

Do Office Reviews in the Federal Probation and Pretrial Services System Do What They Were Intended to Do? Not Yet

Christopher T. Lowenkamp¹

Administrative Office of the U.S. Courts

Travis C. Pratt

University of Cincinnati

Harris County Community Supervision and Corrections Department

Alexander M. Holsinger

University of Missouri—Kansas City

COMMUNITY SUPERVISION—including the challenging tasks of managing, supervising, and treating justice-involved people—is hard (May & Pratt, 2022; Smith et al., 2018). Adding to the difficulty is the fact that such challenges are met with widely different degrees of effectiveness from one jurisdiction to the next (Jalbert et al., 2010; Lowenkamp, Latessa, & Smith, 2006). Indeed, community corrections agencies—even those that operate under the same state system of laws and policies—vary considerably with respect to how well they deliver correctional services to the justice-involved populations they serve (Pratt & Turanovic, 2019; Van Deirse et al., 2020; Viglione, 2019).

The federal probation system—which encompasses 94 federal districts—is no different (Sloas et al., 2019; see also Hughes & Henkel, 2015). And due in part to variation in the quality of outcomes from district to district, the Probation and Pretrial Services Office (PPSO) within the Administrative Office of the U.S. Courts (AO) has long been charged with the task of providing oversight regarding the work of the probation officers in each of the United States district courts

(Chandler, 2015; Sheil, Doyle, & Lowenkamp, 2016) in the form of federal district “office reviews” (Whetzel & Sheil, 2015). Currently, these reviews are intended to support the federal system’s efforts to reduce the risk of recidivism among correctional clients. It is, however, unclear whether this has been the case historically. It should also be noted that office reviews are, and will continue to be, statutorily required to measure certain activities regardless of their relationship to recidivism and public safety. Regardless, such reviews typically entail the analysis of a sample of processed cases, the use of performance metrics (e.g., client risk profiles, rearrest and recidivism rates), site observations, and interviews. The review process was revised in 2014 to include additional staff training, new performance metrics (including a new policy/program questionnaire), and a district self-assessment (Whetzel & Sheil, 2015).

Recent evidence suggests that PPSO has indeed been supportive of federal probation district offices (U.S. Government Accounting Office, 2023). To be sure, PPSO has helped offices to develop strategic plans, to draw up policies and evaluation tools, to train and educate staff, and to assist with budgeting and resource allocation. Such support has

certainly resulted in a more consistent—and arguably more effective—integration of evidence-based practices (EBP) into the federal probation system (Goldstein, 2020), including the development of the pretrial risk assessment (PTRA; see Cohen, Lowenkamp, & Hicks, 2018), the post-conviction risk assessment (PCRA, see Lowenkamp, Holsinger, & Cohen, 2015), and EBP-based training and support efforts such as Staff Training Aimed at Reducing Rearrest (STARR; see Robinson et al., 2011, and Lowenkamp et al., 2014) and the Criminogenic Needs and Violence Curriculum (CNVC; see Goldstein, 2020).

Yet what we know about the impact of office reviews is confined exclusively to these issues of “process” (e.g., implementation of best practices, staff training, risk instrument development). The problem is that we know nothing about the office reviews’ actual relationship to the original “outcome” goal(s) of federal supervision, whose advancement the office reviews were created to evaluate and encourage: that is, reducing recidivism and enhancing public safety. To address this issue, we analyze data from 70 office reviews conducted between 2017 and 2019. In particular, we examine the nature of the relationship between district-level outcomes

¹ Please direct all correspondence to Christopher Lowenkamp at christopher_lowenkamp@ao.uscourts.gov.

(post-conviction revocation, post-conviction rearrests, pretrial release recommendations, pretrial release rates, pretrial failure to appear, pretrial rearrest, pretrial violation, and pretrial revocation rates) and the office review items and overall office review compliance scores. Our broader purpose is to shed light on the degree to which federal probation's system oversight is linked to fulfilling its goals of fair administration of justice and public safety.

Method

Sample

The sample for this study involves all office reviews conducted between 2017 and 2019, which are the first years for which data on the office review findings are consistently available. This sampling process led to a sample of 70 office reviews, of which 12 are on pretrial-only offices. Due to incomplete data, the total number of office reviews was reduced to 69. The sample of offenders and defendants for this study includes all offenders or defendants activated within two years prior to the date that the office review was conducted. This led to a total sample of 60,521 post-conviction observations and 108,369 pretrial observations.

