Probable Causation, Episode 9: Michael Mueller-Smith

Jennifer [00:00:00] Hi, Probable Causation listeners. Before we start today's episode, I want to let you know that we've launched a Patreon account for the show. Making this podcast is fun and rewarding, but unfortunately, it is not free. By becoming a supporter on Patreon, you'll help make this endeavor more sustainable. For just five dollars a month, you can enjoy the warm glow that comes with supporting a public good. You'll also get occasional bonus episodes. And we've got some fun stuff in the works there. For ten dollars a month, you'll also get extended cuts from our interviews with early access to our interview schedule and the opportunity to submit questions for the guests to answer during those extended interview segments. You can subscribe at either level at patreon.com/probablecausation. That's p-a-t-r-e-o-n dot com slash probable causation. If you're interested in sponsoring the show at a higher level, please email us at sponsor@probablecausation.com. We'd love to chat with you. In any case, I hope you'll consider becoming a supporter of the show at any level. Thanks in advance for your financial support. And as always, thanks for listening. And now onto this week's show.

Jennifer [00:01:14] Hello and welcome to Probable Causation, a show about law, economics and crime. I'm your host, Jennifer Doleac of Texas A&M University, where I'm an Economics Professor and the Director of the Justice Tech Lab.

Jennifer [00:01:25] My guest this week is Mike Mueller-Smith. Mike is an Assistant Professor of Economics at the University of Michigan and the Director of the Criminal Justice Administrative Records System, or CJARS. Mike, welcome to the show.

Mike [00:01:38] Thanks, Jen. Congrats on the podcast. I've really enjoyed listening for the last few months.

Jennifer [00:01:42] Oh, well, thank you. So, yeah. So we're going to talk today about CJARS as well as a new paper you've written using CJARS data on the effects of the criminal justice system on inequality and economic outcomes. So to set the stage for us, could you tell us about your research expertise and how you came to work in this area?

Mike [00:02:01] Yeah. So I'm an Assistant Professor of Economics at the University of Michigan. My research focus is twofold. I'm interested in quantifying both the scope and impact of the justice system in the United States. The idea being, can we assess whether policy and practice are delivering their intended objectives of creating safe, productive and healthy communities in the U.S.? And second, behind all of this work - whether it's studying incarceration or diversion in Texas, domestic violence in New York City, or the multi-state inequality project we're discussing today - I'm really excited to about how we can leverage administrative records for research purposes in the social sciences. So every day, businesses, government agencies, and nonprofits go about their daily work, logging individual level transactions on a multitude of important social and economic phenomena. Whether they realize it or not, these records can be used as data for research purposes. And what I find really attractive about this idea is that it's a potentially low cost, sustainable resource that has transformed a variety of fields of research today. So in this vein, I'm the co-founder and director, as you mentioned, of the Criminal Justice Administrative Records System, or CJARS, which is trying to bring this revolution in information technology to the space of criminal justice research.

Jennifer [00:03:16] And I know that a lot of listeners out there are interested in data - especially new data, who doesn't love new data? - and using data to better understand the

criminal justice system. So I do want to spend some time today talking about the CJARS project that you direct. So, yes, so tell us more. What is CJARS and how long has this project been underway?

Mike [00:03:35] So CJARS is the first nationally integrated criminal justice data repository that can follow individuals through the through the criminal justice system, as well as link with existing social and economic data held by the federal government. It's a joint collaboration between the University of Michigan and the U.S. Census Bureau, where my Co-PI, Keith Finlay, is based. And we're currently collecting individual level digitized administrative records from police, sheriffs, criminal courts, institutional facilities and community supervision programs like probation and parole. The project was started soon after I arrived at Michigan in 2015 with pilot funding from the Laura and John Arnold Foundation. In the time since, we've shifted to being supported by the Bill & Melinda Gates Foundation and we've recently finalized a new four year funding agreement with the National Science Foundation.

Jennifer [00:04:24] So how many states do you currently include in your dataset? And I guess more importantly, certainly, I'm sure many people are wondering, how have you obtained those data?

Mike [00:04:35] So we currently have over 700 million lines of raw data from 14 states that cover roughly 57 million unique criminal justice events for, for instance, like court charges or prison entries. And this covers over 17 million unique individuals. The bulk of the information, though, comes from 8 states where we have both current population information as well as historical records dating back, often multiple decades, into the 1980s. We collect our data through three mechanisms. First, when agencies post data to the public domain through their websites, we seek to harvest that information through a technique known as web scraping. Essentially, we write algorithms that directs servers to search and document any published caseload information from criminal justice agencies on the Web. Second, when data is not in the public domain but is legally considered public, we file Freedom of Information Act or FOIA requests to purchase similar information about caseload composition and criminal justice outcomes. Finally, when information is not legally public, we reach out to agencies to establish data use agreements, which provides us access to confidential caseload information. The production and maintenance of criminal justice records in the U.S. is highly decentralized down to the state and often county level with a wide range of access rules and procedures. And so we try to use a flexible approach to data collection through adapting to local environments to ensure that we cast the broadest net possible.

Jennifer [00:06:08] Are you able to tell us which states make up the bulk of your data at this point?

Mike [00:06:12] So the bulk of the data currently comes from eight states. And if I can remember this correctly, it is Florida, Michigan, Texas, North Carolina, Nebraska, New Jersey, Pennsylvania, and Wisconsin.

Jennifer [00:06:27] OK. And I assume the goal is to eventually get to all the states. Is that true?

Mike [00:06:33] The goal is to have this be a national platform covering all of the United States. We'll talk a bit later about some of the challenges associated with this. But that is on the horizon, although it's a distant horizon.

