Dr. Orchowsky: Good afternoon everyone. My name is Stan Orchowsky and I am the research director for the Justice Research and Statistics Association, and it’s my pleasure to welcome you this afternoon to our webinar, Developmental Estimates of Subnational Crime Rates based on the National Crime Victimization Survey with Lynn Langton and Bob Fay. We are all here in snowy DC, and we’re glad you can join us.

Dr. Orchowsky: Before we go any further, I want to thank our friends at the Bureau of Justice statistics for helping to make this webinar possible. Before we get started, I just want to cover a couple of logistical items. We will be recording today’s session for future playback. The link to the recording will be posted on JRSA’s website and email to attendees.

Dr. Orchowsky: Today’s webinar is being audio cast via both the speakers on your computer and teleconference. You have speakers on your computer and headphones and are not a presenter. We would recommend you listening to the webinar using your computer speakers or headphones. To access the audio conference, select audio from the top menu bar, and then select audio conference.

Dr. Orchowsky: Once the audio conference window appears, you can view the teleconference call information or join the audio conference via your computer. If you have questions for the presenters or would like to communicate with JRSA’s staff, we encourage you to submit them using the chat feature on the right side of your screen. Please select host from the dropdown menu next to the text box and you can send those questions in anytime you like.

Dr. Orchowsky: The session is scheduled for an hour and a half today, and so we should end promptly at 3:30 Eastern if not earlier. If you have technical difficulties or get disconnected during the session, you can reconnect to the session using the same link that you use to join initially. You can also call WebEx tech support at 1-866-229-3239.

Dr. Orchowsky: In the last five minutes of today’s webinar, we will ask you to complete a short survey. The information you provide will help us to plan and improve future webinars and meet our reporting requirements. All right, so there is a small but enthusiastic group of us who have been waiting with bated breath for this moment to learn about the sub national estimates that have been based on the NCVS that have been released by BJS. We’ve been hearing that this is coming for a long time and we’re very excited to finally see those happening, and all of those of you out there in the stage should be very excited as well. By the end of today you’ll know why you should be excited.

Dr. Orchowsky: So I’m going to turn this over to our presenters. Let me tell you a little bit about them first, our first presenter is Dr. Lynn Langton, who’s the senior statistician in the victimization unit at US Department of Justice Bureau of Justice Statistics. Dr. Langton has published nearly 20 reports based on data from the National
Crime Victimization Survey on topics ranging from hate crime to sexual victimization, identity theft and interactions between the police and the public.

Dr. Orchowsky: She oversees BJS' National Victimization Statistical Support Program, used to enhance the NCVS methodological research program and assist in the development of the NCVS Subnational Estimation Program. She's also heavily involved in the current NCVS instrument redesign project and the day to day operations of the NCVS. And she got her doctorate in criminology law and society from the University of Florida but she's okay anyway.

Dr. Orchowsky: Dr. Fay is a senior scientist and senior statistical fellow at Westat, where his work since 2008. Bob is also a research professor in the joint program for survey methodology at the University of Maryland. Previously, he worked at the US Census Bureau where he held the position of senior mathematical statistician since 1990. Over his career, he has worked in several areas of survey design, estimation and analysis, and he received his PhD from the University of Chicago about which I have no jokes at all.

Dr. Orchowsky: With that, we will thank both of our presenters for being here and we'll turn it over to Lynn.

Dr. Langton: All right, thank you for that introduction, Stan, despite the job. So we have quite a bit of material to cover today about the NCVS subnational program in general, and then more specifically about the recent R&D report released by BJS and authored by Dr. Fay and his colleague, that presents victimization rates for all 50 states, DC, and other select counties and metropolitan statistical areas. Without further ado, I'm going to go ahead and jump in.

Dr. Langton: Quick outline of this presentation to give you a roadmap of where we're going. I'll start by giving a brief background on the National Crime Victimization Survey for those who aren't familiar, then I'll talk about the development of NCVS subnational program about the utility of having state and local NCVS estimates and about the different approaches that BJS is taking to generate these estimates, including the approach that Dr. Fay is going to talk about in more detail, the development of model based or small area estimates.

Dr. Langton: I'll turn it over to him to talk more about how it's worked for started his methodology and then some of the key findings that are presented in the new report. And we'll wrap it all up by talking about how these developmental estimates can be used, where you can find the report and where you can find the data tables to conduct your own analyses.

Dr. Langton: Okay, so the NCVS is an omnibus crime survey, meaning it collects data on a wide range of violent property crimes. These are all non-fatal crimes since the data are collected through interviews. The survey provides annual counts and
rates of victimization and it's used to measure changes in the crime rate from one year to the next and then over a more extended period of time.

Dr. Langton: It was created back in the 1970s in response to the understanding that many crimes go unreported to the police, and therefore we have what we call the dark figure of crime. So beyond collecting data on crimes both reported and not reported to the police. Another benefit of the NCVS design is that it involves a standardized interview in data collection protocol nationwide. This is in contrast to what we have with official crime statistics where you may have variations in statute from one place to the next. And then on a national level, variations in what and how often local police departments report to the FBI, UCR.

Dr. Langton: Also, the NCVS is an incident and attribute based collection, it's incident based in that we collect information on the characteristics of up to six victimizations during the reference period. So we have data on the nature of and responses to specific incidents of crime. And it's attribute based and that we generally don’t ask respondents, for example, whether they were robbed because we know that these terms are tough and someone may have had their house broken into, but they say they were robbed rather than burgled.

