DOI: 10.1111/1745-9133.12635

CRIMINOLOGY & Public Policy

ORIGINAL ARTICLE

Revisiting the effectiveness of HOPE/swift-certain-fair supervision programs: A meta-analytic review

April Pattavina[1] | **Joshua S. Long**[1] | **Damon M. Petrich**[2] | **James M. Byrne**[1] | **Francis T. Cullen**[3] | **Faye S. Taxman**[4]

[1]School of Criminology and Justice Studies, University of Massachusetts Lowell, Lowell, Massachusetts, USA

[2]Department of Criminal Justice and Criminology, Loyola University Chicago, Chicago, Illinois, USA

[3]School of Criminal Justice, University of Cincinnati, Cincinnati, Ohio, USA

[4]Schar School of Policy and Government, George Mason University, Fairfax, Virginia, USA

Correspondence
April Pattavina, School of Criminology and Justice Studies, University of Massachusetts Lowell, Lowell, MA 01854, USA.
Email: April_Pattavina@uml.edu

Abstract

Research Summary: Originated nearly two decades ago in Hawaii by Judge Steven Alm, a community supervision-court model known as "Project HOPE" proposed to reduce probation failure by responding to violations with immediate but short jail terms. Despite negative evidence from Lattimore and colleagues' 2016 Demonstration Field Experiment (DFE) across four locations, advocates continued to trumpet programs based on Project HOPE's core principles of swift, certain, and fair (SCF) sanctions, arguing that these deterrence-oriented interventions—now known under the acronym SCF programs–reduce recidivism. To assess this claim, a meta-analysis was conducted of 18 studies reporting on 24 separate evaluations of programs falling under the Project HOPE/SCF umbrella. The analysis revealed that the intervention had a statistically significant but substantively small impact on recidivism (the main overall effect = −.058). Moderator analyses revealed weak to null findings across variations in methodological and HOPE/SCF program characteristics.

Policy Implications: At present, evaluation evidence is weak and not robust enough to support the contin-

ued government funding and implementation of SCF programs in their current form on grounds of recidivism reduction. Such deterrence-oriented programs may be based on a flawed theory of recidivism that fails to identify criminogenic risk factors for change. SCF programs might prove more effective if integrated with treatment modalities, though this remains to be demonstrated. More broadly, a range of community supervision approaches now exist that emphasize building relationships with individuals under supervision and guiding their prosocial development. These alternatives might offer a more promising avenue for reform than current programs based on SCF principles.

KEYWORDS
community supervision, correctional effectiveness, Project HOPE, swift-certain-fair programs

1 | INTRODUCTION

Project HOPE—Hawaii Opportunity Probation with Enforcement Program—was invented in 2004 by Judge Steven Alm (see Duriez et al., 2014; Hawken, 2010b). This initiative was a response to a perceived problem in the judge's court: Many probationers were at risk of revocation and a prison sentence following multiple violations of supervision conditions for which they had received no consequences. Concerned about their plight, Alm proposed to knife off repeated violations through the use of consistent discipline—much as a caring parent would do for their children. The key was to greet each and every violation—starting with the first one—with an immediate sanction. The goal was to make punishment "swift" and "certain." An advocate of therapeutic jurisprudence, Alm (2016) had no desire to only "get tough" with and inflict pain on deviant probationers. His goal, after all, was to save them from a prison term. The sanction, thus should be proportionate to the infraction—a short, 2-day stay in jail. In this sense, the punishment was "fair." Together, then, the sanctioning philosophy of Project HOPE was governed by three principles: swiftness, certainty, and fairness. Eventually, this reform would shed the HOPE label and be known by its acronym of SCF probation and parole (Kleiman, 2016; Kleiman et al., 2014). Through the use of a meta-analysis of extant studies, the purpose of the current study is to assess the effectiveness of the HOPE/SCF intervention.