Jennifer [00:06:46] So some people might be thinking, you know, we have the Bureau of Justice Statistics that already collects lots of data. So why don't we talk a bit about kind of how this differs from what already exists? Could you compare the resulting dataset that you've put together to other available data like the NLSY and the National Corrections Reporting Program and the state court processing statistics and all the other datasets collected and maintained by BJS? And I guess that the overarching question here is what can we do with CJARS data that we can't do with those existing datasets?

Mike [00:07:23] Yeah. So I want to start by talking about the constellation of data collection efforts by the federal government and the Bureau of Justice Statistics in particular. So, BJS funds a number of efforts to report on the criminal justice system that most people working in this area will already be familiar with. There's the Uniform Crime Reporting Program or UCR, the Censuses of State and Federal Adult Correctional Facilities that support the production of the annual reports on the size of the correctional population, the Annual Parole and Probation Surveys, the National Correction Reporting Program, just to name a few. The data collection efforts from the BJS often target a single stage of what we think of as the complete criminal justice system. So UCR tackles offense counts and arrests. The survey efforts target institutional confinement or community supervision. With the one exception of a project that's linking multi-stage data from the federal criminal justice system specifically, the reporting on state criminal justice systems is segmented into the specific independent parts. At the same time, the majority of this information is not longitudinal in nature. So what I mean by that is we can't follow the same offenders over time, even within, for instance, arrest data or within court data. And we can't integrate this information with non crime outcomes like earnings and employment trajectories or household composition and family structure, things that might be impacted by criminal justice interactions or be a cause of criminal activity in the first place.

Mike [00:09:01] The state court processing series, which you mentioned, which is one of the most important information- sources of information on criminal court proceedings in the U.S., actually has in fact been dormant since 2009. So even if the issues of integration and longitudinal data structure were solved, BJS doesn't necessarily have the resources at hand to consistently maintain all the information they seek to collect. So in the absence of kind of comprehensive, longitudinal, integrated data, there have been a variety of nongovernmental efforts that have come to the front and have been playing an important role in research. So, for instance, the National Longitudinal Survey of Youth, while not intended as a criminal justice specific dataset, does collect information on interactions with the justice system. And so there's information like NLSY or the Fragile Families dataset, which have become really important pieces of data in terms of enhancing our knowledge about the interactions between the criminal justice system and society overall. Now, these are limited, though, in the fact that they're one cohort and nationally representative. And so the level of detail that you can get into is somewhat constrained.

Jennifer [00:10:17] Yeah, and the NLSY data, I assume Fragile Families also, is self reported right in terms of criminal justice involvement.

Mike [00:10:25] That's exactly right. So, you know, and that actually brings up another source of data that BJS collects. So we have, for instance, the National Survey of Inmates, which is the source of data that we use to calculate how many children are affected by parental incarceration in the United States. And so we ask inmates how many minors they have parented, and from that extrapolate out to a national figure. And if there are reasons why people might have concerns about accurately reporting that information, whether it's

because they're ashamed of the fact that they might be in prison and have minors on the outside, or they think that this might influence their ability to get parole, then these pieces of information could be biased. And so survey information can be really valuable in that it collects a lot more detail than we might see in administrative records. But it does raise this concern about one, who can we actually observe in the survey data — so, who responds to surveys — and then two, how accurate is that information if we might have something like social desirability bias influencing it?

Jennifer [00:11:35] Right. Yeah. I mean so administrative data, as many will quickly point out, doesn't necessarily tell us all of the crimes someone commits, say, but it will tell us all of the interactions that someone has had with the system, the actual arrests and the actual convictions. So you don't have to worry about losing people if you're surveying them over time and and all that, so admin data are really nice. It is really amazing, I feel like every time I talk with policymakers or, you know, any just citizens who are interested in criminal justice but aren't researchers themselves, I think that one of the first things I have to clarify is how little we know, right. Like how incomplete all of these data efforts are. And, you know, there're the people who work at BJS and in these other other spaces obviously are doing the best they can, but it is just remarkable for such an important set of programs across our country that we just have so little data. And I imagine this is something you also wind up talking about a lot, just like pointing out the the big hole here.

Mike [00:12:38] Yeah, so just to follow up on that for a second. You know, a lot of people are surprised when I talk to them about the fact that we don't currently have a data system in place to say how many people in this country currently hold convictions or felony convictions. We don't know how many people have ever been incarcerated. We don't know how many children have ever had parents involved in the criminal justice system. There are researchers who have made attempts to try to get at this information. And often these will take assumptions in terms of using imperfect methodologies, but to try to approximate the best that we can, given the available information. But we currently don't have a system in place to actually track this information.

Jennifer [00:13:21] Right. Right. It's just amazing. So so kind of along those lines, I mean just to get a bit more in the weeds for a minute, you're collecting data here. I mean, the actual logistics of doing this are you know, it's just it's really difficult, right. So you're collecting data from across the country that's produced and maintained often at the county level or city level without a common data standard or format. Everyone, you know, has their own data system. So you can't even link usually within a county like across arrest to convictions to incarceration, like even that is really hard. So for listeners out there who have not themselves had the pleasure of working with administrative criminal justice data, explain why exactly this is so difficult and what challenges are unique to this enterprise that other data collection efforts typically don't face.