Dr. Langton: So instead, we ask them to tell us about the incident and then based on the attributes they provide, we classify the type of crime as a robbery or a burglary. NCVS data are collected for BJS by the US Census Bureau. The Census Bureau selects a nationally representative sample of households and then sends field representatives out interview all persons aged 12 or older in the sample households.

Dr. Langton: The interviews are conducted with the same household every six months over a three and a half year period for a total of seven interviews. The first is typically conducted in person and subsequent interviews are conducted over the phone. And we enjoy a very high response rates in the 80 to 90 percent range, which anyone in survey research can appreciate.

Dr. Langton: The key takeaway from this slide is really that the NCVS was designed to generate national estimates of victimization based on a nationally representative sample. Now, in the past five years or so, BJS has been looking to expand the NCVS from being able to generate only national estimates of victimization to being able to produce subnational, so state and local estimates of crime.

Dr. Langton: We see a lot of value in having subnational victimization estimates. For one, the NCVS can go beyond police statistics and begin to get at the dark figure of crime. And this gives a better sense of who's at risk for victimization, what crime problems exist that may not be coming to the attention of the police, and so on. These estimates also have a lot of utility for resource allocation. Within the DOJ
for instance, we have grant programs, JAG, the Justice Assistance Grants being one.

Dr. Langton: For which funding is initially allocated to state based on UCR crime data. So you could have two states with similar reported rates of crime that then get similar amounts of funding. But the problem is that one has a much higher rate of reported crime, unreported crime, and potentially that state really needs more resources than the other state.

Dr. Langton: Subnational estimates can also be used to examine local crime policies and local interventions. Again, the NCVS is a standardized instrument and data collection process. So this allows for the comparison of places, be at cities or states to see how they differ and how these differences contribute to differences in crime rates.

Dr. Langton: Also, if you have a baseline measure of crime in a particular area and then a new policy or program is implemented, because you have a standardized measure you can isolate the effect of that program or policy on changes in crime rates over time.

Dr. Langton: Then finally, another value of the subnational victimization estimates is that NCVS data at the local level can be linked up with other local level sources of data on economic and social characteristics, on local law enforcement resources and policies, on victim service provision within the area. And of course, official police statistics. And this allows for the development of a much richer understanding of the criminal justice system within that area as well as gaps and shortcomings in the functioning of the system.

Dr. Langton: So there's a lot of justification for producing subnational NCVS estimates. Why don't we just do it? Well, of course it's never that easy because the NCVS sample is selected to be nationally representative. It's not necessarily representative of particular states or counties. So in New York for example, we may have lots of sample in New York City, but no sample in upstate New York. And the composition of people living in upstate New York may be very different from the composition of people in Manhattan.

Dr. Langton: Additionally, we may have a lot of sample in New York state, but then we have no sample or very little sample in other states. So consequently we can't say much about subnational rates of crime with the current NCVS design.

Dr. Langton: So BJS is taking a multipronged approach to developing a subnational estimation program. I'll talk in a bit more detail about the course to approaches, direct estimation and the low cost companion study. And then I'll turn it over to Dr. Fay to talk about the recent batch of estimates he generated using the last approach.
The first approach to generating subnational NCVS estimates is to boost and reallocate the core sample, so that we have enough sample in particular areas to generate reliable estimates and that sample is distributed such that those estimates are representative of the population of that area.

So we’re literally just adding enough sample cases so that we have the right locations to be able to generate direct state or MSA level estimates from the survey. And this is being done in large states where we already have quite a bit of sample to begin with. The advantages of this approach, in theory, we have all the variables from the core NCVS and the supplements available at the state level in states that received the boost.

Now, I say in theory, and I’ll get into that in a second. Another advantage is that the estimates are based on direct survey data, so they’re easy to understand and relatively easy to replicate, and because we have the full content of the NCVS available at the subnational level, BJS can provide state and local estimates that are not currently available through official statistics.

For example, we collect data on victim-offender relationship and this allows us to put out direct state level estimates of intimate partner violence, which is something that you can’t get through the UCR. Data on the proportion of crimes reported to the police is another big one, another big advantage.

In terms of limitations, cost is a big one here. Boosting the sample isn’t cheap, especially since a lot of where we’ve boosted is in fairly rural areas and the interviewers have to drive further to conduct the interviews and we also may not be able to share interviewers with other major federal surveys if those surveys aren’t in those same areas as well.

At this point, we think that for most states in MSAs where we’re boosting, we’ll need about three years of data to be able to generate estimates of violent crime with relative standard errors around 10 percent. If something were to happen with BJS funding two years in, that could also have implications for the ability to generate direct estimates.

And then another limitation with direct estimates is that we have to be concerned with disclosure. So we would likely not put out a public use data file with state identifiers. These files would be restricted use where researchers would have to go to the Census Bureau, Research Data Centers, present a proposal to conduct statistical work using the data and basically swear on their firstborn child that they’ll maintain the confidentiality of the data.