1.1 | Advocacy and concern

Project HOPE might have remained a local reform had it not been evaluated by Angela Hawken and the late Mark Kleiman. In an emerging era of evidence-based corrections (Cullen & Gendreau, 2001; MacKenzie, 2006), showing that an intervention "worked" boosted its appeal. Their

design involved a quasi-experimental "pilot program among high-risk probationers" which was expanded to a "randomized controlled trial among general-population probationers." The HOPE group was compared to those assigned to "probation-as-usual" (Hawken & Kleiman, 2009, p. 4). Participants were followed for 12 months. Hawken and Kleiman's (2009) report showed that compared to those receiving "probation as usual" (PAU), HOPE subjects not only were more likely to meet supervision conditions but also had fewer arrests and incarceration days sentenced. Equipped with a novel idea and evidence to support it, Judges Alm, Hawken, and Kleiman were committed to bring this innovation to the US mainland, changing the "H" in HOPE from "Hawaii" to "Honest." To give one example, following his retirement as a judge, Alm moved to the DC area "to work as a legal consultant to states, the Department of Justice and Congress" and to ease the burden of traveling for talks once or twice a month (Blair, 2016). His family and peers noted that Alm "talks endlessly about his creation"—HOPE—and "is a born salesman, a boxer as a kid growing up in Honolulu who doesn't easily give up on what's important to him" (Blair, 2016). He was elected in 2021 as the Prosecuting Attorney of Honolulu.

The final ingredient for the broader expansion of Project HOPE into agencies was money, with federal funding serving as a major driver of the program's expansion. The Justice Department—via the Bureau of Justice Assistance (BJA)—did for HOPE probation what previous federal government administrations did for boot camps, electronic monitoring, intensive supervision programs, residential treatment, and other "get tough" intermediate sanctions initiated in the early 1990s: They funded their expansion by providing grants to state and local agencies. Since 2014, they have provided approximately $15 million in funding to spur the development of these programs in more than 30 states across the country (Fleming, 2021) and the federal government set up a technical assistance center to guide swift and certain grantees. For fiscal year 2022, for example, BJA allocated $4 million in support of new SCF programs, with additional funding provided for a resource center providing technical assistance.

Project HOPE was widely embraced in large part because it was seen as a liberal reform—as a way to keep probationers out of prison by using only a slap on the wrist, albeit one that occurred shortly after every violation. Still, three concerns surfaced about the HOPE model.

First, although Hawken and Kleiman were credible researchers, their evaluation was never published in a peer-reviewed journal. Questions were raised about the study's design (e.g., limited follow-up period and random assignment procedures) and about the methodological details provided (Duriez, et al., 2014). Most significant, advising agencies to implement the HOPE model based on a single study was risky. For one thing, Project HOPE was implemented by a charismatic, caring judge in a unique community. Whether similar outcomes would be produced in other jurisdictions with more punitive jurists and a different racial/ethnic composition of the probation population was unknown. For another thing, the replication crisis in science cautions that initial positive findings often are not reproduced in subsequent studies (Ritchie, 2020). In fact, Hawken et al.'s (2016) follow-up evaluation almost 10-years later revealed less promising results. HOPE probationers still had a significantly lower number of drug charges, but no significant differences were found for property, violent, and social disorder charges (Hawken et al., 2016).

Second, Project HOPE was a variant of specific deterrence and control-oriented supervision (Nagin, 2016). Correctional history provides little evidence that this approach is effective. Most notable, this theory informed the control-oriented intensive supervision programs that emerged in the 1980s. Evaluations, however, showed that watching persons closely and threatening and imposing sanctions did not reduce misconduct. It was only when ISPs included a treatment component that reductions in recidivism were achieved (Byrne et al., 1992; Petersilia & Turner, 1993).

Third, the proposed HOPE model severely limited rehabilitation-type services. Hawken (2010a) noted that only those who failed multiple drug tests would receive treatment—what she referred to as "behavioral triage." Restricting intervention to this group was seen as economical. What was unclear, however, was how services would be provided to high-risk clients who committed misconduct for which no urine test existed (e.g., other crimes). Moreover, no consideration was given to improving the human capital of people on supervision such as through education, social supports, and job training programs. Project HOPE represented a dramatic turn away from the rehabilitative ideal and toward a punishment ideal in which the goal was to suppress noncompliant conduct with swift, certain, and fair sanctions. Scholars cautioned that downplaying treatment and embracing punishment had rarely, if ever, produced good results in corrections (Duriez et al., 2014; see also Cullen & Gilbert, 1982).