Mike [00:14:10] So every National Academies review of criminal justice in the past 30 years has noted the lack of a national integrated justice data system in the United States and that without such data, research policy and practice have been potentially narrow in both scope and effectiveness. But the problem has been that agencies are often siloed, operating largely independently, frequently at different jurisdictional levels. And so there hasn't been an easy solution of integrating this data given the legal and financial barriers. So one challenge that you alluded to is simply acquiring the data. You know, we're really proud of the data collection efforts that we've had to date. But realistically, getting to 50 states comprehensively is going to be a 15 to 20 year challenge at least. We only have so many people that we can be engaging with at a given time for data use negotiations. And

so reaching this 50 state goal will take a number of years. Second, the data that we collect doesn't contain unique identifiers. People often joke about how, you know, there're different qualities of data across different areas of social policy. Health data is often raised as one area that's thought of as very bad. But criminal justice's data I think is universally considered the worst.

Jennifer [00:15:32] We've got them beat, yeah.

Mike [00:15:34] So, you know, we don't have something like a Social Security number that's available across all of these different datasets that we're collecting. Often the best personal identifying variation or information that we have are just names and dates of birth in the data. And if you think that somebody might be using an alias or there might be typos when the handwritten records are transcribed, we have to think strategically about how we could actually try to recover all of the different criminal justice events for one person across time. So we deploy algorithms to figure out who the same person both within and across these different datasets. And one very cool thing that I'd like to talk about is that we have an entity resolution algorithm, which is just a fancy way of saying probabilistic matching, which has been trained on a subset of our records with biometric identifiers.

Mike [00:16:28] And so what that means is we have data where we have a true fingerprint based ID where we know which of the records actually belong to the same person. Yet we have PII, or Personally Identifying Information, like names and dates of birth that are recorded as they were originally entered into the criminal justice data system. This is somewhat unique. Often when we're looking at jurisdictions that have these biometric IDs, they go in and they overwrite any variation in the PII to have it match according to who they think the true person is and what their name and date of birth is. But in these instances where we have this original variation, all of the flaws and messiness in the admin records but this measure of the true underlying match with the fingerprint based ID, we can develop a machine learning based probabilistic matching model that's trained on this. So we can decide how to compare across different variations of name information or date of birth information how much weight should be placed on whether or not the last name matches or the last name is off by a few letters versus the date of birth.

Mike [00:17:47] And so with that, we've been able to achieve things like precision rates, which are- is a statistics used in the Matching literature, which is a measure of how many of the matches that we make are true matches when we're going through and deploying this machine learning algorithm. And we find performance statistics around 99.5 percent, which is really, really exceptional in the space of administrative data. So we're really excited about this innovation, and I think that this is one of the places where criminal justice administrative records can actually give back to the rest of other types of research on administrative records, whether it's SNAP or other types of social assistance programs or labor market data where, you know, you wouldn't typically see something like a fingerprint. So this is pretty exciting.

Mike [00:18:39] Finally, each dataset that we acquire has its own unique way of coding stored information. Sometimes this isn't a big deal. So, like, if a date is stored as year-month-day versus month-day-year, that's pretty easy to handle. We can code that up pretty quickly. In contrast, issues like what type of crime the offense was can be really complicated when fields may be entered as freehand text. And have, you know, potentially over 100,000 unique values for all these different- across all of these different datasets. So what we're working on right now, and this is something that's very much still in progress, is leveraging another type of machine learning algorithm where we are using an existing

crosswalk that was developed by the Bureau for Justice Statistics, where they've taken the offense codes in the originating data for the National Crime Reporting Program, which tracks entries and exits into state prisons and harmonized that into a single nationally harmonized offense code. So it's a subset of codes that maps these 100,000+ unique values into one common format across all of the different states. And when we're using that to leverage, you know, things like mapping the predictive value of a specific word or phrase or similar types of combinations of words. So, for instance, "possession of" could be an indication that a crime was a drug possession crime, even if we don't have the exact description mapping into the current BJS crosswalk. So we're working on developing that right now and I know that other groups are also working on trying to develop these offense coding algorithms. NORC is one group, Measures for justice is another group. And we're excited to see this field develop.

Jennifer [00:20:36] Yeah, and I think just the standardizing of the offenses over places is really interesting. And another thing I think most people don't think much about. You know, we have the Uniform Crime Reports, which is, you know, in many cases the best data we have that standardized across all places in terms of kind of counts of the number of offenses in different categories. And even then, it's a relatively small number of categories. And I think, you know, one of the reasons for that is that it's actually really hard to standardize the definition across places and have a robbery mean the same thing across different jurisdictions. And like getting everyone on the same page about that to upload the data is actually like a huge logistical challenge. So this is fascinating.

Mike [00:21:15] Yeah. So one of the things I know that you want to talk about data access in a moment, but one of the things just thinking about what we're trying to do as an organization, we're going to try to do the best job that we can, and I'm sure that there are others who can add more to this. When we go through to create an external access mechanism, our intention is to provide, you know, the coded information as we've pulled it together to lower barriers to access so that it doesn't feel overwhelming to work with this data. But we're also intending to provide access to the raw, original, very messy offense code information and the same would correspond for a variety of other variables: Provide the original source format so that if there's some sort of specific detail — you often see this in legal research on how certain crimes are prosecuted in one way versus another — that variation will still be in there, even if we've scrubbed over that in the algorithms that we've deployed.

Jennifer [00:22:16] Right. And we will get to data access in a minute, but I do want to talk a little bit more about the possibility here. So talk a bit more about your partnership with the U.S. Census Bureau. That aspect of all this seems especially exciting in terms of what these data will be able to tell us.