Then one other limitation that isn’t shown here is that with the direct boost, we’re not able to go back in time to examine trends and changes over time. So any trend analysis based on the direct boost estimates will essentially start with 2013.
Dr. Langton: In July of 2013, BJS and the Census Bureau began a pilot test of boosting the sample in the 11 largest states. With this pilot boost, we are able to test assumptions about costs, data quality, design effects, field operations, variability in victimization rates, relative standard errors, and so on. The pilots has lasted for two and a half years and just ended at the end of December, and we hope to rehab a report on the findings from that by the end of this year.

Dr. Langton: But in the meantime, we've been assessing the data as we go and making necessary adjustments. And then beginning in January we began implementing the boost on a full scale in the 22 states that you see highlighted in this map. These 22 states make up about 80 percent of the population of the US and also of the crime.

Dr. Langton: Now, in addition to the pilot boost and now the implementation of the 22 state boost to collect direct estimates, another approach that we're taking to generate subnational estimates is a low cost companion survey. This is separate from the core NCVS and it's administered in particular area to produce local area estimates that are comparable to the NCVS. So basically it's a self-administered mailed survey that collects information on victimization, but then also on a perception gap and satisfaction with the police, perceptions of neighborhood safety and other questions along these lines about disorder and fear that we're calling non-crime questions.

Dr. Langton: The companion survey approach was piloted in 2014 and originally the idea was to try to blend the estimates with the core NCVS, but because of how complex the NCVS is, that proved to be virtually impossible. So the emphasis shifted to focusing on a survey that would parallel the NCVS and could be administered in multiple locations and over time to compare differences say pre and post to certain treatment in one area.

Dr. Langton: In phase two, a field test is being conducted using two different versions of this self-administered mail survey. The first is a simpler survey that can be used to generate prevalence estimates while the other is more complex and is intended to capture incident level information.

Dr. Langton: The field test is being administered both in 2015 and 2016 in 40 of the largest MSAs, and part of the focus is on whether we can use this approach to detect changes over time. So can we pick up changes in the crime rate from 2015 to 2016 in these areas? And then also can we detect differences in victimization across place?

Dr. Langton: And another key feature that's being tested is the feasibility of generating subsidy level estimates. So the ability to look at within city differences. In three of the MSAs, Chicago, LA, and Philly, we oversampled and we stratified the sample across police districts to see whether we can compare victimization rates and measures of police legitimacy across these areas.
Developmental Estimates of Subnational Crime Rates Based on the National Crime Victimization Survey
January 27, 2016

Dr. Langton: Wave one of the field test just wrapped up at the end of December with a response rate around 45 percent overall. So pretty good for a mailed survey and we're doing some analysis of the data now and wave two will go out later this year.

Dr. Langton: So that covers two of the three approaches, and with that I'm going to turn it over to Dr. Fay to talk about the third approach to generating subnational estimates.

Dr. Fay: Thank you. So the third approach is more labor intensive and less capital intensive. To set the history here, in 2008 of panel of National Academy of sciences recommended to BJS that they consider among other improvements to the NCVS, development of small area estimates. Now, many of you may not have encountered this sort of estimation before, I'll describe it as a methodology that's different from the direct estimates than the bread and butter of the NCVS.

Dr. Fay: In the first phase, there was worked on actually at Westat doing a broad review of small area estimation possibilities. My colleagues Cantor, Krenzke, Stukel, and Rizzo submitted a report and that's available on the BJS website. But I'm going to speak about the second phase, which was not just conceptual but to actually produce the estimates. We're going to show you and under a grant to me and to my colleague Mamadou Diallo, we developed the estimates I'm going to talk about. Most of our research focused on states, but in the end of the project we're able to produce four large counties and large metropolitan areas.

Dr. Fay: Dr. Langton has already mentioned the constraints of working at the Census Bureau. This work was carried at the Census Bureau itself and use the internal files with the geographic information. We're both had special sworn status and then any numerical results we wanted to take out of the Census Bureau. We submitted and we're approved by the Census Bureau Disclosure Review Board. We started really the work just in August, 2011 and over a period of time have been reporting progress papers that proceedings the first was that the federal committee on statistical methodology in that January in 2012.

Dr. Fay: Now, over the course of the project we really enjoyed collaboration with the BJS staff to get directions to appreciate what might be needed or most valuable. And so there had been a process of showing disclosure review board approved results to BJS in a preliminary where by before we've come out with this final set. Among the collaborations I've mentioned that one of the papers Dr. Michael Planty of BJS was one of our collaborators of paper we gave in 2013.

Dr. Fay: So where we got to is a technical report that's now come out, and I give the link here. They've already seen the title. The webpage also has both these links to the report and also a link to Excel spreadsheets with the complete set of estimates that we've produced. We have documented our methods besides
these papers directed at a statistical audience of the methods are embedded in
the software which, some which is publicly available through our, and then
others to BJS.

Dr. Fay:

So I’m going to talk about what we tried to estimate. One of the things about
small area estimates is that there are only produced for certain characteristics
as opposed to being able to cross classify and dig into data. So the
characteristics I’m going to show you are both violent crime by type, and one of
the, just again, to stress limitations of the approach to get this much geographic
detail. We were forced to combine rape and sexual assault with aggravated
assault.

Dr. Fay:

Not because they're equivalent, but only because the first category of crime was
so rare when we couldn't even model it. As reported in NCVS I should add. But
the NCVS reports were too thin to produce this estate level model. Robbery is
another category and the NCVS measure simple assault, that's the largest single
source of violent crime in NCVS, and that was the other component we model.

