A Rejoinder to Dressel and Farid: New Study Finds Computer Algorithm Is More Accurate Than Humans at Predicting Arrest and as Good as a Group of 20 Lay Experts[1]

Alexander M. Holsinger
Criminal Justice Coordinator, Johnson County, KS
Christopher T. Lowenkamp
Administrative Office of the U.S. Courts
Edward J. Latessa
University of Cincinnati
Ralph Serin
Carleton University, Ottawa
Thomas H. Cohen
Administrative Office of the U.S. Courts
Charles R. Robinson
Administrative Office of the U.S. Courts
Anthony W. Flores
California State University, Bakersfield
Scott W. VanBenschoten
Administrative Office of the U.S. Courts

IN A RECENT article published in *Science Advances*, Dressel & Farid (2018) presented results from their recent study that they believe call into question the accuracy and fairness of the COMPAS risk assessment tool specifically and all statistically-based prediction tools more generally. In reaching these two conclusions, Dressel and Farid made the arguments that laypeople are as accurate (or better) and as fair in their prediction of reoffending as statistically based risk assessment instruments empirically designed to predict reoffending.

It is interesting to note that Dressel and Farid came to these conclusions by analyzing the same data used by Angwin, Larson, Mattu, and Kirchner (2016) just two years earlier. Angwin et al. concluded the COMPAS was biased against African Americans—and were subsequently taken to task in several rejoinder articles for failing to understand such basic statistical concepts as base rates, percentages, and statistical significance (see Chouldelkova, 2016; Flores, Bechtel, & Lowenkamp, 2016; Spielkamp, 2017). Now, it seems Dressel and Farid are traveling a similar path.

Several of us have devoted much of our professional careers to developing and studying the use of risk assessment in the field of criminal justice and were dismayed and disappointed when *Science Advances* published the Dressel and Farid study. While we normally applaud instances when researchers challenge accepted conventions, we also expect those offering critiques to do their homework and to offer compelling evidence. Unfortunately, we saw neither in this study.

In the following pages, we closely examine the authors' premise, methodology, and conclusions, focusing on some omissions and incorrect assumptions. In addition, while Dressel and Farid focus on the binary decision of "future crime" (yes vs. no), we also argue that risk assessment has important justice-related objectives beyond merely predicting

new criminal conduct. We also think it is worth noting that none of us has any ties to COMPAS or its parent company Northpointe. This rebuttal is not meant as an endorsement of COMPAS.

Criminal Justice Decision Making

It is hardly an overstatement to say that criminal justice case processing is driven by decision making. At every stage of the criminal justice system decisions must be made—decisions that may have serious, lasting, and reverberating effects for criminal suspects, victims, defendants, inmates, probationers, and parolees, as well as their families and friends. From the decision to arrest (or not) at the "beginning" of the criminal justice system, to the decision whether or not to grant parole at the "end," criminal justice professionals including police officers, prosecutors, judges, parole boards, and community supervision officers make important decisions on a daily basis. These decisions obviously affect those who are justice-involved as well as their families, but they also can affect our communities.

One could argue that concerns relating to public safety, in some form, are central to professional decision making in criminal justice. Whether the central focus of the system is a suspect, a defendant pending release from pretrial detention, a prison inmate, or a person on community supervision, the prevention and suppression of crime and future criminal behavior is undoubtedly one of the primary interests of criminal justice decision making. As a result, much attention is paid—and rightly so—to decision making in justice case processing. Although ensuring public safety is an important goal, we are also concerned about ensuring fairness, justice, transparency, and due process. Indeed, the use of objectivity and evidence-based risk assessment in criminal justice decision making have been emphasized for many decades (see for example Gottfredson, 1987), including recently (see Desmarais, Johnson, & Singh, 2016) as a means of promoting a fairer and more equitable way to make decisions. Objective evidence and the influence of research can be seen clearly via the development, implementation, and testing of actuarial risk assessment instruments designed to aid criminal justice decision making.