Mike [00:22:33] Yeah, so I think the partnership with the Census Bureau is one of the things that really sets our project apart. So all of the data that we collect at the University of Michigan, we go through, we harmonize it, we link up the same individuals. We try to map out what we call the criminal episodes so we can say which arrest led to which court charge, which court charge led to which period of sanction. And once we have all of that data cleaned up, we take that and we ship it off to the Census Bureau. We've already, in the data use agreements that we've signed, we've negotiated the right to redistribute this information to the Census Bureau. And so we're- this is- kind of reaching that step of being able to transfer this information, although it sounds simple, is actually pretty complicated and took a lot of work. But we send the information off to the Census Bureau and it goes through another round of probabilistic matching to integrate our PII into what's known as a

PIK, or I believe this is the protected identification key, which is kind of like a hashed Social Security number that's held internally at the Census Bureau. It's never published anywhere. There's no reason anyone would ever interact with their PIK, although you and I and everyone else in the United States has one. With the addition of the PIK, you can take CJARS then and integrate it with a variety of different types of data held by the federal government. So one of the things that we're doing in the project we're talking about today is we've linked CJARS data at the individual level to individuals own responses in the American Community Survey, as well as the decennial census. But there are a variety of other types of data that you could think about, things like administrative measures of earnings and employment, like the LEHD or the Longitudinal Employer Household Dynamics Program, or there's data held on public assistance programs, or a variety of other things.

Mike [00:24:33] One of the things that's really challenging in negotiating data use agreements with criminal justice agencies, with this idea that we're going to be hosting their data for researchers to request and access it, is concerns about data security. And what's really valuable about the partnership with the Census Bureau is that they've basically already solved this problem and we don't need to reinvent the wheel. So we are using the federal statistical research data center system to provide a mechanism for external researchers to request, access, and analyze CJARS integrated with a variety of other data held by the federal government. What's exciting about this RDC system is that there are a number of both physical as well as information technology safeguards that have been put in place that really help instill confidence with criminal justice agencies, that we're taking the protection of their records very seriously.

Mike [00:25:33] Finally, one of the things that I think is really important and exciting about this project, and through our partnership with the Census Bureau, is that all of the records that we're collecting will become part of a permanent federal archive of criminal justice data that could be requested by you, me, any researcher who can go through a federal background check and acquire what's known as special sworn status with the U.S. Census Bureau. And I think this is really important. I think we're at a time where there's a lot of interest and excitement about reforming the criminal justice system in the United States. I think that we have a responsibility and an obligation in the research community to make sure that we have the best and the brightest minds working on this. And so a major goal in CJARS is democratizing access to this really high quality, integrated administrative data. For too long, I think that the research on criminal justice within the economics community has been dictated by who has access to what data and how individual preferences have set the research agenda. And I think it's really important to think about ways in which we can really support all individuals accessing the data so that data access isn't a limitation on the knowledge that's being produced.

Jennifer [00:26:58] Yeah, that is- it's an incredible public service that you and Keith are doing to make all these data both exist and available to everyone. Do you have a timeline in mind in terms of how soon researchers might be able to get access to these data in any form?

Mike [00:27:14] So we have committed to launch our external request mechanism by the end of 2020. So we are working diligently right now to really kick the tires on the data that we have. There are a variety of research projects that we're doing with the data. And I view this as a way to really assess the nuances and the problems that come up when practically somebody goes to use the dataset itself. And so we're really just trying to do our best job right now in terms of providing the best quality data, the best documented data, the best

information prior to setting that launch date at the end of 2020. As part of our new funding agreement with the National Science Foundation, I'm excited to report that we'll be publishing a Title 13 proposal guide to help individuals who are not familiar with the RDC proposal process to learn how to apply to access this data, how to make the arguments on what would be data that's in support of Title 13 purposes. And so that's something that I think will really help people actually, not just to have the data theoretically available to them, but practically available to them. Another thing that is being funded through this grant from the National Science Foundation is that we'll be hosting a national fellowship competition where we will sponsor the access fees for individuals to work with CJARS data. And so people should stay tuned probably next year we'll be announcing that and we'll be taking applications for that program, which is exciting.

Mike [00:28:55] And then finally, because there is a lot of work that goes into working with data in the RDC, our plan is to put out a synthetic dataset that is modeled off of the information that's contained in CJARS and resembles the same essential structure that we have in CJARS so that individuals can work on exploring the data, analyzing it to a limited extent before going through the big push of submitting a proposal to the federal government to work with the data. You could also think that this might be a useful tool if you're teaching a class on the economics of crime and you want to — or in criminology or in sociology — and you want to give your students a readily available, well-documented dataset that they can go in and work with. The production of the synthetic dataset will probably take a bit longer than our launch of the external research mechanism. And the reason is because the information that we're collecting in CJARS is so detailed. You know, you think about linking up this person was charged in early 1992 with robbery, and then in 2001 they were charged with burglary, and then in 2011, they were released from prison. This creates very unique fingerprints on who somebody is and what their trajectory is over time. And our goal isn't to violate the privacy and the confidentiality of individuals who are involved in the criminal justice system. And so what we need to work on in the coming years is how to effectively maintain the relevant policy analysis variation in the data while still restricting the personally identifying variation in the data in the synthetic dataset that we'll be putting out. And so there will be a public data version that will be available, but that probably won't be until 2022, roughly speaking.

Jennifer [00:31:01] Just around the corner. OK. So one last question before we move on to this paper that we want to talk about, so where can people learn more about CJARS if they're interested and if they want to contribute in some way, especially if they want to give you data, how how should they get in touch?