Office Review and Office Configuration

The office review process involves a site visit by a team of probation administrators and practitioners from the field. The team reviews district practices and measures those practices against nearly 200 benchmarks included in the office review instrument.² While the office review instrument has changed over the years, during the research time frame the office review instrument included some combination of the following areas: firearms and safety; location monitoring; post-conviction supervision; procurement; pretrial services investigations; pretrial supervision; substance use disorder and mental health; and post-conviction low-risk policy. Each area listed above receives a compliance score representing the percentage of items adhered to by the district. The office review instrument is not publicly available, yet the items are relatively pedestrian and audit-based in that they measure the mere presence of—rather than the

quality of—activities. In the U.S. Probation and Pretrial System, some districts are “combined” in that both probation and pretrial services functions are carried out by one office with one chief. In other districts the probation and pretrial offices are “separate” in that separate offices operate each service, and each is overseen by a chief dedicated to that office. In regression models a measure capturing whether a district is a solely pretrial district or a combined district was also included.

Offender and Defendant-based Measures

Data from the Probation/Pretrial Services Automated Case Tracking System (PACTS) were used to identify cases activated during the time periods of interest and to develop district-level measures of case composition and certain outcome measures; namely, revocation, detention recommendations, actual release, failure to appear, and violations. Rearrest measures were based on data from the FBI's computerized criminal history database. For each district, individual-level data were aggregated to create measures that captured the percentage of defendants or offenders that were male, White, U.S. citizen, charged with or convicted of a violent offense, a drug offense, a firearm offense, average risk scores, and average age. Outcomes for the pretrial analyses were created by calculating the percentage of cases that were recommended for pretrial release, the percentage of cases released, the percentage of cases with a violation, the percentage of cases revoked, the percentage of cases with a failure to appear, and the percentage of cases with a rearrest. The outcomes created for the post-conviction analysis are the percentage of cases with a revocation and the percentage of cases with a rearrest.

Analysis

To analyze these data, we estimated proportions and standard errors for each of the six outcomes of interest. Because we have proportions only for outcome measures, rather than for treatment effects with a comparison or control group, it is important to control for differences in the composition of cases across districts (e.g., some districts have, on average, higher or lower risk cases, which could be related both to how a district scores on the office review and how that district performs in terms of outcomes). As such, we ran regression models using Stata meta regress.³

³ While the results should be identical to models using WLS in Stata, we chose meta regress because meta regress presents the results in a format that is easier to manipulate and export to additional files

When we ran models predicting pretrial outcomes, the control variables included percentage male, White, charged with a violent offense, charged with a drug offense, charged with a firearm offense, average age, average risk level, and whether the district was a combined district or not. Models estimating the relationships between the office review instruments and post-conviction outcomes included percentage male, White, charged with a violent offense, charged with a drug offense, charged with a firearm offense, average age, and average risk level. We estimated 750 regression models for pretrial (6 outcomes and 125 office review measures and 334 regression models for post-conviction (2 outcomes and 167 office review measures).

Results

The results of the pretrial regression models indicate that the office review-related measures are statistically significant ($p \leq .05$) in 37, or about 5 percent, of the models. These statistically significant findings are distributed across the six outcomes, with about half of the relationships being negative and about half being positive. Given the pattern of results, the sheer number of statistical tests, and the small number of statistically significant relationships, we conclude that these findings are likely due to chance alone. For instance, Figure 1 (next page) provides a visual display of the coefficients plotted against the p-value. A dashed reference line has been added to the chart at a p-value of 0.05. As can be seen from the chart, the overwhelming majority of coefficients—roughly 95 percent—are not significant at $p \leq .05$; similarly, most of the coefficients deviate only slightly from a value of zero.

Regression models predicting the post-conviction outcomes reveal a nearly identical trend. One hundred and sixty-seven models were run for each of the two post-conviction outcomes (rearrest and revocation). This process generated 334 regression coefficients of interest. Of those 334 regression coefficients, only 19 (again roughly 5 percent) were significant at the $p \leq .05$ level. Further, of those 19 coefficients, 5 are positively related to a post-conviction outcome and 14 are negatively related. Most (12) are related to revocation. And similar to the results seen in Figure 1, those in Figure 2 indicate that almost 95 percent of the coefficients are not statistically significant at the $p \leq .05$, and the coefficients values rarely deviate statistically from zero.

for subsequent analysis.

² There are 167 benchmarks that relate to post-conviction operations and 125 benchmarks that relate to pretrial operations. These figures include the individual items, domain scores, and overall scores. There is some overlap between these two sets of benchmarks (e.g., risk domains like antisocial attitudes, alcohol/substance use problems, and employment).