The Development of Actuarial Risk Assessment in Criminal Justice

The development and use of actuarial assessments are perhaps most advanced in the correctional environment. Indeed, risk assessment has evolved from "gut feeling" intuition-based (and often bias-ridden) decision making to fourth-generation assessment tools that not only allow for objective risk management, but also facilitate case planning and the measurement of change in dynamic risk factors. In correctional settings (secure and/or community-based), the time investment that comprehensive risk and need assessment requires is regarded as an essential component of an evidence-based decision-making process. For example, those placed on probation are often given a sentence that can range from some months to several years, with the presumption that they will receive some interventions along the way that are designed to help them address some of their dynamic risk factors, such as substance abuse treatment or employment training. As such it makes sense that probation officers would apply their time and expertise to learn as much as they can about the individuals they will be supervising for sometimes lengthy periods (Miller & Maloney, 2013; Lowenkamp, Lovins, & Latessa, 2009). In probation and other correctional settings it is not uncommon to find assessment tools used to help make a variety of decisions, including the types of services and programming required, the intensity of supervision, or even whether the individual requires more restrictive placement. Indeed, there is a large and growing body of literature demonstrating the effectiveness and benefits of actuarial assessment in correctional environments.

The use of actuarial assessment tools is found at every stage of the court and correctional system. For example, it is not uncommon to find actuarial risk tools in place in pretrial settings—instruments designed to assist with decisions about pretrial release and community supervision for defendants (Lowenkamp, Lemke, & Latessa, 2008; Lowenkamp & Whetzel, 2009). Actuarial tools also have a long history of use in jails and prisons to help classify and make appropriate housing placements, and are used by some probation and parole agencies to help determine how best to handle violation of supervision. Recently actuarial risk assessment has also become a point of discussion at the sentencing stage (Monahan & Skeem,

2016; Scurich & Monahan, 2016).

The advent and proliferation of actuarial assessments has largely been viewed positively and with continued optimism. In fact, "assessment" in and of itself is typically regarded as an essential component of evidence-based practice (Andrews & Bonta, 2015). The incorporation and use of standardized and objective measures are seen as an improvement over purely qualitative and unstructured one-on-one clinical assessment, which may invite bias into decision making (intended and/or unintended). Given the gravity of criminal justice decision making, and the importance of pursuing justice, the constructs of objectivity and standardization are key. Further, despite the evidence supporting the effectiveness of actuarial assessment, seldom if ever has the recommendation been made to "blindly" follow the risk score; in other words, people make decisions, not the instrument alone. Yet, use of actuarial risk assessment tools should help guide those decisions, in part because they help summarize and sort large amounts of information in a systemic, objective, and standardized way (Shlonsky & Wagner, 2005).

Nonetheless, actuarial assessment in justice settings has come under renewed skepticism regarding all the potential ills it is typically designed to address, most notably inaccuracy (via "false positives" or "false negatives") and bias (circumstances where an algorithm is not as effective for, and/or over-classifies specific demographic groups). One central question has driven the controversial discussion surrounding risk assessment thus far, and yet remains in some pockets: Is a risk assessment instrument or algorithm "better at" predicting the likelihood of future criminal behavior than a human being?

The Research on Risk Assessment

For an article devoted to comparing the predictive effectiveness of risk assessments to that of human judgement, it is puzzling and somewhat surprising that the authors either omitted or ignored an entire body of literature illustrating the capacity of actuarial devices to outperform human decisions in risk prediction. The literature on the predictive capacity of risk instruments dates back to the 1920s (see Burgess, 1928) and over time, our knowledge of how actuarial risk assessments outperform clinical or professional judgments has been augmented by hundreds of individual studies and more recently by a number of meta-analytical studies. Meta-analyses essentially entail

the pooling of many studies, and these research efforts have empirically demonstrated the extent to which these instruments consistently and uniformly generate predictions of risk that surpass those of humans (see Ægisdóttir et al., 2006; Grove, Zald, Lebow, Snitz, & Nelson, 2000; Meehl, 1954). Curiously, the authors do not mention any of these seminal works, though they do briefly discuss one meta-analysis on sex offender assessments where the authors concede that these instruments provide more accurate predictions than "unstructured measures" in the task of assessing which sex offenders represent a danger to the community.