Dr. Fay:

We could also break down violent crime by relationship to the perpetrator. This
was a second goal that was added during the course of the project. So intimate
partner violence has been a key interest of the research community and some
reports by BJS nationally about this type of violence.

Dr. Fay:

Secondly, we separated out violence committed by strangers on, or more
strangers, and then had to group together all other types of crime that could
both be other relatives, friends, acquaintances, and a small category of
unknown. The other characteristic we estimated was property crime and not
broken down by type. These three types of burglary, larceny, any type of theft,
except motor vehicles theft as a third category, motor vehicles theft was
somewhat in and hard to model.

Dr. Fay:

Property crime in total, however, is much more frequently reported in NCVS
that is violent crime. It was an easier problem, but perhaps the less interesting
one. Additionally, what we did is to try to provide or do are providing estimates
for 50 states in the district of Columbia, 65 large counties and for 51 large
metropolitan areas or CBSA's, and these are only the largest ones.

Dr. Fay:

We didn't dig down further as you may listen to the talk, you'll maybe
appreciate we weren't, by the time we got to counties of metropolitan areas,
didn't want to model areas that we might not even have any sample in NCVS.
The other thing is we're estimating for time periods here, so we've formed the
estimates into three year averages even though the model was originally
developed for single years of age. But these three-year averages we shall
provide more stability, and it's the same strategy unless it's been revised that
will be used for publishing estimates in the boosted states at least I understand
the three-year averages will be offered there.
Now, I think we've learned in the US that users can or are willing to work with three and five year estimates given the American Community Survey experience. So, this notion of averaging is distress, we're looking more at longer term trends and very short term changes in the level of crime. For the estimates we've produced, we consider a window of 15 years' worth of NCVS data and produce 13 sets of overlapping three-year estimates.

We were able to estimate states up to 2013 but because of availability, we cut off our estimation of 2012 for the county and metro models we've produced.

So I'm gonna give you four slides of sort of a crash course on small area estimates to say that the characteristics of the estimates we've produced are like in some ways other small area estimation projects that have occurred. Most surveys and NCVS, other federal surveys, practically all of the products would produce estimates we call direct estimates. And typically you can obtain the estimates by adding up weighted estimates on a data file, or doing relatively simple operations to estimates that you have from the data file.

The direct estimates for an area or domain used data from the domain only, so that estimate for New York state would depend on data that were collected in New York. But as a consequence, the reliability of these estimates depends on how big the sample sizes are in the domain and question.

Now, small area estimates in contrast, I know I'll just mention as we did, even though it was building on other ideas of collaborated with a late colleague, Roger Herriot in '79, a few years before then produced these estimates and then we published an article about applying these ideas to the 1970 census long form data. And so that was a sort of large initial application. It was only for very small places where the long form data would be unreliable. Now, there are several efforts going on producing specialized small areas statistics in a number of federal surveys.

And the characteristics of these through models, they use data from other domains or sometimes other time periods to form estimates. So the estimate for New York will not in application might depend on what's observed in other states as well as New York. And as a group, they all depend one way or another on some implicit or explicit statistical models, much more so than direct estimates.

And the idea is to produce or improve the reliability of the estimates, but in a sense, because it's a model, it's maybe doing well on average, predict general patterns among the domains.

Now, the key ideas also to incorporate auxiliary data from external sources, without auxiliary data these small areas of estimates don't do too much. Now, but paradoxically, because especially if you have the auxiliary data, then you can
even produce estimates through small area models for areas that don't have any sample at all or have very small sample. And very typically, you get much smoother estimates out of a small area model then the variability you will see with sample estimates based on moderate to small samples, where much of the variability is to simply to sampling error.

Dr. Fay: Just to give you a kind of a metaphor of this. This meant to be informal term borrowing strength is used a lot and it just says that these models borrow strength from other observations across the survey to make predictions for the domains of interest. Also the term has been used indirect and these models can either be for a single point in time and use information across domains, the model. There are other models and applications that can look across time to help form the estimate. And you can combine both ideas and look across domains and time. And that's what the estimates were presenting do.

Dr. Fay: That was general small area estimation, an overview, but what's specific about this application? Well, again, as I mentioned, we're using the geographic information on the Census Bureau's internal NCVS files and we are able there and behind the firewall to compute the direct estimates for states or the other domains, but their input to the model. And we're also use auxiliary data from the FBI uniform crime reports by type of crime. Now, these are the most readily available statistics they published from this an annual series, the key statistics we use are forcible rape, robbery in terms of violent crime, then burglary, larceny and motor vehicle theft this property crime.

Dr. Fay: Now, we tried and found no gain really from including aggravated assault, which is a key component of the UCRs violent crime reports. But statistically at an area level and trying different levels of analysis found very little correlation between the two series. I'm speaking only as a statistician and mentioned this amongst subject matter experts to see if we could somehow figure out more about this.

Dr. Fay: But moving along, I've already mentioned we're borrowing across the main some time. If you want a note to technical details, we produced a variant of the value model which was published over two decades ago. Also, I can provide a citation to John Rao's book in 2003 years, collaboration with Isabel Molina, the second edition, Small Area Estimation, which is kind of the go-to book for statisticians in this field.