Discussion

When it comes to the criminal justice system, making sure that policies and practices are doing what they were intended to do is both important and difficult (Pickett, 2019). To be sure, it is important to determine whether agencies are living up to their charge and to hold them accountable—and to make the necessary changes—when they are not. And that is certainly the case with federal probation

district office reviews, where such reviews were intended to identify potential threats to public safety and to help probation offices reduce recidivism and other key probation outcomes. But did these reviews actually do that? Based on our analyses of data from 69 district office reviews and over 160,000 offenders under their supervision, three conclusions are warranted.

First, the scores on the office reviews failed

to consistently predict any of the outcomes we assessed. Indeed, from rates of revocation to rearrest, the office review instrument scores were unrelated to district outcomes. So for these key outcomes, office reviews fell well short of what they were intended to do. Now granted, this is the first attempt at linking office reviews to outcomes in the federal system, so we assume that a learning curve is likely in place and that it is possible that such reviews have improved over time. Nevertheless, the results we have presented here do not paint much of a positive picture of the effectiveness of such reviews when it comes to the kinds of outcomes—like recidivism rates—that the public tends to be most concerned about (Petrich et al., 2021). Two obvious potential explanations prevail whenever any instrument or system does not predict the intended outcome(s): either the instrument measures “the right” things but is being done incorrectly, or the instrument is measuring “the wrong” things, in which case how well it is being done does not matter. In the current instance, under the assumption the results presented above have validity, a re-examination of the office review content and process is in order, if the objective is to differentiate sufficiently effective districts from less effective districts and aid the improvement of the less effective ones. Ideally, the office review process would produce a quantitative measure of a district’s effectiveness and would in turn identify improvements that can be made and benchmarks to work toward.

Nevertheless, our second conclusion is that it is important to note that systems of accountability—like these office reviews—serve other administrative purposes beyond their potential to influence things like revocation and recidivism rates. For instance, it is often necessary politically to have a system of accountability in place when it comes to law enforcement to preserve a sense of “legitimacy”—that is, the belief that the exercise of power is right and proper (Zelditch, 2006)—with respect to the justice system in general (McLean, Wolfe, & Pratt, 2019). Further, there is administrative value in measuring process, procedure, and output/activity. Even if the office reviews were ineffective when it came to the outcomes we examined here, there is a structure in place that, if changes are made, could perhaps enhance the effectiveness of these reviews. In their current state, however, effectiveness (measured in terms of accountability and public safety) is nonexistent.

And to that end, our third conclusion is that concrete changes can and should be made

FIGURE 1.
Scatter plot coefficients and p-values for pretrial measures and outcomes.

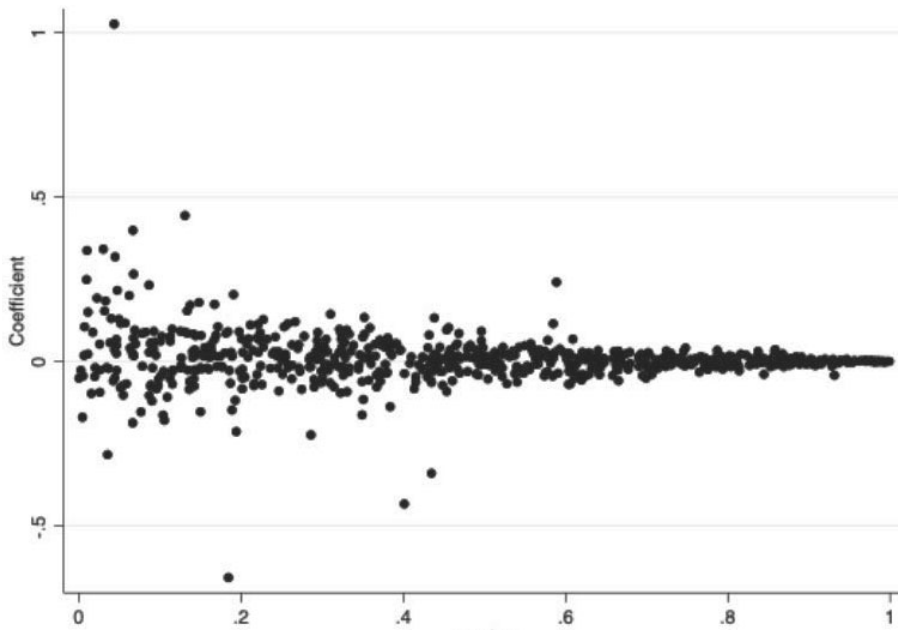
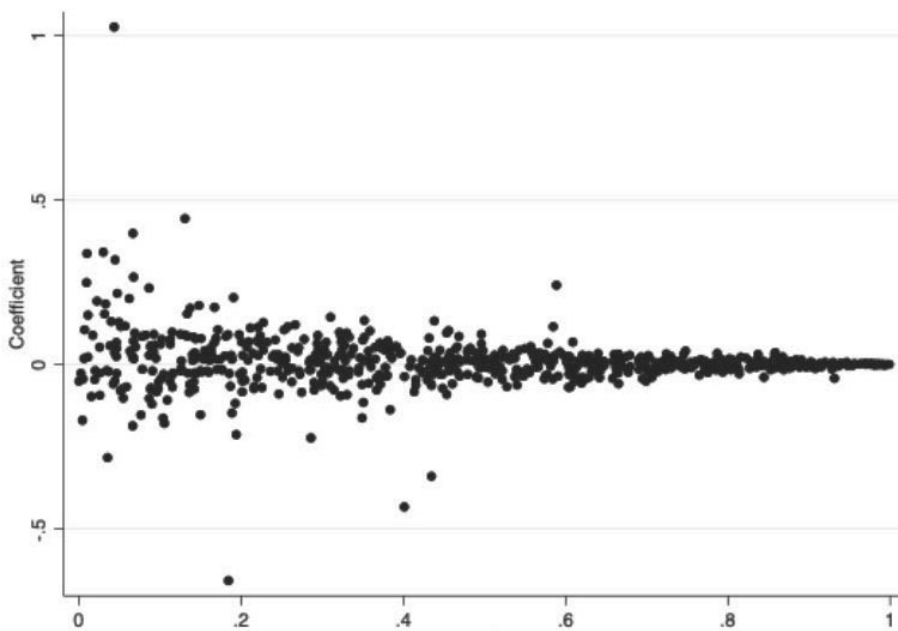