Critique of Dressel & Farid

Accuracy

Perhaps the most damning shortcoming of this study is that the authors provided laypeople with an "edge" in predicting recidivism. This edge was provided in two ways. The authors restricted the information that was presented in estimating risk to risk-relevant measures. As is discussed below, this process of restricting the information to risk-relevant factors is a known method in estimating risk that has been shown to be about as effective as actuarial approaches. The authors also provided an unfair advantage to the participants in that the participants were "trained" as they worked through the data, potentially learning what is predictive and what is not.

The description for each vignette was limited to risk-relevant information. It included the defendant's age, sex, offense type, offense severity, adult convictions, juvenile felony charges, and juvenile misdemeanor charges—virtually all of which are robust risk factors for recidivism (see, for example, Gendreau, Little, & Goggin, 1996). In real cases and real settings, decision-makers must deal with reams of (often risk-irrelevant and biasing) information. Perhaps wildly different predictions would have been made if things like residence, education, employment, family relationships, family structure, stress, depression, self-esteem, mental illness, physical well-being, veteran status, attitudes, peers, substance abuse, treatment episodes, past performance on community supervision, infractions while incarcerated, socio-economic status, financial holdings, a victim's impact statement, sentencing law, and the like were presented in the vignettes along with age, sex, and criminal history.

A large body of research indicates that structuring human judgment (by providing checklists of relevant risk factors, for example) yields much more accurate violence predications than unaided human judgment (see Skeem & Monahan, 2011). In essence, Dressel and Farid structured laypeople's decision processes—which may have scaffolded their accuracy. The authors wonder whether laypeople's predictive accuracy "would improve with the addition of guidelines that specify how much weight individual features should be given." That would amount to transforming a structured clinical decision (which is more accurate than unaided judgment) into something like an actuarial decision (which is also more accurate than unaided judgment, and akin to the COMPAS).

One of the central problems with "expertise" is that experts, like judges, rarely if ever receive feedback about the decisions made and the resulting outcomes. This feedback over time would certainly impact one's decisions. Participants in the current study were given two forms of feedback after each answer: whether the response was correct and their average accuracy. So, in essence, though the authors assumed the participants had little to no experience in criminal justice, each participant was given years of "experience" via the provision of potent and risk-relevant factors and feedback after each response. A vast literature indicates that people—like other sophisticated organisms—learn, with feedback. The process that Dressel and Farid used amounts to "human learning" rather than validation. It would be valuable to re-test the layperson after he or she had been "trained." To validate the laypersons' abilities to predict without feedback on a new set of cases would have been much more akin to statistical validation or model confirmation. In essence, the process used likely "overfitted" the model to the data. In the absence of any feedback and learning process, predictions are likely to have been much less accurate.

Dressel and Farid recruited their research subjects via Amazon's Mechanical Turk (MTurk), an online crowdsourcing marketplace that pays volunteers to complete online tasks. Once enrolled in the study, participants were given a series of vignettes with information pertaining to an actual person who had been charged with a crime. After reading the vignette, subjects were asked to indicate (yes/no) whether they thought the person would commit another crime within two years. After each answer, study subjects were told whether they were correct in their assessment. It is important to point out (see below) that study

subjects were paid $1.00 for completing the task and a $5.00 bonus if their overall accuracy was greater than 65 percent.

The authors assert that the individuals who made predictions on MTurk are "nonexpert" (page 2) and can be "assumed to have little to no expertise in criminal justice…" (pages 1 & 3). This assertion is dubious for a number of reasons. First, the title and description of the task on Amazon's Mechanical Turk (MTurk) may have had direct bearing on a person's decision to take part in the survey. The task title was "Predicting Crime" and the description given was "Read a few sentences about an actual person and predict if they will commit a crime in the future," with the following key words: "survey, research, criminal justice." Rather than assume that participants had no background or expertise, why didn't the authors explicitly ask participants? It seems reasonable to assume that people might search for tasks that fit with their training, interest, and/or expertise. Without knowing the backgrounds of the participants, it is reasonable to assume that at least some of the participants had some level of training or expertise in criminal justice prior to taking part in this study, or at least an above-average interest in the subject at hand.