Dr. Fay: We've not wandered very far from standard methodology, but we tweaked it a little bit. Now, many of these I see, in this small area estimators form a composite of predicting from a model such as regression model, but also waiting in some amount of what the direct estimate is. In this case, for example, the estimate from New York for a three-year period is coming both from what's observed in New York and other times, especially recent times and through regression coefficients that are modeled based not only on New York but all the other states relating the UCR data to the sample estimates.
Developmental Estimates of Subnational Crime Rates Based on the National Crime Victimization Survey
January 27, 2016

Dr. Fay: So I'm going to show you some results for a while before I talked about their reliability, but these are some of the primary results that I'm giving you. These are all coming out of the report I mentioned was available, I've just changed them from portrait to landscape if I need to make that disclosure, but otherwise they're the same figures. I'm showing in blue, I used blue to indicate lower than average national compared to the National Crime Rates and red, a red scale to indicate elevated or above the national average.

Dr. Fay: On the left. It's hard to say but on the left are the estimates we've produced, these small area estimates I stayed and on the right ... and some contrast or what the UCR. Now remember, the two sources are measuring different things, the simple assault is being reflected in the estimates on the left but not the right.

Dr. Fay: I don't want to make too many contrasts because there are multiple reasons for differences, but it is kind of striking. There's somewhat more crime being showing up in the northwest states, especially Washington, Oregon, compared to the portrayal in the UCR. The South Atlantic states along the atlantic starting, well, Virginia is low in either series but moving down the coast. There's a whole series of states estimated to be low, whereas the UCR gives us a different picture.

Dr. Fay: And one thing I've noticed a long time ago is Tennessee which is sometimes commented on is having a very high rate of violent crime in the UCR does not show up that way at all on these estimates. So even though the model is using the UCR data, it's also using direct data from the NCVS averaged over time to form these estimates.

Dr. Fay: Now, let's see, this is just for the most of recently available three year period. This is the picture for property crime, somewhat similar in terms of the contrast, especially both this higher crime in the northwest, then we might've thought from the UCR, and lower crime along the South Atlantic states.

Dr. Fay: Now also, I'm showing figures what happens if you average these small area estimates over a period of time in contrast the ... and that's the full 15 year period here in contrast to what the UCR looks like over that whole period. And the story is somewhat similar, somewhat to a little different. And similarly for property crime, the shift a bit to the northwest, even and west, and this picture compared to the South Atlantic states. Rape and aggravated assaults, I'll show you some pictures now here again, we're comparing NCVS on the left, and UCR on the right.

Dr. Fay: Maybe, I'm only commenting what you can see. This again, shows some differences with the South Atlantic states showing up in the northwest. That's kind of one of the most persistent differences. Robbery, now I'll pause on
robbery because with this model, the UCR and the NCVS were closest in alignment of all the violent crimes.

Dr. Fay: So the regression coefficient on robbery was quite substantial and the patterns look fairly similar. Still, a little bit of difference in the south central states or South Atlantic states. Simple assault, now I've changed, there's no longer UCR on the right. I put two graphs up here, but the left is just the most recent period on the right you'll say what happens if we average these small area estimates over the entire period. So there's some contrast between the answers for both. It was only recently the Pennsylvania shows up as a high crime state with simple assaults and that's now over a longer period of time.

Dr. Fay: We tried our best and hear what shows up in terms of intimate partner violence and what the model predicts. Again, I'm predicting this, the recent period, showing the recent period on the lefthand side. The longer term average on the right hand side. I don't have, being always statistician, I have an interpretation for what we're showing up here.

Dr. Fay: I do have the national estimate showed an intimate partner violence is frequently have the type of assault, it is most frequently reported as some form of assault. But this is another feature of the small area estimates is the data are too thin to cross classify. So we've told the story two ways, we've told it by type of crime and here we can tell it by the perpetrator, but there are no estimates of the cross classification by state.

Dr. Fay: Here's violence by strangers. Again, there's no UCR data to show as a reference and we get similar pictures over the short term and the longer term average. I wanted to show that we're also providing in the report, all the graphs state-by-state I'm picking on Massachusetts that's been shown before in a previous talk. These graphs show a trend over time so we take off, first of all violent crime and in the graph there is again a comparison to the UCR.

Dr. Fay: Now, if it goes to the UCR first, Massachusetts, this entire period has been a very average state with respect to violent crime. Because of an apparent dip in violent crime during the early 2000s we see this portrayed as a depth and simple assault is seems to be the biggest contribution to that, but then the estimates come back together again.

Dr. Fay: Another, we can also offer violent crimes by the perpetrator. Now, this total crime and block is still the same values as in the preceding graph, but the type of crime here, you'll see that the intimate partner violence as the smallest of the three components of blue at the bottom, it's really quite average, and there's not too much to distinguish these graphs except the difference maybe seems to come out of maybe others more than impossibly, a little bit from stranger violence during that period where the estimates small area estimates seem to dip.
Dr. Fay: Here's the results for property crime in Massachusetts. These are on page B-64 of the report. You'll see the UCR really points to Massachusetts as having a much lower property crime rate than a nation as a whole. And especially due to theft or larceny. And actually our small area estimates presented fairly similar picture.

Dr. Fay: So in these cases, although I have other cases, drawing a contrast between the UCR and NCVS small area estimates piece. In this case, you would say it's a fairly similar picture. Of course we're seeing here the kind of dropping crime that showed up in the NCVS more so than the UCR during this early period of 2000 and some years before that.