Mike [00:31:20] So our website is cjars.isr.umich.edu, that's C-J-A-R-S dot I-S-R dot umich dot edu. We are working to provide kind of regular updates as we have them on our website, as well as when we have the technical documentation that will get published there as well. We also hire both undergraduate RAs, and we've recently started having postdocs work with CJARS. I'm really excited that one of your students, Jen, Brittany Street, will be joining us next year and we will be hiring another postdoc again. And so for all of those graduate students out there who are going on the market, please keep your eye on the CJARS job listing in JOE, we welcome your application. And we also welcome data donations. So, you know, one of the challenges that we run into is that as criminal justice agencies are bombarded with more and more requests, one of the responses that they've had is to really push up the price of the data to try to throttle this excess demand for access to their records. And so one example that we ran into this past summer is we've requested a set of data from the New Jersey court system that a colleague of mine had successfully acquired in June of last year for about 1,000 dollars, which we were quoted

roughly 82,000 dollars for the exact same set of data. And we were told that there was a new pricing model. And so, you know, very generously — and I want to give credit where public goods are being donated — Alex Moss and Amanda Agan donated their archive of New Jersey criminal court records to the CJARS project, which we're really excited about. And, you know, I think as we think about ways to speed up the acquisition of data from around the country, if there are people out there who have worked with administrative data, they have either FOIA-ed or have another form of data collection that permits the legal redistribution of their records to the criminal justice administrative records system. You know, we really welcome having that conversation about how we can give your data a new life. You know, you might be sick of it, but we would love to host it and help other people get excited about it as well.

Jennifer [00:33:51] That's awesome. OK. All right, well let's talk about the paper now. So the first paper that you produced based on these amazing new data is titled "Inequalities in U.S. Criminal Justice and Economic Outcomes." It's coauthored with Keith Finlay, your Co-PI at CJARS. My first question is just, you know, there are a zillion projects you guys could have worked on with these data. So why is this the first project you've invested in?

Mike [00:34:14] So one of the things that's really exciting about this project is that it addresses some very low hanging fruit issues that we were talking about earlier. That cumulative- we don't even have very good estimates on the cumulative exposure risk of things like felony convictions or the risk of being imprisoned. So we don't know- we don't have very good estimates on how many people will be- will receive a felony conviction by, say, age 25 or be in prison by age 25. And so what we can do with this is, one, develop really good new estimates on this using the wealth of data we've collected in CJARS. But we can take it a step further and say not just in the states covered by CJARS overall, but by where you were born or when you were born, or whether or not you're a black man or a white woman or a Hispanic woman, how did these rates change? And so there is kind of a fundamental gap in reporting on the criminal justice system that we're addressing with this project that we're really excited about. And then we can think about how this variation ties into a variety of social and economic phenomena that we've observed over the last half century in terms of changing income inequality and family structure in the United States and what potential role the criminal justice system might be playing in that process. And so I think that this is a really nice way to introduce CJARS to the research community overall.

Jennifer [00:35:50] And you had alluded a little bit earlier to, you know, people have tried in previous studies to estimate some of this stuff. And so so talk a little bit about those studies. What had we previously known about this topic?

Mike [00:36:03] Yes. So, you know, I think a lot of what we get in terms of the reporting on the criminal justice system, we can think about it coming in kind of three different forms. One is a large volume of Point-in-Time estimates from the federal statistical agencies. So, for instance, we have reports that say that 2.6 percent of adults in the United States were under some form of correctional supervision at year-end 2016. Or that, you know, in the middle of 2007, 52 percent of state inmates and 63 percent of federal inmates reported having minor children. And this accounted for about 2.3 percent of the resident population under 18. So these are Point-in-Time estimates, which really don't track the cumulative exposure to the criminal justice system. In the absence of having really solid data to estimate this, there are a couple different approaches that have emerged. One is our model-based estimates. And so we have life table approaches. You can think of the work by Becky Pettit and Bruce Western, which suggests that cumulative risk of time in prison is 3 percent for whites, 20 percent for blacks by their early 30s for those who were born in

1965 to 1969. Or there's work, a paper that I cite myself, Shannon et al 2017, which says that 8 percent of all adults currently have felony convictions and 33 percent of African-American adult men have felony convictions. And this is an approach that uses case filings and convictions, as well as an assumption about recidivism rates and how that would, in steady state generate a prevalence rate of felony convictions in the population. And so in both of these approaches, it's really dependent on whether or not the assumptions that are being satis- that are being made are satisfied in the true underlying data. And then finally, we have the longitudinal surveys. So again, Bruce Western and Sara McLanahan have reports on, for instance, from NLSY that the cumulative risk of incarceration is 4 percent for whites, 17 percent for blacks and 10 percent for Hispanics, for the cohort that's being studied in the NLSY79. Now, this is limited in the way that it's one estimate that's nationally representative and a single cohort. So we can't really dig in on which states might be driving these rates and how this might have evolved over time.

Mike [00:38:40] Now, on the latter half of what we're studying in terms of the relationship between changes in the criminal justice system and economic outcomes, there have been a number of efforts that have sought to explore how criminal justice system change, specifically the period of mass incarceration, how this has influenced a variety of outcomes like earnings and employment, family structure, changing marital patterns within the black community. And a large focus in this is often been trying to relate the contemporaneous incarceration rate to a variety of these economic and social phenomena. And that is likely both an important question to ask, but it's probably not the only question that's important to ask. If we think that something like a felony conviction — which there is a variety of audit studies that would suggest that having the mark of a felony conviction as opposed to having spent some time in prison or jail — is the predominant feature that might be impacting labor market outcomes, we need to go further in terms of integrating court records. But not just contemporaneous court records, cumulative court records so that we can see not just who got a felony conviction in 1993, but what share of the population up to 1993 has received a felony conviction and integrate that into the analysis.