Another concern is that participants had a financial incentive to make accurate predictions. As noted above, those taking part in the study were paid to read and make determinations, with a $5 bonus for achieving 65 percent accuracy or higher. While $5 is not a lot of money, it certainly increased the chances that some might consult the internet or an old textbook to learn more about the best predictors of criminal behavior. A quick internet search reveals that age, prior criminal history, and being male are all good predictors of criminal behavior (see Figure 1, next page). A little more searching on the internet leads one to know that early onset of criminal behavior (i.e., a juvenile record) is also a good predictor. Again, while $5 may not seem like much money, one must realize that the participants in the "Predicting Crime" task on MTurk were already reading 50 vignettes for $1 each. As such, it is certainly possible that some of the participants might spend a little time trying to increase their accuracy, given that an acceptably high rate of accuracy paid more than simply completing the task. It may sound like quite a leap to think that participants would go to more trouble to increase accuracy for a mere $5, but perhaps no more so than it was for the authors to assume they

were random laypersons.

In short, specifying relevant risk factors and providing feedback on the degree to which they predict recidivism "loaded the deck" for laypeople to predict more accurately than experts would in the much more complicated context of a real criminal case—with relevant and irrelevant information provided by the defense and prosecution and with no feedback about recidivism after an individual leaves the courtroom.

Fairness

In what seems to be an afterthought, the authors indicate that "…differences in the arrest rate of black and white defendants complicate the direct comparison of false-positive and false-negative rates across race (black people, for example, are almost four times as likely as white people to be arrested for drug offenses)." Stated differently, once an instrument demonstrates predictive parity (which the COMPAS does), it mathematically follows that different base rates of the criterion measure (in this case rearrest) across groups (i.e., white and black defendants) will necessarily lead to different rates of false positive and/or false negative rates. This is problematic as this issue is not resolved in the Dressel and Farid paper, and false negative and false positive rates are used as the measure of bias or figure into the calculation for bias in most of the measures reported throughout Tables 1 and 2 in the study (see pages 2 & 4 of Dressel & Farid, 2018).