Dr. Fay: Let me say a few words on what we did to evaluate these estimates. The first thing we did was a more empirical, I hope more intuitive kind of comparison I'm going to show you. This was very constrained by disclosure review board requirements, in my own concern not to ask too much. So what we did is to average, first of all, a preliminary version of the small area model over time and just for total violent crime and property crime.

Dr. Fay: We also did the direct estimates by state and average them over 15 years. Now by averaging survey estimates over that longer period of time and showing only rates, there was essentially very limited chance of any kind of disclosure you could ... The fact that crime occurs in every state tells us nothing about any individual. So I maybe push this as hard, but I can show you what I asked for. I do think that disclosure review board generally would not allow publication of annual estimates from small states.

Dr. Fay: The presumption and I assume this has already been asked to boosted states, there'll be some things coming out. But as Dr. Langton mentioned, probably not micro data files for public use. So this was a special requests they approved and I also drew a cutoff at two million. Again, just to be sure that the sample in a small state wouldn't cause a concern.

Dr. Fay: So then I'll show you the same kind of picture we've been looking at by put first in gray, all the states I've included. Now in hindsight, maybe I could've gone down to one million because we will tell you, I worry even more about states below one million and their accuracy is below two million. But just to be safe, the grey states are essentially unknown, although I could've shown you the averages by less than gray too, to make the comparison visual comparison easier.

Dr. Fay: In the other states, I wanted to show you that very close agreement between taking an average of the NCVS over 15 years and what's the model estimates did over 15 years. Now, this is kind of inherent in the modeling. So the model is trying to use the sample data that extended can and over 15-year period any
place where the sample data's being coming reliable, it's being built into the estimate over time.

Dr. Fay: So, I hope this graph shows you the estimates we're producing are made out of just whole cloth and totally dreamed up. But have correspond to what the survey's been saying and even are somewhat high estimates of crime in Washington and Oregon are matched equally or somewhat more so by the direct sample data. There's actually a table of these results in the report as well, if you want to look at this specific figures.

Dr. Fay: So that's one evaluation. The other measure that's often are almost always now produced for a small area estimation products is to produce the mean square error or its square root. The root mean square error, and the RMSEs are easier to show people and work with. So, we produced those for all the estimates and they occur, they're shown in all the spreadsheets.

Dr. Fay: So we've looked at them in order to get an idea how accurate these are, if I could tell you that these estimates are so accurate, the boosting is unnecessary, then I'd be proud as a statistician, but that's not the case at all. The estimates have more error than will be obtained from the direct estimates that will be obtained at the state level from the boosting and by quite a lot as opposed to maybe getting a 10 percent error in boosted states, we're seeing something closer to 20 percent.

Dr. Fay: And yet we're still seeing some pictures, some picture that's been hidden in the NCVS data, at least from public view. So we have these available. Now, we've seen combat as the statistician to the root-mean-square error, and the very smallest states are quite large. With states below a million population, I can tell you even a statistician looking at the design from the outside with no privileged information could say, "It's virtually impossible that the current NCVS design could be in every state," because they only had nationals estimation objectives and anyway I could go through a long argument.

Dr. Fay: But the short of it is, you can almost be sure that some states aren't in the NCVS at all. And then those other states that are in with less than a million population don't have that much sample. Particularly for violent crime, but even for property crime in very small states, if you have sample, it might be an only one county and it's representing other counties in the state.

Dr. Fay: You can do that in New York state and take primary sampling units, or California or Texas. But in small states, that's really pushing the limits. But in between, I think these estimates, getting above a million we can make some sense of them. And for at least a period of time, at least projected out, we've honed even have boosted estimates for some time. And many states and even in the states about to be boosted 11 states, these will be the best estimates we have for a little while.
Dr. Fay: So I see this as one of a fitting into one of the three overall goals contributing and not taking away from the other efforts that BJS is doing to produce small area resolves. Looking from a distance and with a disclaimer here is my own opinion, but it may be these methods may provide useful in the future as BJS needs to integrate the low cost companion survey and the 40 CBSAs. What the NCVS may be saying there's a core NSVS.

Dr. Fay: So that might be something in the future and it might help us to understand what are the differences between the low cost survey going to be guild in the core NCVS. With that was that small look to the future, I'd like to turn it back to Dr. Langton.

Dr. Langton: So before we turn to the Q&A's I just want to talk quickly about the utility of these SAE estimates, and also where you can find the data. And Dr. Fay has touched on all of this already. So this is just a sort of a summary at the end here.

Dr. Langton: Dr. Fay put up the web location for the report, but for those who aren't able to quickly jot that down, if you go to the BJS website, www.bjs.gov. On the home screen, you'll actually see the report in the list of new releases. If you click on that title, they'll go to a page that has an abstract, a link to the PDF file and also has a link that says Excel. Those Excel tables contain all of the data that Dr. Fay has discussed here and more.

Dr. Langton: So for each state, you'll have estimates over time, in those three-year rolling periods for a range of violent and property crime. There are also some, a point estimate the numbers of crimes. And then also for select counties and CBSAs, you'll find that information as well. Now again, these are developmental estimates and you know, this methodology has been vetted and Dr. Fay talked about he's gone through steps to verify and validate what he's done.

