call that test bias. It means it's lacking test bias if it's functioning in the same way regardless of your demographic characteristics.

Another definition of fairness is that the instrument be interpreted in the same way regardless of what demographic group you belong to. So the instrument should not systematically be leading to harsher treatment for one group than another, given the same scores. Any instrument should be examined for this type of test bias to make sure that it's operating in the same way. I think one area where we may be most apt to find a lack of fairness in the equitable use of tools, it's been attributed to the people who are using the tools to make the decision. So the frontline staff, detention staff might be filling out tools, pretrial officers, probation officers, but they're making a recommendation to a court. Where we would see something being unfair is if the officer says, "Hey, this person scored a five. That puts them in the moderate risk category," and then the decision maker, whoever that may be says, "Oh yeah, but they committed a really serious offense and therefore I'm going to override them." That's when we start considering the use as unfair.

Holley Davis:

Before, we mentioned that risk assessments can be a little divisive and it seems like the field seems to be a bit split. Some groups think that they're a very objective tool while others believe that risk assessments further disparities for minoritized groups. Why is that and what's your take?

Gina Vincent:

I think this goes back to Spencer's comment about the fact that people who are putting out disseminations have done a great job of confusing practitioners when it comes to risk assessment. Some work that came out from Angwin and others from ProPublica came out with a very harsh statement that these tools are biased. Once they put that out, several big organizations who prior had

been helping places put risk tools into place, stopped doing that and came out with statements that they're not going to do it anymore because these instruments are biased and hurting people of color. I feel like some of those statements were not well-informed because they were basing it simply on this piece. So I feel like that's what has led to a lot of the split when it comes to thinking about risk assessment instruments, but for those who are studying them, I can say that the work that has come out on risk assessment instruments has shown the vast majority do not have test bias. Part of the problem is that Angwin and folks we're defining test bias based on false positive rates.

The other issue is that there has been some misuse of risk instruments because there's been a lot of overrides due to policy, whether it's statutory policy or local policy, leading to individuals of color getting treated more harshly, but that's not because of the tool. That's because of a lot of policy overrides.

Holley Davis:

It seems like public opinion was set 10 years ago and there hasn't been that same groundswell to change it or provide any sort of context for how research has changed, how application has also changed.

Spencer, I'd love to hear your thoughts on this.

Spencer Lawson:

When you're talking to practitioners and you use the term bias, I think immediately they think of discriminatory practices and that can be a challenge because if you're talking about test bias or predictive bias, really we're talking about the statistics. And I feel like more often than not, I'll hear a practitioner say, "Well, the score is the score." As the individual that's completing it, we're not doing it in a way that disadvantages Black individuals compared to white individuals. We have to reorient them where there's discriminatory practices, that's one conversation, but when we're talking about predictive bias, it's something different. And so I think the education piece is constantly there when we're doing this type of work and interfacing with our community.

Holley Davis:

Spencer, you conducted a systematic review, which means that you pulled and analyzed data from numerous published studies. What did you examine?

Spencer Lawson:

Yeah, so this was a team effort. We had some inclusion criteria that we applied to the literature, and after we did all of that, we ended up with 21 articles that we examined that really intervened on this topic of disparate impact, and those articles were focused on 13 risk assessment instruments and a variety of different legal outcomes. We really did open up the systematic review across the entire criminal legal continuum. For us, that disparate impact term is defined as mean score differences between subgroups on an instrument that led to the unfair application of the scores or unequal equal outcomes.

In terms of what did we examine, about half of those 21 articles would be classified as peer reviewed articles. The remainder were either government or technical reports, master theses, dissertations. Eight of the 13 risk screening and assessment instruments that were included represented validated tools. What we mean by that term is there was a research team that went in, conducted a study and examined the psychometric properties of the tool that was being used.

And then, lastly, a majority of these articles focused on tools that were implemented pre-adjudication. Also, a best practice when conducting a systematic review is assessing the quality and rigor of the methodology being used. So our team also did that as well. We used the quality assessment tool for quantitative studies as our means of assessing that methodological quality.

Holley Davis:

Fantastic, and so let's get to my favorite part, the results. What did you find in your research?