Perhaps most egregiously, Dressel and Farid actually found that the COMPAS did outperform laypeople in predicting recidivism—at a statistically significant level. The authors downplay this finding—claiming that laypeople's accuracy is "just barely" lower than that of COMPAS. For justice-involved people who are the subject of recidivism predictions, "just barely" lower probably matters. Again, recall that this is really a comparison of structured judgment to fully structured/actuarial assessment—and fully structured/actuarial assessment "won." Not by much, but again how important that small margin is depends on which side of the "correctional desk" you sit on.
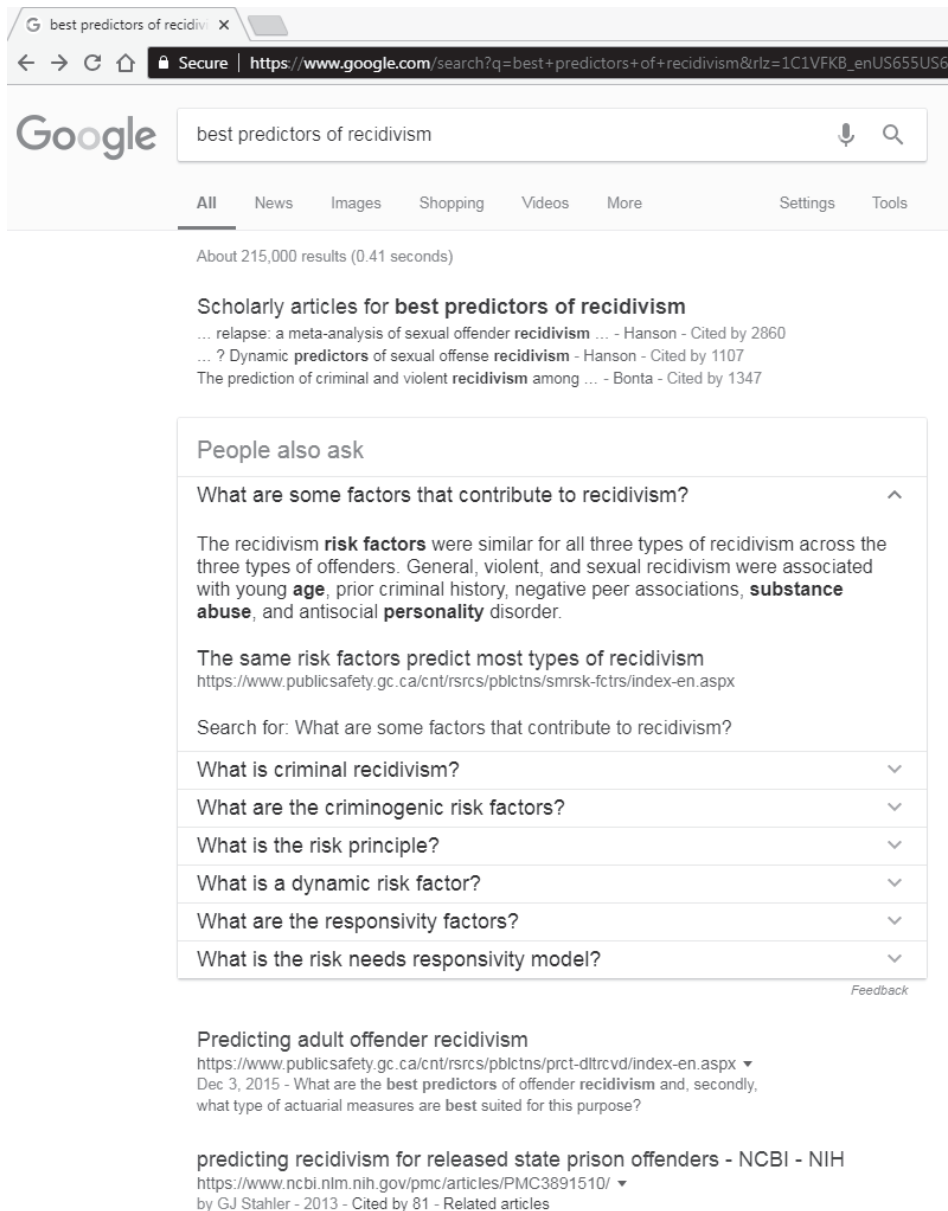
A Note on a Relevant Omission

In addition to omitting an entire literature on the topic of risk prediction, the authors speculate that it "remains to be seen whether the addition of dynamic risk factors would improve predictive accuracy" and then note that the integration of dynamic factors into COMPAS has not resulted in improved prediction for this instrument. First, this statement is somewhat puzzling, because although COMPAS does collect information on a large number of dynamic risk factors, relatively few of these contribute to the instrument's overall risk score; hence, it's not possible to say with certainty that "integration of dynamic factors into COMPAS" has failed to improve this instrument's predictive capacities. Moreover, these statements imply that human predictions based on a narrow range of factors that are generally not amenable to change (e.g., age, sex, and criminal history) are all that is required to assess recidivism risk. Such notions are unfortunate because they reduce the concept of risk assessment from one involving a holistic approach aimed at assessing an individual's recidivism risk and identifying crime-driving factors that, if changed, could help with the reintegration of offenders back into society into one where risk prediction is circumscribed to just a few static items. This restricted view of risk assessment has been superseded by the development and evolution of dynamic risk assessment. The topic of dynamic prediction is of crucial importance in the risk assessment and community corrections literature.

Over time, risk assessments have evolved from instruments that primarily assessed risk on static factors such as the ones used in this article (i.e., age, criminal history) to actuarial devices that can measure changeable factors

FIGURE 1
Results of Google Search for "Best Predictors of Recidivism"

that are criminogenic, meaning that they are empirically correlated with crime such as the presence of procriminal attitudes, the lack of prosocial associates, the manifestation of antisocial personality patterns, the existence of poor family relations, the inability to find and maintain employment, and the struggle with substance abuse problems. While these dynamic factors may be more difficult to measure, assessing their presence and tracking whether they are changing over time is really important if we want to try to reduce risk and protect the community.