Dr. Langton: But you know, the reality is we're still wrestling with how to disseminate all of these different types of subnational estimates that we're generating, and any new information that added to the model or if we pull in the data from the boosted states, this could change the estimate slightly and so that's something to take into consideration.

Dr. Langton: But even with that caveat, we think there's a lot of value to having these numbers and to getting a better understanding of where your state stands relative to the national averages to the national average. And then relative to other states with similar characteristics. They have utility, these estimates have utility for examining differences in crime trends between states.

Dr. Langton: Beginning to try to understand why one state's crime rate may go up at a certain point in time while others go found. Dr. Fay mentioned, these are not particularly good, these estimates don't have a lot of utility for examining short term changes over time. For example, the impact of a particular program or...
Developmental Estimates of Subnational Crime Rates Based on the National Crime Victimization Survey
January 27, 2016

policy on crime rates. But we can go back over the longterm and look at longer trends from the late 1990s and on.

Dr. Langton: Then another use, as I mentioned earlier, is resource allocation. Starting to think about whether if resource allocation for law enforcement and for other criminal justice entities were based on victimization rates that include the dark figure of crime rather than on official police statistics, what would that picture look like?

Dr. Langton: And of course, you know, the one caution here that has also been mentioned is that because the UCR data figure heavily into the models themselves, it's difficult to really use these estimates to compare to rates of reported crime across states. But otherwise, the data are there on the BJS website and we would certainly encourage you to download them and to begin exploring what you can do with them.

Dr. Langton: Here is the contact information for me and for Dr. Fay and we can take some questions now. But then also, you know, feel free to email us if other questions come up after the fact.

Dr. Orchowsky: Thank you. If you do have questions for our presenters, please email them in, chat them in to our chatline. Let me ask you a couple of things as a non-statistician. I guess my first question is, do you look at other possible inputs to the models other than UCR? And it just seems like there would be like an infinite number of those or a number of additional things that one could look at to model the NCVS estimate. Do you look at anything else?

Dr. Fay: Early on, we did look at a few other characteristics, not extensive but one of them suggestion of a colleague of mine was perhaps to look at murder rates, but that turned out not to contribute to what the modeling these characteristics we've produced. That would have also been from the UCR, but the notion was maybe they would be less variable in terms of the standards and criteria.

Dr. Fay: We looked at a very simple finding from just looking at the national tabulations, is that simply dividing people into owners and renters. You've got almost a factor of two, not quite on the on violent crime. Now, that's not controlling for other things. So that seemed at a national level to be a clue and yet putting that in the model of proportion of renters again, didn't do much for us or anything.

Dr. Fay: Are we limited our data snooping in a sense the things that would make substantive sense? And there's some other work that was ongoing because at one point I was also helping think through what characteristics would be good for modeling county crime rates. For use in the design of the NCVS because the Census Bureau, traditionally every redesigned has tried to fit a model. Modeling the relationship between the NCVS and other census characteristics that would be available.
Dr. Fay: That didn't produce a very long list either and again even rent or did a little bit at the county level but not very much. Interestingly, we went back looking back through unpublished material, found that Census Bureau had no luck modeling using aggravated assault from the UCR as a predictor of violent crime in NCVS, a decade previously. I think during the course of the project I was looking for suggestions and there may be someone can send me a suggestion, although or send BJS a suggestion if the model is fitted again.

Dr. Orchowsky: What about BJS is paying for a lot of our members to do state level victimization surveys and a number of states have done these and some states have done them every few years. Would that serve as an input or how would those results be expected to? How would you expect those results to compare to these state level estimates from NCVS?

Dr. Fay: I was just aware of for example, it came across a reference to nhanes recent release of their state level. And I don't know how, I think the questions to ask the first questions I would ask are whether you get a uniform methodology across states and can that be ... the more uniform the methodology, the more easily it fits into some kind of compositing or small area model. The small area estimates kind of try to use the same just as NCVS as the primary input to the small area model has the same questionnaire across the country.

Dr. Fay: I think the matter would be how uniform the state results might be from one state to the next. And that can be affected not only by the questionnaire but the motive collection response rate. It's kind of a tricky business sometimes to collect these data.

Dr. Orchowsky: You talked about, I'm just trying to conceptually you talked about the root-mean-square error and the other measure. We know that both of these things are inaccurate in the sense, we know that they contain error, and we know pretty much at this point what the nature of that error is for both sources.

Dr. Orchowsky: So how do you know if you've improved, I guess would be my question on ... or is it not a matter of improving on what we've got? Is it a matter? I guess so it's like partly it's a question of how do you evaluate, how do you know if your estimate is better? And because you're on the phone, I'm using air quotes on better. But also ultimately the purpose of doing these estimates, I think a lot of us assume that, at least for some of these offenses, like sexual assault that we're going to get more accurate responses from the victimization survey because so many people don't want to report to the police.

Dr. Orchowsky: So how do you know if that has had in fact happened to? Does that make sense?

Dr. Langton: So I think you're getting at sort of the comparison between the different subnational estimates using direct versus using like a model based and how do
we assess, which one is correct? Again, using the air quotes, is that along the lines of what you're getting at?