FIGURE 2.
Scatter plot coefficients and p-values for post-conviction measures and outcomes.



in the future to the office review process to make them more effective. For example, we should quickly note that there are no known psychometric properties of the office review instrument, and the instrument itself fails to adequately capture the known correlates of program effectiveness (see, e.g., Lowenkamp, Latessa, & Holsinger, 2006; Wright et al., 2012). A solid next step, then, would be to develop (or adopt) an office review instrument that adheres to what is currently known regarding effective behavioral interventions within a community supervision environment. A number of options already exist that could serve as models, such as the Canadian Risk-Need-Responsivity (RNR) model of corrections (Andrews & Dowden, 2007), the Community Corrections Score Card (Serin, Lowenkamp, & Bourgon, 2021), or even the other instruments that the office reviews have had a hand in creating already (e.g., PCRA, STARR, CNVC).

The bottom line is that precedents exist for doing better, and there is nothing stopping the federal probation system from making improvements. The process could be straightforward: determine what the most important outcomes for pretrial are (e.g., are they maximizing release rates, and/or ensuring appearance in court, and/or minimizing rearrest during pretrial release?) and then identify and test measures that could be associated with obtaining those outcomes in pretrial settings. Special attention can and should be devoted to the psychometric properties of the new office review process and instrument, including indicators of internal consistency, interrater reliability, and validity. And it will be critical to make sure that districts are able to access the office review instrument in order to self-assess their progress.

This study was instrumental in helping to inform the AO of the performance of national oversight activities in relation to system outcomes. Additionally, system-wide feedback pertaining to office reviews further persuaded the AO to reconsider the long-held and institutionalized processes and procedures governing office reviews. Within the AO, efforts to redesign a national oversight model are underway and will be guided by research and data. The AO is committed to ensuring that national oversight of the federal probation and pretrial services system is guided by the use of evidence-based practices and related research that aims to improve system outcomes. When considering the very question that prompted this study—“whether office reviews in the federal probation and pretrial services system do

what they were intended to do”—perhaps “not yet,” but the AO is on a journey to get there.