One caveat needs to be stated. In reacting to negative experimental evidence on the effectiveness of HOPE (Lattimore, MacKenzie, et al., 2016), Judge Alm (2016) argued that this model was never intended to be a "sanctions-only program"; to see it as such was a "misunderstanding" (p. 1202). In Hawaii, probation officers were trained in evidence-based practices (including cognitive-behavioral therapy, motivational interviewing, and use of a risk-need assessment tool) and officers were encouraged to develop a "therapeutic alliance" with clients (p. 1202). "The HOPE probationers," Alm (2016) observed, "knew I cared about them and wanted them to succeed" (2016, p. 1203).

Although undoubtedly sincere, Alm's attempt to pull HOPE back from a focus on SCF principles—which comprise a sanctioning program—was not embraced by other SCF proponents. The writings of Hawken and Kleiman do not advocate the principles of risk, need, and responsivity—the RNR model that advises to focus interventions on high-risk clients (R), to target for change criminogenic needs or risk factors (N), and to use treatment modalities that are responsive to or capable of changing needs that predict recidivism (R) (Bonta & Andrews, 2017). Neither do they advise training of officers on core correctional practices (Bonta & Andrews, 2017) or the development of empathetic, caring "alliances" or quality relationships with supervisees (see Skeem et al., 2007). As Doherty (2016, p. 291) documents:

> The imposition of swift and certain sanctions, moreover, has eliminated any space for the kind of individualized "concern for the client" relied on by the Supreme Court in *Gagnon*. Kleiman has warned supervisors (whether judges or probation officers) against exercising any favorable discretion for a probationer who has violated a condition covered by the program. He has emphasized: "The temptation on the part of probation officers and judges to cut an erring probationer some slack 'just this once' can be disastrous; when consistency is the name of the game, mercy is toxic." Sanction hearings are to be quick and summary.

1.2 | Evaluation

Despite the three misgivings discussed, Project HOPE remained popular. By 2016, it was estimated that this program had been implemented in 160 jurisdictions, including 31 states (Alm, 2016; Bartels, 2016). In August 2012, a team of researchers headed by Pamela Lattimore embarked on an evaluation study of HOPE. In essence, the goal was to replicate the Hawaii assessment (Hawken & Kleiman, 2009) to determine if these promising findings were robust in the US mainland and in diverse agencies. This project was initiated when John Laub, the noted criminologist, was Director of the National Institute of Justice. At the time, one US Senator was prepared to introduce

legislation mandating HOPE in all 50 states—a bill that Laub recommended delaying until the evaluation data were available (personal communication, John Laub, August 20, 2018). The study was funded by National Institute of Justice in partnership with the BJA, and it carried the title of the Honest Opportunity Probation with Enforcement Demonstration Field Experiment (HOPE DFE) (Lattimore, 2021). The DFE was conducted in four locations in which HOPE probationers were compared with those receiving PAU (Lattimore, 2021; Lattimore, Dawes, et al., 2016; Lattimore, et al., 2018). This study was significant because it was led by unbiased and experienced evaluation researchers who ensured program integrity and because it provided the highest quality experimental evidence.

The findings, which were much anticipated, were published in 2016—again, at a time when HOPE programs were flourishing. Stunningly, the DFE found that Project HOPE had null findings (MacDonald, 2023). Lattimore, Dawes, et al. (2016, p. 2, emphasis in original) concluded that "overall, there were *no significant differences ($p < .05$) between the HOPE and PAU groups in the likelihood of primary outcomes including arrest, probation revocation, or new convictions*." These results led the Department of Justice's technical assistance website, CrimeSolutions.gov,[1] to rate the HOPE program as having "no effects." Despite this fact—and as noted—the program has continued to receive DOJ funding, to be implemented, and to be trumpeted by its advocates (Cullen et al., 2018). DOJ also funds a technical assistance center to help sites implement HOPE.