In fact, several studies show that if probation officers correctly identify the existence of dynamic criminogenic factors through the application of risk assessment and then attempt to ameliorate them through appropriate interventions, they can reduce an offender's likelihood of recidivating. The nexus between change in measurable risk characteristics and offender recidivism outcomes can best be understood through studies conducted using dynamic risk assessments including the Level of Service—Inventory (LSI-R) and the Post-Conviction Risk Assessment (PCRA). These studies have clearly shown that an offender's risk scores and characteristics can change over time and that changes in risk scores are associated with changes in an offender's likelihood of committing future crimes. Specifically, they show that offenders with decreasing risk scores are less likely to garner new criminal arrests after reassessment, while offenders with increasing risk scores are more likely to recidivate post reassessment (See Cohen, Lowenkamp, & VanBenschoten, 2016; Labrecque, Smith, Lovins, & Latessa, 2014; Miles & Raynor, 2004; Raynor, 2007; Schlager & Pacheco, 2011; Vose, Lowenkamp, Smith, & Cullen, 2009; Vose, Smith, & Cullen, 2013). Moreover, meta-analyses done on the predictive accuracy of dynamic risk factors have shown that these factors perform at least as well as static domains (Gendreau, Little, & Goggin, 1996). Finally, it is critical to note that when empirically constructed risk instruments capable of identifying dynamic criminogenic factors are not being used by community corrections staff, officers will often engage in supervision practices that focus on addressing issues uncorrelated with crime. It is unfortunate when officers target non-criminogenic needs, as the result is often a waste of corrections resources with no reduction in an offender's proclivity to recidivate or enhancement in community safety (see Lowenkamp, Latessa, & Holsinger,

2006; Oleson, VanBenschoten, Robinson, Lowenkamp, & Holsinger, 2012).

The bottom line is that without dynamic risk assessments, officers will be unable to assess the presence of changeable risk characteristics, devise and implement strategies to address these characteristics, and monitor in a systematic and quantifiable fashion whether offenders are improving, remaining unchanged, or worsening while under supervision.

Last, this study overlooks an entire body of literature where community corrections professionals working in the criminal justice field discard risk assessment recommendations for their own "seat of the pants" judgments. In the risk assessment field, we call these decisions "supervision overrides." Basically, an override occurs when a community corrections officer decides that the risk assessment instrument their department uses has incorrectly assessed an offender's propensity to recidivate and decides to supervise that offender at a level of intensity diverging from the risk instrument's recommendation. Most overrides involve an officer's decision to supervise offenders at levels of intensity higher than recommended by the original risk classification. Though overrides typically occur with sex offenders, this is not always the case, and not surprisingly, we find that overrides typically result in reducing the validity of the tool in predicting risk to reoffend. In other words, in criminal justice contexts where officers decide to ignore an actuarial risk recommendation and exercise their own discretion or judgment, the officer's decisions are usually not as predictive as that of the actuarial classification. We find it surprising that this literature was not acknowledged or discussed in this article (see Cohen, Pendergast, & VanBenschoten, 2016; McCafferty, 2017; Wormith, Hogg, & Guzzo, 2012).

Final Thoughts

A judicious read of the paper suggests it was primarily intended to challenge the utility of actuarial risk assessments, with purported evidence regarding the accuracy of laypersons' ability to "conduct" risk assessment, albeit using a markedly constricted range of factors. Concurrently, it seems the authors also raised issues regarding the need for transparency in risk assessment (i.e., public access to algorithms in commercialized risk scales) and extolled the merits of mechanical learning over other statistical analytic approaches for selecting and weighting risk factors. From our read of this research, mechanical learning fails to offer any particular advantage

in terms of predictive accuracy, despite its increased prominence in recent risk assessment research.