Spencer Lawson:

The take-home message is no strong evidence of differential treatment by the legal system based on race or ethnicity as the result of risk instrument use. We had 21 articles. 18 of those articles revealed evidence that risk instrument use, either we found a null effect on disparities or there was some sort of beneficial impact. So disparities were reduced by the use of the risk screening and assessment tool. When we say null effect, what we mean is that risk instrument use did not lead to the creation of disparities in legal decision making, nor did it worsen existing ones. That means then the remainder, those eight articles, they actually found some sort of advantageous or beneficial impact with their use. So disparities were being reduced.

A really important finding is going back to that methodological quality. So out of those 18 articles, only five of them actually had strong methodological quality. So now pivoting then to those remaining three, those were the articles that we actually did find evidence of disparate impact on risk assessment instruments. And a note, only one of those articles actually had strong methodological quality. So overall, the current evidence signals that risk instruments are not resulting in harsher treatment for individuals with minoritized racial and ethnic identities.

Gina Vincent:

I think the take-home message is there's no good evidence that these tools are leading to a disparate impact or treating one group more harshly than another, and in fact, there's more evidence showing that it's helping reduce disparities than there is evidence showing that it's increasing disparities.

Holley Davis:

Spencer, in your paper, you've noted that there were not many high quality articles for you to select from. Why do you think that is?

Spencer Lawson:

Yeah, great question. I would like to point to two potential reasons. One is just higher quality research is difficult to do. Quality research on the disparate impact of risk screening and assessment requires both pre-implementation data and post-implementation data. Scholars need data on the percent of people in each racial and ethnic group who receive the legal decision-making outcome in question before a risk tool is put into place, and then comparing that to the data after the risk tool is put into place.

The second reason is the importance of examining disparate impact of risk tools has received attention relatively recently. So I would like to point to a journal called *Criminology*. They have a 2016 article by Jennifer Skeem and Christopher Lowenkamp. This article by them really elevated the concept of disparate impact within the field. So it's still a relatively new area that us scholars are exploring within risk tools.

Holley Davis:

That's really helpful to hear.

Gina, what role do you think that education and training play in the effective use of risk assessments?

Gina Vincent:

It's very important to have education and training related to how to complete the risk assessment instrument. So the judges, the officers who are going to be completing the instrument, the institutions that are going to be looking at the data, the attorneys, so that we're sure that they have buy-in to the process and everybody's voice is heard, it's critical that everybody be educated about the why have their input and understand how this needs to work in order for it to effectuate any change.

There's two really good guidebooks that I'm going to refer to here. There's *Risk Assessment in Juvenile Justice: A Guidebook for Implementation*, which is something that I put out with Laura Guy and Tom Grisso where we talk a lot about where education factors in and all the implementation steps for putting a risk tool in place effectively. And then there's also the one by Sarah Desmarais at PRA, David D'Amora and Lahiz Tavaréz, and that's called *Advancing Fairness and Transparency: National Guidelines for Post-Conviction Risk and Needs Assessment*. Risk assessment, as we talked about, is done in multiple steps. So these two guidebooks together talk about every step at which risk assessment may be implemented and where education falls into place. The most effective implementation strategy is going to involve getting buy-in from all the partners. And education is not a one and done. It's never a one and done. It's constant messaging, not only because we have turnover, but also because we're humans, we forget. And so constant messaging is going to be critical.

Holley Davis:

What sort of cadence do you recommend for that messaging?

Gina Vincent:

Every six months minimum, but that's where you actually want to intentionally put out a message about it or do another training about it. I think anytime that an officer is presenting results of a risk assessment instrument or a risk needs instrument to the court or to an attorney, it's important to remind them about risk need responsivity, which is really the framework that we're largely trying to follow here.

Spencer Lawson:

I think many of the legal actors that we involve ourselves with when it comes to risk screening and assessment, they wear multiple hats. One hat is being a screener, but they also are wearing this education hat because they themselves are getting in front of the judge or their peers and they're having to explain, like Gina mentioned, the value of risk need responsivity, infusing into their process.

Holley Davis:

Thanks so much, Spencer.

Gina, what kind of policy or practice changes do you believe could improve the perception of fairness and accuracy of risk assessments?

Gina Vincent:

I think the education piece is an important one. Ensuring that you have examined your instrument, ensuring that it's been examined for fairness, reminding folks that this has been tested and we know that it works for these different groups is going to be critical.