Facing disappointing results seemingly falsifying the HOPE model, Judges Alm, Angela Hawken, and Mark Kleiman chose not to retreat but to continue to promote this approach (Cullen et al., 2018). Their responses were nuanced but also seemingly split the alliance into two divergent pathways (Cullen et al., 2018). Departing from his co-advocates and as mentioned above, Judge Alm (2016) argued that the programs tested in the DFE did not include a sufficient commitment to rehabilitation, as intended in his original HOPE model. By contrast, Hawken (2016) and Kleiman (2016) doubled down on the punishment principles of SCF. Beyond the narrow DFE test of Project HOPE, they noted that it was essential to consult a broader literature that assesses—favorably—programs using SCF principles. "A variety of previous reports," noted Kleiman (2016), "have shown success in programs implementing SCF principles" (p. 1185). Going forward, they chose to use the Swift-Certain-Fair phrase to label their intervention—not "HOPE," which they see as falling under the umbrella of SCF supervision. The Bureau of Justice Assistance now has a website titled *Swift, Certain, and Fair Supervision*. SCF principles have been defined in this way (see BJA, 2018):

- **Swiftness**—responding to behavior promptly so that people under supervision connect the response to their behavior.
- **Certainty**—ensuring that sanctions and rewards are applied with consistency and predictability.
- **Fairness**—making sanctions proportionate to negative behavior and rewards appropriate to positive behavior.

1.3 | Current study

In light of the fact that these programs continue to receive substantial public funding and are being widely implemented, the goal of the current study is to take stock of the available evidence on their effectiveness. In doing so, we use meta-analytic techniques to synthesize evaluations of Project HOPE and other probation programs based on the SCF principles. The policy implications of our findings are explored in the final section of the article.

It should be noted that Cullen et al. (2016) conducted a limited meta-analysis of Project HOPE studies. This work was undertaken in response to an invitation by then-editor Daniel Nagin to provide a commentary on evaluations to be published in *Criminology & Public Policy*. As a commentary as opposed to a research article, the analysis was intended to provide suggestive results. The sample of studies included "the analyses of the Hawaii program (Hawken and Kleiman, 2009; Hawken et al., 2016), the evaluation from Delaware (O'Connell et al., 2016), the one from Washington State (Hamilton et al., 2016), and the pooled four-site evaluation by Lattimore, MacKenzie, et al (2016) ($N = 4$ independent datasets)" (Cullen et al., 2016, p. 1219). The findings were disappointing, with the Project HOPE overall mean effect size computed to be .070, with a confidence interval ranging from –.039 to .181. The effect size for studies outside Hawaii (from "elsewhere") was only .033 (p. 1219). When Cullen and colleagues presented the findings at the 2016 meeting of the American Society of Criminology, panel discussants Andrea Hawken and Mark Kleiman noted—credibly—that the results were suspect because the meta-analysis did not include all SCF studies, many of which did not fall under the formal umbrella of HOPE and arguably reported positive effects (see also Kleiman, 2016). The current project takes up this challenge of reporting a comprehensive meta-analysis of relevant evaluation studies.

Subsequently, Cullen et al. (2018) undertook another limited assessment, this time examining SCF studies that Hawken and Kleiman cited in their writings as showing the effectiveness of these principles. They tracked down 20 studies and then judged whether a program's "HOPE treatment effect" was statistically significant and differed appreciably from the overall effect size of .033–.070 reported in the Cullen et al. (2016) meta-analysis. They concluded that only one study met these criteria, leading them to confirm their view that "HOPE-style supervision is unlikely to achieve meaningful change in behavior" (Cullen et al., 2018, p. 20). Although these findings were consistent with the extant research, the study lacked rigorous standards for including past evaluations in the analytic sample, did not compute precise effect sizes, and did not explore moderator effects. The current project addresses these prior limitations.

2 | METHOD

2.1 | Literature search and analytic sample

To conduct the systematic review and meta-analyses presented in subsequent sections of the current article, the research team gathered relevant studies via Google Scholar, Google search engine, and university libraries. Figure 1 provides a visual overview of the search and study selection processes, including (1) study identification, (2) screening, (3) eligibility, and (4) selection or inclusion. Databases were searched using combinations of the keywords "HOPE" and "probation." Over 191,000 studies were initially identified, though few of these referred to the Hawaii Opportunity Probation with Enforcement model of community supervision (many of these were articles that included the verb "hope"). Records were reduced to a manageable number when the search term "opportunity probation" was paired with "HOPE." We also identified an additional 63 articles through searches of the reference sections of major HOPE evaluations. This allowed us to discover any studies that might have been overlooked in initial searches. At this stage, only peer-reviewed academic articles and evaluations published on government websites were retained. When duplicate studies were removed, the titles and abstracts of 255 research articles were screened for relevancy. Of these, 237 were excluded because they were not evaluations