Interestingly, the field of risk assessment has previously seen debates regarding which risk assessment instrument is most effective (Baird, 2009; Gendreau, Goggin, & Smith, 2002; Hemphill & Hare, 2004). Such debates have now been replaced with the recognition that risk instruments tend to share many similarities, both in terms of accuracy (Yang, Wong, & Coid, 2010) and content (Kroner, Mills, & Reddon, 2005). Indeed, the discussion is more about which measure best fits an agency's needs (i.e., costs, training requirements, synchronization with internal datasets, time taken to complete, and accuracy) and fidelity of assessments. Moreover, as noted earlier in this paper, primarily static risk assessment scales are being augmented by more dynamic risk measures (such as the COMPAS, LSI-R, and PCRA), and also risk measures intended to assess changes in acute dynamic risks (Serin, Chadwick, & Lloyd, 2015). This evolution in risk assessment has led to incremental improvements in accuracy, as well as refinements in case planning and supervision (Serin, Lowenkamp & Lloyd, in press).

In our view, Dressel and Farid largely "rediscovered" what has been well-established in a large body of risk assessment literature: Compared to unstructured human judgment, structured human judgment and actuarial approaches are more accurate. Structuring decisions limits consideration and unnecessary emphasis on factors that are unrelated to risk of recidivism (i.e., bias).

This "rediscovery" is not apparent in Dressel & Farid's article (or news coverage of it). For example, see the *New York Times* headline "Can Software Predict Crime? Maybe So, But No Better Than A Human."[2] This oversight potentially negates the advances of past decades. A return to unstructured risk assessment, a natural conclusion from their paper, will actually increase bias and potentially lead to capricious decision-making, hindering accuracy, jeopardizing public safety, and risking fairness for clients.

As noted earlier, their own data do not support that laypersons are more accurate or that machine learning is more accurate than the COMPAS. We believe the field of risk assessment cannot advance beyond the status quo when the focus remains on minutiae

---

[2] https://www.nytimes.com/2018/01/19/us/computer-software-human-decisions.html

(i.e., seeking meaning from differences of .02 gains in comparisons of AUCs resulting from different risk measures). Regardless of which validated risk assessment measure is used, more promising pursuits and discussion would include recognizing the need for fidelity in scoring risk measures, strategies for developing greater understanding regarding the merits and limits of risk scores, and using risk assessment results more specifically in case planning and supervision.

References

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., Nichols, C.N., Lampropoulos, G.K., Walker, B.S., Cohen, G., & Rush, J.D. (2006). *The meta-analysis of clinical judgment project: Fifty-six years of accumulated research* on clinical versus statistical prediction. *The Counseling Psychologist, 34*, 341-382.

Andrews, D.A., & Bonta, J. (2015). *The Psychology of Criminal Conduct*. Routledge, NY.

Angwin, J., Larson, J., Mattu, S. & Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Burgess, E.W. (1928). Factors determining success or failure on parole. In A. Bruce, E. Burgess, & A. Harno (Eds.), *Prediction and classification: Criminal justice decision making*. Chicago, IL: University of Chicago.

Chouldechova, A. (2016). *Fair prediction with disparate impact: A study of bias in recidivism prediction instruments*. Retrieved February 2, 2018, from the arXiv database.

Cohen, T.H., Lowenkamp, C.T., & VanBenschoten, S. (2016). Does change in risk matter?: Examining whether change in offender risk characteristics influence recidivism outcomes. *Criminology and Public Policy, 15(2)*, 263-296.

Cohen, T.H., Pendergast, B., & VanBenschoten, S. (2016). Examining overrides of risk classifications for offenders on federal supervision. *Federal Probation, 80(1)*, 12-21.

Desmarais, S.L., Johnson, K.L., & Singh, J.P. (2016). Performance of recidivism risk assessment instruments in U.S. correctional settings. *Psychological Services, 13(3)*: 206-222.

Dressel, J., & Farid., H. (2018). The accuracy, fairness and limits of predicting recidivism. *Science Advances, 4*(1).

Flores, A., Bechtel, K., and Lowenkamp, C. (2016). False positives, false negatives, and false analyses: A rejoinder to Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. *Federal Probation, 80(2)*: 38-46.

Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology, 34(4)*, 575-607.

Gottfredson, D.M. (1987). Prediction and classification in criminal justice decision making. *Crime and Justice, 9*: 1-20.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*, 19–30.