Jennifer [00:40:10] So, as you mentioned, a major contribution of this paper is simply documenting this variation across place-by-cohort groups in the likelihood that individuals will wind up with a felony conviction or in prison at some point in their lives. And just to walk through that a little bit more directly: so you're defining a cohort here as anyone of the same race and sex who's born the same year, and then the place-by-cohort groups then are anyone in the same cohort who's born in the same commuting zone as the geographic unit you're looking at. So, for instance, all black men born in the Houston area in 1972 would all be in the same place-by-cohort group. Do I have that right?

Mike [00:40:49] That's exactly correct.

Jennifer [00:40:51] OK, great. And so you're going to interpret this variation as someone in that group's risk of a felony conviction or incarceration. So to start out, tell us more about why this type of variation is interesting. Like how should we think about the variation across place-by-cohort groups that you're documenting here? And what do you think explains this variation?

Mike [00:41:14] So I think that that's a really good question. I think that there just some really descriptive, new, important pieces of information that, at least to me, I find really exciting. So one is that we observe that the risk of getting a felony conviction by age 25 has roughly doubled from those born in 1960 to those born in 1980. You know, people who are roughly two decades apart in terms of their birth cohort from 4 percent roughly to about

8.5 percent. Now, later birth cohorts, those who are born after the 1980s, the risk has gone down slightly. And this might be reflective of recent efforts by states and localities to employ a variety of approaches of reforming the criminal justice system, you could think of the criminal justice reinvestment strategies for example. When we drill down within different demographic groups, we can say, for instance, that the average black male felony conviction rate is about 20 percent by age 25 within the states and cohorts that we cover. Whereas, the same average rate for white men is about 5 percentage points. And you can look across time and space. You could think about doing the mental exercise of taking everyone who was born in the locations that we're covering across all of the years and line them up and rank them according to their geographic risk of getting a felony conviction — not just geographic, based off of their place and time and their demographics, their risk of getting a felony conviction. And what you can see when you do this exercise is that the 10th percentile of risk of getting a felony conviction based off of when and where you're born for black men is around 10 percentage points. For white men, the 90th percentile is smaller. It's 8 percentage points. So there's a really dramatic difference that we're documenting within this project in terms of the exposure- a fundamental difference in terms of the exposure to the criminal justice system based off of the differences of these different demographic groups and how that risk has changed over time.

Mike [00:43:35] Now, black men in particular drive a lot of the nash- the overall variation that we see year by year. So the 90th percentile of risk that we measure in the dataset is about 27 percentage points. And so there's, you know- and that's only by age 25. So this would only go up over time. We limit it to twenty five so that we can have broader coverage within the geographies and years of birth. But there's quite a degree of variation that we're now newly documenting in this project. And increase in the risk of criminal justice contact, as documented in terms of felony conviction risk, is something that's really fundamentally new, at least in my mind. There's a lot that's been written about and discussed in terms of the experience in the United States of mass incarceration, as well as, for instance, the crack epidemic and the decline in criminal activity since the early 1980s- sorry, late 1980s, early 1990s. But this idea that individuals who are roughly around our age right now, maybe a bit older, are at the highest risk of having had a felony conviction, I think is something that's really new and really important to document.

Jennifer [00:45:00] Which demographic groups are you able to focus on using these data and how do they rank in terms of their exposure to the risk of criminal justice contact?

Mike [00:45:08] So so. So one of the things that's really important about using population level coverage is that we can actually speak to more narrow demographic groups than might be covered, for instance, in survey efforts. And so one thing that we can observe is that there's basically two pools of individuals. There are minority men, mainly- specifically black men, Hispanic men, and Native Americans who observe elevated risk, significantly elevated risk of contact with the criminal justice system, whether that be felony convictions or risk of imprisonment. Black men are at the highest end of risk. And then Hispanic men and Native American men are at an intermediate margin. At the low end are white men and Asian men and then most female groups which have very limited contact.

Mike [00:46:05] I want to get back to the question that you raised about what do I think explains this variation? I think that there are two kind of fundamental sources of input that go into whether or not somebody has a felony conviction. One are factors that influence the underlying risk of criminality. So we can think of things like whether or not you are exposed to lead as a child — which might impact the risk of aggressive behavior —, childhood education systems, the contemporaneous opportunities that you have. So if you

don't have a job or access to modes of sustaining yourself and your family, you might turn to crime. Or if there's a new drug that comes on the market that makes it attractive to use drugs in a way that you wouldn't have otherwise, that is something that could pull you into committing crimes when you wouldn't have otherwise. The second is a variation in how the criminal justice system responds to given criminal activity. And so this is influenced by policing behavior, prosecuting behavior, judicial behavior, variations in penal codes and sentencing guidelines. And so what we're focusing on in this project is using a research design, which as best as we understand, is eliminating a lot of the variation that stems from the factors that increase your criminality. And in the data that we're observing appears to be more driven by responses to state policies.

Jennifer [00:47:45] OK, so let's talk more about that. So, OK so you're you're documenting this variation. And then as the paper goes on, you're using it to estimate the effect of one's risk of a felony conviction or incarceration on a variety of economic outcomes, as you mentioned. And more specifically, what you're doing here is regressing those economic outcomes on that place-by-cohort risk factor that we just talked about. And your goal is to measure the causal effect of that risk factor. And even more ambitiously, as you were just alluding to, you argue that this is likely the causal effect of policy variations, such as changes in policing and three strikes laws and other criminal justice practices, that have changed substantially over time and across cities and states. So this leads to the question of what we need to believe in order to interpret the results in your paper as the causal effect of this sort of criminal justice system driven risk of a felony conviction or incarceration. So talk us through that a bit more. What other factors are we worried about here? You just mentioned things like lead or employment, there's stuff- and stuff like that. And then what checks do you do to convince yourselves that you're measuring the impact of policy variation and not variation in the underlying criminal behavior of the people in those cohorts?