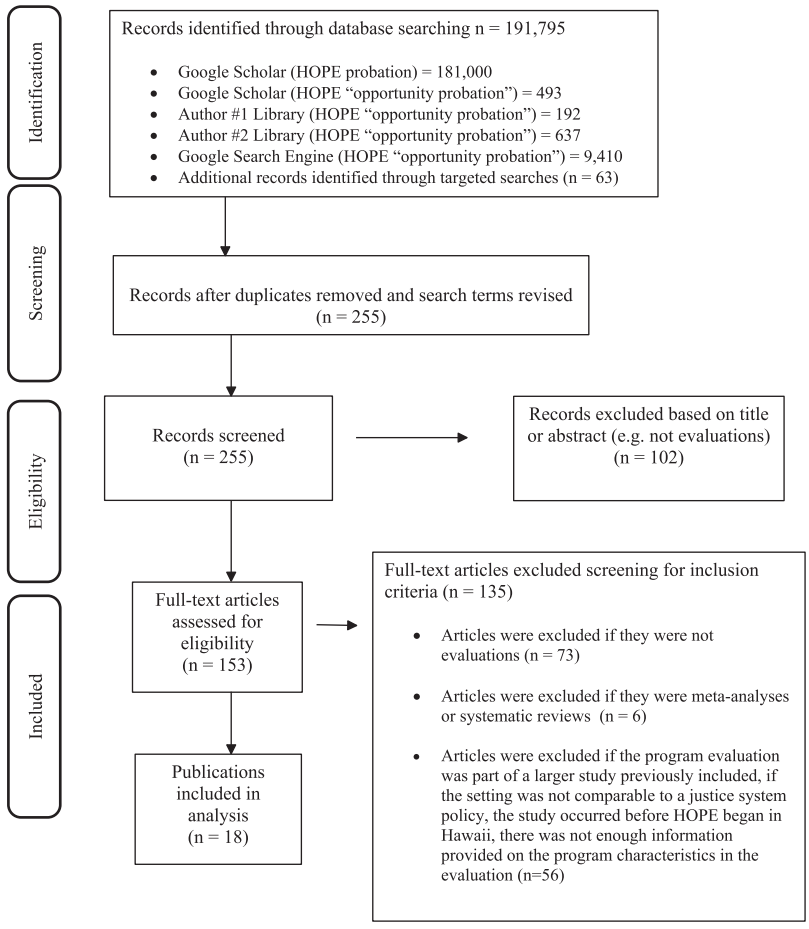

FIGURE 1 Publication inclusion flowchart.

of HOPE (for example, they may have cited HOPE or SCF interventions in a literature review), they were meta-analyses or systematic reviews that did not include an original evaluation, or they were reporting the results of a larger study that was included in our analysis. The remaining 18 studies were found to be evaluations of HOPE or SCF interventions that contained enough information about each program to be included in our analysis.[2] (For a summary of these studies, see Supporting Information Tables S1 and S2.)

The 24 evaluations deemed eligible for our review included a total of 451 experimental and quasi-experimental comparisons of HOPE/SCF supervision to probation-as-usual or otherwise. Forty-five of these comparisons did not convey enough information to calculate an effect size. The final analytic sample was thus comprised of 406 effect size estimates. Each study contributed between 1 and 51 effects given that most included analyses of multiple types of recidivism, varying follow-up lengths, and comparisons of different groups within individual primary studies. Two of the study authors coded each study for details on effect size estimates and other study details on sample characteristics of participants, research methods, recidivism outcome type, follow-up period, and model details as potential moderators. The intercoder reliability rating for study

characteristics after an initial pass was 86%. The two coders discussed any discrepancies and reached agreement on the final codes for each study.

2.2 | Description of studies

The analytic sample for the current study comprised 18 studies of HOPE or SCF programs that began after the Hawaii HOPE program was launched in 2004. Three of these studies provided evaluation details on more than one independent sample. The Hawken and Kleiman (2009) and Hawken et al. (2016) studies included results from two separate evaluations, and the Lattimore et al. (2018) study included four experimental site evaluations and an overall multijurisdictional evaluation bringing the total to 24 evaluations that we included in the analysis.