But there is another really important policy and procedural change that I think we have hit on recently in our work with PRA that's been funded by the MacArthur Foundation, and that is that unhook the instrument, the risk instrument from the policy. So this was blown wide open by Jennifer Scheme Montoya and Chris Lowenkamp. In 2023, they came out with a widely publicized piece, looking at the federal pretrial tool and found that 79% of racial bias in federal detention settings is coming from mandatory policies and criminal history, not the risk tool that's being used. And the kinds of mandatory policies that we're talking about, the common one is type of offense. Did the person come in on a warrant? Did they come in on a violation of some sort? Those are the kinds of policies that are leading to the bigger racial disparities.

Alex Piquero, the former director of the Bureau of Justice Statistics said this in a webinar recently that we really need to consider, are we putting people in detention because we're mad at them or because we're scared of them? And if we're putting people into detention because we're mad at them, this is often where these policies are coming from. These policies do not, at least the ones we have studied, the ones that are tied to criminal history or current offense, are not predicting recidivism. And I think every state should really take a look at what their mandatory hold policies are and unhook that from the tool when they're doing their analyses. The tool may be fair, but the policy may not be.

Holley Davis:

There's so many things to balance.

Gina Vincent:

There's a lot to balance there. I'm glad that you raised that because I don't want to come across as someone commits murder, but they're low risk, so we should let them go. I don't want to come across that way because obviously the people of the court have this high hazard that they're worried about, but we need to really look at our policies and think about, is this here because this person is a risk to public safety, they're not going to come back to court, or is this here because it's longstanding issues about who we're upset with?

Spencer Lawson:

And I'll just chime in there real quick. I would caution jurisdictions from knee-jerk reactions to an event that happens within their community and then that leading to the creation of a policy or mandate that hasn't had the same level of input from the community as a whole.

Holley Davis:

Really eye-opening. Thank you. Spencer, what is next for the field?

Spencer Lawson:

We need more strong pre, post, quasi-experimental studies on disparate impact of risk screening and assessment, particularly at other points besides the pretrial setting, right? It's the pre-data that we as scholars need to be mindful of when we are starting a project with our community partners to make sure that infrastructure is in place to be able to collect that information. More research on races other than Black and white individuals. I want to caveat that by saying we don't want to stop the research of looking into predictive bias and disparate impact of risk tools among Black and white individuals. However, we need to expand to other racial and ethnic identities. I just want to make a plug for the team at PRA because they are doing this. They recently got to their article accepted within this special issue of law and human behavior that was looking at predictive bias within a pretrial risk tool among a Native American population.

More research with the voices of those that have lived experience within the legal system, infusing those perspectives into the work that us as scholars are doing is invaluable. Scholars need to get on the same page. Only five of the 21 articles that were included in our systematic review actually used this term disparate impact. I think that us not using a similar language, a share taxonomy in this work only creates theoretical indeterminacy in the literature and it also confuses practitioners because when Gina goes to jurisdiction A and uses disparate impact, but Bob goes to jurisdiction B and uses another term, are we touching upon the same concept or is it something different? And that's just confusing for practitioners that are wanting to engage in meaningful work and reform efforts within this space.

Holley Davis:

Excellent. Thank you so much for sharing your research with me today. I've enjoyed learning more about how risk assessments can contribute to reducing disparities in the criminal legal system.

This has been an episode of *Data Points*, a production of Policy research. Learn more about us by visiting us at prainc.com. If you have questions or comments about this episode, email us at communications@prainc.com. *Data Points* is available via Spotify, Apple Podcasts, and SoundCloud.

This research and this episode were made possible by the John D. and Catherine T. MacArthur Foundation. The content and opinions expressed are solely the responsibility of the authors.

This episode was produced and hosted by Holley Davis, engineered by Nicole Vincent-Roller and Elianne Paley, and edited by Elianne Paley. Until next time, keep creating positive social change.

About

Policy Research Associates, Inc. is a Women-Owned Small Business (WOSB) founded in 1987. We offer four core services: research, technical assistance, training, and policy evaluation. Through our work, we enhance systems that assist individuals with behavioral health needs on their journey to recovery.

Contact

Policy Research Associates, Inc.

🏠 433 River St, Suite 1005, Troy, NY, 12180

✉️ pra@prainc.com

🌐 prainc.com

☎️ 518-439-7415