Dr. Langton: I think that we would say in general that the direct estimates are the gold standard. I mean that's what we'd want to do, if we could across the country and across all counties. But that's not feasible. And even with crimes like rape and sexual assault for example, which are relatively rare. Producing subnational estimates is a challenge because you have to be able to get a sufficient number of victims responding in those areas to be able to put out reliable estimate.

Dr. Langton: There may be situations in places in the NCVS where we have little or no sample even for particular subgroups. That's something that's still up for discussion and for particular crime types too, where it may be that we actually need to rely on more of a model based approach, or it's a matter of rolling up the direct, the data from the boost over an extended period of time to be able to get sufficient sample sizes to be able to produce reliable estimates.

Dr. Langton: So then it's a balance of, "Do you want the estimate tomorrow, or can you wait three years to be able to get it?" The dissemination strategy and how to integrate and use these different estimates in concert is something that we're still working through. But I think there's a lot of value in each of them and they're very complimentary. I mean, again, for example, you know, with the boost data, we're never going to be able to go back in time.

Dr. Langton: So if we want to do that, we have to rely on something like what Bob's doing or trying to use existing sample and reallocating it and doing things like that. So I mean it's a very good question and these are still things that we're wrestling with. But like you said, there's error in all of these estimates and I think, you know, we assume that what we're getting through the boost is the gold standard and we'll sort of a set everything based on that.

Dr. Orchowsky: You mentioned that was just completed ...

Dr. Langton: The pilot test was just completed.

Dr. Orchowsky: So you have another chance to look at that.

Dr. Langton: No, we got to file based on 18 months of the data and we started looking at that, but then we really wanted to get that two and a half years, and even then ideally we think three years to get the CVs, the relative standard errors that we want. So we'll be putting out a report on that later on this year.

Dr. Orchowsky: How does that met help the estimation? I know it was with the estimation but we'd be able to take another iteration of that based on the boosted data, and then hopefully this will model that boosted data even more accurately.
Dr. Langton: Right. I mean, these are discussions that we're having and then in addition to that, and I'll let you jump in and one other thing to mention is that we're going through an expensive instrument redesign with the NCVS right now. And one of the things that we're considering adding in, like I talked about the companion survey, are these non-crime questions.

Dr. Langton: So getting at things like perceptions of the police, fear of crime, perceptions of safety in the neighborhood and those types of variables. If they're collected in a systematic manner through the NCVS could also potentially feed into a model.

Dr. Fay: Yes, until I ad. The boosting, just fitting this model. Now that the models developed in the software. Lynn is right, they would take tweaking to reflect the changes in let's say the variances in copd, sampling variances and covariances because I've model uses one various generalization now, but there needs to be a small change to reflect the increase sample size in these states due to the boosting.

Dr. Fay: But when that work is done and it's not a huge deal, I can say in advance that the model estimates that might come out would really look very similar because of the compositing to the boosted estimates themselves because the boosted estimates in the boosted states will reach your level of reliability that model isn't. Just waiting together as the sample estimate and the regression model would recognize this and the outcome very similar to what boosted estimate is.

Dr. Fay: But there's an indirect benefit, first of all, to be able to see a retrospective historical series that's now anchored into a more definite boosted estimate. In terms of what we've seen today, I'd love to see Pennsylvania when that comes out. That's one of the original 11, to see if that was just a fluke that the violent crime rate went up there particularly recently. So once the data are available, the rerouting the model would kind of anchor the experience of Pennsylvania into what's more known for the most recent period.

Dr. Orchowsky: I am still looking for our first question from our participants, and I'm not seeing one. So I guess we've covered everything there is to cover. We're going to go ahead and put up the poll, we would appreciate your participation. I don't know if you can answer this or not, but you did mention funding and I think it's remarkable that you managed to do this considering, the amount of funding that BJS has had in recent years.

Dr. Orchowsky: What is the thinking about, where this is going in the future, say the next five years? Well, maybe even three years, are you confident you're able to pull off at least the boosted piece and ...?

Dr. Langton: Well, I mean, we've gone up to 22 states and the Census Bureau has also changed their sample design to make it more flexible so that if we need to expand or contract the number of states that we're boosting in, we have the
flexibility to do that now on a relatively short-term basis where we wouldn't have been able to do that before.

Dr. Langton: Yes, we've boosted now in 22 states with the longterm plan of this being the way forward. We have support from other federal agencies that are also very interested in having these estimates. There's a lot of value, for example, in terms of understanding the victimization from the perspective of the office of violence [inaudible 01:11:25], office of victims of crime. So we have a lot of other support. So yes, I mean, we're thinking of this as a longterm program.

Dr. Orchowsky: Have you had a chance to present this to anybody who does control the purse strings? I would think that there would be interest in this from congress and at least this administration and those about the next. But because of the potential to have such a rich source of data that we just have not had in the past. Do you have any inklings? But somebody's going to Andrew a big old fat.

Dr. Langton: You're asking me questions that are above my [crosstalk 01:12:16] right now.

Dr. Orchowsky: No, you need a crystal ball really all over.

Dr. Langton: Right.

Dr. Orchowsky: All right, well thank you all very much for joining us. thank you. Dr. Langton, Dr. Fay for presenting this information. And the emails are up there, if you all are apparently are too shy to ask your questions here. But today, if you'd like to email those folks, they're happy to answer your questions. Thank you for joining us this afternoon. Have a great day and look for our next JRSA webinar.

Dr. Langton: Thank you.