Mike [00:48:58] Yeah, thanks, Jen. So what we're doing in this paper is we're looking at the correlation between these risk factors based off of when and where you were born and what demographic group you are and your economic outcomes and socioeconomic outcomes. There's no reason necessarily that the variation isn't fundamentally driven by both of these factors. Issues that drive criminal activity, as well as issues that drive the response to criminal activity by public agencies. What we employ are a fixed effects research design, where we're going to demean by your place of birth as well as your year of birth. And look at the residual variation in both the outcomes and the risk factor and look at the correlation. Now, even there, there's still strong arguments that could be made that there are dynamic factors within cities that could be violating an assumption that this is the causal effect of the criminal justice system on economic outcomes. That there's some other contemporaneous adjustment that's going on in terms of potentially, let's say, education policy or reorganization of cities in light of gentrification. And so what we do is look in the data at the degree to which we can assess these- whether or not these relationships are persistent after we've put in these fixed effects. So one thing that we do is we look at whether or not childhood home environment for the cohorts that are destined to go on to become high risk of having a felony conviction are systematically better or worse than the childhood environment of those who are not destined to go on to have high risks of felony conviction. So what we're doing is we're putting on the right hand side of the regression with fixed effects, the felony conviction risk. And then on the left hand side, whether or not your family owned your home, what your- the income of your parents were when you were between the ages of zero to six years old. Whether or not you started school on time or had started school between the ages of zero to six, which could be a measure of whether or not there had been some sort of mental delay that could be

associated with lead exposure, for example. So we link at the geographic level all of the future variation that will happen to this co- these different cohorts and their felony conviction risk to their individual level responses in the decennial censuses from when they were between the ages of zero to six. And we don't see any systematic correlation. So we can't- we don't find that the individuals who were from very poor backgrounds are the ones who go on to commit higher rates of felony convictions later on in life. So that's one thing that we do to explore.

Mike [00:52:10] Another set of analysis that we do is we look at the contemporaneous environment when somebody turns 18 and whether or not, conditional again on these fixed affects, if that is associated with higher or lower risk of a felony conviction. We look at two different major sources of variation. One are a variety of labor market statistics supported by the Local Area Unemployment Statistics program. So we look at, for instance, the labor force to population ratio or the prevailing unemployment rate when somebody turns 18. And we don't find any systematic correlation there as well after applying these fixed effects, which is surprising in some ways because you would think that the economic opportunity would be a major factor that contributes to the risk of having a felony conviction at some point between the ages of 18 and 25. The other variable that we look at are the prevailing crime rates as reported in the Uniform Crime Reports, the UCR. And whether or not that is systematically correlated with the cumulative exposure rate between ages 18 to 25. And interestingly, we don't find systematic correlation. We don't find any statistically significant correlation between the prevailing crime rate when somebody turns 18 and what their risk of getting a felony conviction or going to prison is by the time that they reach age 25. And we think that this prevailing crime rate is something that really helps us capture a variety of phenomena that would be fundamentally impossible to capture using the administrative records and other types of survey data we have available. We think that the crime rate is probably capturing the prevailing norms or the opportunities for crimes, which might be things that would be really hard to document by themselves.

Mike [00:54:03] And so these are the ways in which were seeking to rule out or basically quantify whether or not we think that the when we go to look at economic outcomes, it is factors from the first category — the influences that encourage criminality — that are driving the economic relationship that we'll be talking about later versus the policy variation. And we give two examples of policy variation in the paper. One is we look at the number of police officers that have been awarded to a local jurisdiction where an individual is born between the ages of 18 to 25 by the Community Orienting Oriented Policing Services Program, or COPS, which has awarded hundreds, if not thousands of police officers- funding for police officers around the country. And one observation that we have is that when there are more police officers that have been awarded to a local jurisdiction, this increases the likelihood of having a felony conviction specifically for black men. We find a relationship for other demographic groups, but it's definitely most pronounced for black men in the data. And so this is one example of a potential local policy or practice that when there are more police officers on the ground, they're better able to convert criminal offenses into a criminal conviction. A second policy that we look at is whether or not things like the habitual offender laws or three strikes laws are related to this variation that we document. And so we can look at the percent of years between the ages of 18 and 25 for an individual, whether or not that influences the likelihood that they end up with a felony conviction or their probability of being incarcerated. And what we find is that interestingly, or perhaps not surprisingly, while habitual offender laws or three strikes laws really aren't geared around increasing the likelihood of having a felony conviction, it really does dramatically increase the likelihood that black men are incarcerated. And so this is one of

the ways in which we see policy in practice playing out in a way that really ties in with how the law is structured.

Jennifer [00:56:26] OK, so let's get to the good stuff. Walk us through your first set of economic results, the impact of increasing felony conviction risk on economic behavior.