Labrecque, R., Smith, P., Lovins, B., & Latessa, E. (2014). The importance of reassessment: How changes in the LSI-R risk score can improve the prediction of recidivism. *Journal of Offender Rehabilitation, 53*,116-126.

Lowenkamp, C.T., Latessa, E., & Holsinger, A. (2006). The risk principle in action: What have we learned from 13,676 offenders and 97 correctional programs? *Crime and Delinquency, 51(1)*, 1-17.

Lowenkamp, C.T., Lemke, R., & Latessa, E.J. (2008). The development and validation of a pretrial screening tool. *Federal Probation, 72*: 2-9.

Lowenkamp, C.T., Lovins, B., & Latessa, E.J. (2009). Validating the Level of Service Inventory – Revised and the Level of Service Inventory: Screening Version with a sample of probationers. *The Prison Journal, 89*: 192-204.

Lowenkamp, C.T., & Whetzel, J. (2009). The development of an actuarial risk assessment instrument for U.S. Pretrial Services. *Federal Probation, 73(2)*: 33-36.

McCafferty, J. T. (2017). Professional discretion and the predictive validity of a juvenile risk assessment instrument: Exploring the overlooked principle of effective correctional classification. *Youth Violence and Juvenile Justice, 15(2)* 103 – 118.

Meehl, P.E. (1954). *Clinical versus statistical prediction*. Minneapolis, MN: University of Minnesota Press.

Miles, H., & Raynor, P. (2004). *Community sentences in Jersey: Risk, need, and rehabilitation*. St. Helier, UK: Jersey.

Miller, J., & Maloney, C. (2013). Practitioner compliance with risk/needs assessment tools: A theoretical and empirical assessment. *Criminal Justice and Behavior, 40*: 716-736.

Monahan, J., & Skeem, J. (2016). Risk assessment in criminal sentencing. *Annual Review of Criminal Psychology, 12*: 489-513.

Oleson, J.C., VanBenschoten, S., Robinson, C., Lowenkamp, C.T., & Holsinger, A. (2012). Actuarial and clinical assessment of criminogenic needs: identifying supervision priorities among federal probation officers. *Journal of Crime and Justice, 35(2)*, 239-248.

Raynor, P. (2007). Risk and need assessment in British probation: The contribution of the LSI-R. *Psychology, Crime, and Law, 13*, 125-138.

Schlager, M., & Pacheco, D. (2011). An examination of changes in LSI-R scores over time: Making the case for needs-based case management. *Criminal Justice and Behavior, 38*, 541-553.

Scurich, N., & Monahan, J. (2016). Evidence-based sentencing: Public openness and opposition to using gender, age, and race as risk factors for recidivism. *Law and Human Behavior, 40*: 36-41.

Serin, R.C., Chadwick, N., & Lloyd, C.D. (2015). Dynamic risk and protective factors. *Psychology, Crime & Law*, 22:1-2, 151-170.

Serin, R.C., Lowenkamp, C.L., & Lloyd, C.D. (in press). Managing Violent Offenders in the Community: Reentry and Beyond. In *What Works in Violence Risk Management* (Wormith, S., Craig, L., & Hogue, T. Eds.) Wiley-Blackwell.

Shlonsky, A., & Wagner, D. (2005). The next step: Integrating actuarial risk assessment and clinical judgment into an evidence-based practice framework in CPS case management. *Children and Youth Services Review, 27(4)*: 409-427.

Vose, B., Lowenkamp, C.T., Smith, P., & Cullen, F.T. (2009). Gender and the predictive validity of the LSI-R: A study of parolees and probationers. *Journal of Contemporary Criminal Justice, 25(4)*, 459 471.

Vose, B., Smith, P., & Cullen, F.T. (2013). Predictive validity and the impact of change in total LSI-R score on recidivism. *Criminal Justice and Behavior, 40(12)*, 1383-1396.

Wormith, J. S., Hogg, S. M., & Guzzo, L. (2015). The predictive validity of the LS/CMI with Aboriginal offenders in Canada. *Criminal Justice and Behavior, 42(5)*, 481-508.