References

- Andrews, D. A., & Dowden, C. (2007). The risk-need-responsivity model of assessment and human service in prevention and corrections: Crime prevention jurisprudence. *Canadian Journal of Criminology and Criminal Justice, 49*, 439-464.
- Chandler, H. P. (2015). The future of federal probation. *Federal Probation, 79*(2), 4-9.
- Cohen, T. H., Lowenkamp, C. T., & Hicks, W. E. (2018). Revalidating the federal pretrial risk assessment instrument (PTRA): A research summary. *Federal Probation, 82*(2), 23-29.
- Goldstein, R. (2020). Using implementation science to transform policy into practice: The federal probation and pretrial services system's evidence-based journey. *Federal Probation, 84*(2), 44-49.
- Hughes, J. M., & Henkel, K. S. (2015). Federal probation and pretrial services system since 1975: An era of growth and change. *Federal Probation, 61*(1), 103-111.
- Jalbert, S. K., Rhodes, W., Flygare, C., & Kane, M. (2010). Testing probation outcomes in an evidence-based practice setting: Reduced caseload size and intensive supervision effectiveness. *Journal of Offender Rehabilitation, 49*, 233-253.
- Lowenkamp, C. T., Holsinger, A. M., & Cohen, T. H. (2015). PCRA revisited: Testing the validity of the federal Post Conviction Risk Assessment (PCRA). *Psychological Services, 12*, 149-157.
- Lowenkamp, C. T., Holsinger, A., Robinson, C. R., & Alexander, M. (2014). Diminishing or durable treatment effects of STARR? A research note on 24-month re-arrest rates. *Journal of Crime and Justice, 37*, 275-283.
- Lowenkamp, C. T., Latessa, E. J., & Holsinger, A. M. (2006). The risk principle in action: What have we learned from 13,676 offenders and 97 correctional programs? *Crime and Delinquency, 52*, 77-93.
- Lowenkamp, C. T., Latessa, E. J., & Smith, P. (2006). Does correctional program quality really matter? The impact of adhering to the principles of effective intervention. *Criminology and Public Policy, 5*, 575-594.
- Lowenkamp, C. T., Pealer, J., Smith, P., & Latessa, E. J. (2006). Adhering to the risk and need principles: Does it matter for supervision-based programs? *Federal Probation, 70*(3), 3-8.
- May, T., & Pratt, T. C. (2022). Treating offenders with substance abuse problems: Implicit beliefs about addiction and failed drug tests. *Corrections: Policy, Practice and Research, 7*, 107-119.
- McLean, K., Wolfe, S. E., & Pratt, T. C. (2019). Legitimacy and the life course: An age-
- graded examination of changes in legitimacy attitudes over time. *Journal of Research in Crime and Delinquency, 56*, 42-83.
- Petrich, D. M., Pratt, T. C., Jonson, C. L., & Cullen, F. T. (2021). Custodial sanctions and reoffending: A meta-analytic review. *Crime and Justice: A Review of Research, 50*, 353-424.
- Pickett, J. T. (2019). Public opinion and criminal justice policy: Theory and research. *Annual Review of Criminology, 2*, 405-428.
- Pratt, T. C., & Turanovic, J. J. (2019). A criminological fly in the ointment: Specialty courts and the generality of deviance. *Victims and Offenders, 14*, 375-386.
- Serin, R., Lowenkamp, C. T., & Bourgon, G. (2021). Community Corrections Report Card. Carleton University.
- Sheil, J., Doyle, J., & Lowenkamp, C. T. (2016). Imagining sentinel event reviews in the U.S. probation and pretrial services system. *Federal Probation, 80*(2), 33-37.
- Sloas, L., Murphy, A., Wooditch, A., & Taxman, F. S. (2019). Assessing the use and impact points and rewards across four federal probation districts: A contingency management approach. *Victims and Offenders, 14*, 811-831.
- Smith, A., Heyes, K., Fox, C., Harrison, J., Kiss, Z., & Bradbury, A. (2018). The effectiveness of probation supervision towards reducing reoffending: A rapid evidence assessment. *Probation Journal, 65*, 407-428.
- U.S. Government Accounting Office. (2023). *Pretrial supervision: Actions needed to enhance management of the location monitoring program*. Washington, DC: U.S. GAO, GAO-23-105873.
- Van Deince, T. B., Givens, A., Cowell, M., Ghezzi, M., Murray-Lichtman, A., & Cuddeback, G. S. (2020). A randomized trial of specialty mental health probation: Measuring implementation and effectiveness outcomes. *Administration and Policy in Mental Health and Mental Health Services Research, 49*, 415-428.
- Viglione, J. (2019). The risk-need-responsivity model: How do probation officers implement the principles of effective intervention? *Criminal Justice and Behavior, 46*, 655-673.
- Whetzel, J., & Sheil, J. (2015). Accountability and collaboration—Strengthening our system through office reviews. *Federal Probation, 79*(3), 9-13.
- Wright, K. A., Pratt, T. C., Lowenkamp, C. T., & Latessa, E. J. (2012). The importance of ecological context for correctional rehabilitation programs: Understanding the micro- and macro-level dimensions of successful offender treatment. *Justice Quarterly, 29*, 775-798.
- Zelditch, M. (2006). Legitimacy theory. In Burke, P. J. (Ed.), *Contemporary social psychological theories* (pp. 324-352). Palo Alto, CA: Stanford University Press.