Mike [00:56:37] So one thing that I want to address really early is that when we're running these regressions of looking at economic outcomes and how they're impacted by felony conviction risk, this is going to quantify a somewhat complex relationship. This is not just the impact of a felony conviction by itself, but it's also going to be other factors that might go hand-in-hand with it. Like, for instance, if there were reforms to the juvenile justice system, this is probably going to be picking up on that relationship as well. But then also perhaps more complex ideas like if there's statistical discrimination. So if you are born in a cohort where there is a high risk of having a felony conviction, but you yourself didn't have a felony conviction, you still might be harmed by that. So what we observe is that when individuals are born with a high risk of a felony conviction by age 25, it decreases their likelihood of completing high school. It lowers their employment rates systematically. It decreases their wage income so that it looks like the risk of having a felony conviction has really hampered the production and accumulation of human capital within these cohorts in a way that fundamentally limits their self-sufficiency.

Jennifer [00:58:01] And you in the paper go on to focus on the spillover effects, particularly of black male felony conviction rates on other demographic groups. So talk a little bit about kind of the intuition or idea behind this exercise. And what do you find?

Mike [00:58:18] So the idea is that the overall variation that we observe in the data is really largely driven by the disparate treatment of black men across place and time, with the caveat that we already discussed that other minority groups are also being negatively affected by the criminal justice system. The impact, though, of felony conviction doesn't necessarily stay with just the individual who receives a felony conviction. There's been a lot of discussion about how criminal justice policy over the last 50 years has transformed the family structure of black households and could have labor market spillovers within low skill groups. And so we do a couple of analyses to look at how the risk of a black male felony conviction by age 25 impacts both black men, but then other demographic groups. What we find is a pattern where black men- as black men are being pushed out of the labor market, whether it's because they're physically or effectively incapacitated by incarceration or a felony conviction, we see an increase in participation among among black women. And so this is a pattern that's been documented before. But we're really bringing the importance of felony convictions and felony conviction risk to the forefront in this conversation. And so we can see women, black women have increased labor market participation rates; yet overall, the household income declines for both black men and black women. And what could explain this? It's- we see very blatantly in the data essentially the destruction of two parent households in black- in black communities across the United States when there is an elevated risk of felony convictions. There is a decrease in total adults in the household for black men and black women, and there does appear to be some suggestive evidence that there might be an increase in minors for black women.

Mike [01:00:23] Now, we also documents spillovers where nonblack men appear to benefit from the exodus of black men out of the labor market. And I don't think that this necessarily has to be anything nefarious going on, but it makes sense just in terms of a general equilibrium adjustment that as one demographic group is having their labor supply constrained, labor demand will go and seek other demographic groups to employ. And so

we can observe patterns where it appears that nonblack men are benefiting as black men are being restricted from the labor market or excluded from the labor market effectively.

Mike [01:01:09] One of the things that we didn't have a chance to talk about, but I think is quite interesting, is we can compare the relative influence of the probability of imprisonment versus felony conviction. And when we do these comparisons across the different types of analysis that we've talked about, it does appear that felony conviction risk is the predominant driver of the economic relationships that we're observing, rather than what's been the predominant narrative to date, that mass incarceration has been the cause of economic decline and challenges and transformations within family structure, within minority communities. And so I think that this is an important thing to raise and that without the data infrastructure that we have in place in this project, it would be fundamentally impossible to have documented this relationship.

Jennifer [01:01:58] So what are the policy implications of this work? What should policymakers take away in terms of how to improve the criminal justice system?

Mike [01:02:07] One thing that I find a bit challenging in this project is that, you know, we are documenting potentially major changes that have happened within the criminal justice system and how they relate to major changes in social and economic outcomes within American communities. We don't- we are not yet at the point where we can say this specific policy needs to be implemented or changed in order to help us achieve our goals of having safe and productive communities. But one thing that I do think is worth addressing is that what we're documenting is these changes in criminal justice policy and practice have really fundamentally transformed the life experiences of certain cohorts in a way that even if we, for instance, start to reduce the incarceration rate today, it's not necessarily clear that that will help all of the individuals who to date now have felony convictions and no longer have kind of the two parent family and stable income that they might have otherwise gotten to experience because of when and where they were born. And so I think it's really important to think about not just how can we change the practice today, but what do we need to do to support the individuals who have gone through prior practice in the criminal justice system to help them get back on to a better track in terms of having opportunities for self-sufficiency and an employment opportunities to be happy and productive in the United States.

Jennifer [01:03:45] So what's the research frontier here? What are the next big questions you'll be thinking about in the years ahead?

Mike [01:03:51] So I think that there are a lot of new opportunities that are opening up with the possibility of integrating individual level administrative data on the criminal justice system with a large volume of data on employment and earnings, family structure, a whole host of things. The things that I think are really exciting are the intersections between labor markets and criminal justice practice and policy. For instance, I'm working on a project right now that's seeking to quantify how much of the labor market is legally off limits to individuals because of occupational licensing restrictions and other types of industry limitations. I'm working on projects that look at spillovers between labor market demand and criminal justice policy and how that sets people up to engage in criminal activity or not. I think that there's just a whole new generation of research that is really exciting right now that, you know, whether or not it's something that I'm working on or something that other people are going to bring to the table. I'm just really looking forward to what's going to be produced in the next five to ten years.

Jennifer [01:05:02] My guest today is Mike Mueller-Smith from the University of Michigan. Mike, thanks so much for doing this.

Mike [01:05:07] Thanks so much, Jen.

Jennifer [01:05:14] You can find links to all the research we discussed today on our website, probablecausation.com. You can also subscribe to the show there or wherever you get your podcasts to make sure you don't miss a single episode. Big thanks to Emergent Ventures for supporting the show. Our sound engineer is Caroline Hockenbury. Our music is by Werner, and our logo is designed by Carrie Throckmorton. Thanks for listening and I'll talk to you in two weeks.