Table 1 displays the design characteristics of these evaluations. Thirty-eight percent used true experimental designs (i.e., randomized control trials [RCTs]), while the remaining used some type of quasi-experimental design. Table 1 also illustrates variability in several other characteristics of the primary evaluations included in the current review. There is variation in the size of the samples used in the 24 evaluations we included in our review. Most evaluations included multiple types of outcome measures, with 83% of studies evaluating arrest as a recidivism or outcome measure, 67% examining probation violations, 58% for incarceration, and 38% evaluating substance use as an outcome measure. In terms of the length of longest follow-up for measuring these recidivism outcomes, 33% of the evaluation studies used a follow-up period of 12 months or less, 17% used between 13 and 24 months, and the remaining 50% of studies tracked participants' recidivism for more than 2 years. Finally, 13% the included evaluations were reported in peer-reviewed journal articles, whereas the remaining 87% came from non-peer-reviewed technical reports.

Part of our review of evaluation studies involved an assessment of overall design quality. We used criteria from the University of Maryland's Research Review Protocol to do so.[3] The Ranking column displays the results of this assessment and illustrates that 50% of the 24 included evaluations were classified as Level 2 studies, due primarily to questions about the construction and characteristics of comparison groups. Forty-two percent were classified as either Level 4 or 5 and 8% were classified as Level 3 according to our use of the assessment tool. As for the RCTs, evaluations by Hawken and Kleiman (2009) and Hawken et al. (2016) were classified as Level 4 studies due to incomplete details on study procedures. Full information on the research methods utilized in these two studies has never been released. The most we were able to find on randomization procedures, for example, was a brief description in the introduction to the report by Hawken and colleagues (2016):

> A study group of 507 probationers was identified by the probation officers. Of this group, 493 were deemed eligible for inclusion in the study by probation-office supervisors. Third-party batch randomization assigned eligible subjects to HOPE or probation-as-usual (control group). The study groups were well balanced; demographic profiles show no significant differences in age, sex, or race/ethnicity. LSI scores and prior arrests were used to assess any differences in baseline risk. The average baseline LSI score for HOPE probationers was higher than for control probationers and there was no meaningful difference in the number of prior arrests across the study groups or differences in most-serious prior charge (pp. 8–9).

There is also likely variation in the comparability of the treatment and control groups identified in the quasi-experimental studies we have included. As noted in Table 1, only four of

TABLE 1 Study design characteristics.

	Design type*	Treatment N**	Comparison N	Ranking***	Outcomes	Follow-up (months)	Publication type
Hawken and Kleiman (2009)[1]	Basic Match[b]	940	77	2	Drug use, Incarceration, Probation violation	≤12	Report
Hawken and Kleiman (2009)[2]	Experimental	330	163	4	Arrest, Probation violation	≤12	Report
Loudenburg et al. (2010)	Basic Match[b]	1231	29,218	2	Arrest	25+	Report
Grommon et al. (2012)	Experimental	281	248	5	Drug use, Arrest, Incarceration	13–24	Article
Shannon (2013)	Basic Match[b]	307	300	2	Drug use, Arrest, Probation violation	≤12	Report
Lattimore et al. (2018)[1,2,3,4,5,a]	Experimental	743	761	5	Drug use, Arrest, Probation violation, Incarceration	25 +	Report
Hawken et al. (2016)[1]	Basic Match[b]	34	78	2	Arrest, Incarceration	25+	Report
Hawken et al. (2016)[2]	Experimental	330	163	4	Arrest, Incarceration, Probation violation	25+	Report
O'Connell et al. (2016)	Experimental	184	200	5	Drug use, Incarceration, Arrest, Probation violation	13–24	Article
Hamilton et al. (2015)	Propensity Score Matching	4,838	4,838	4	Arrest, Prob/parole violation, Incarceration	≤12	Report
DeVall et al. (2015)	Propensity Score Matching	379	379	3	Arrest, Incarceration	13–24	Report
Midgette and Kilmer (2021)	Basic Match[b]	356	2412	2	Arrest	≤12	Article

(Continues)

TABLE 1 (Continued)

	Design type*	Treatment N**	Comparison N	Ranking***	Outcomes	Follow-up (months)	Publication type
Snell (2007)	Basic Match[b]	371	681	2	Arrest, Probation violation	25+	Report
VA Sentencing Commission (2016)[c]	Propensity Score Matching	63	63	2	Arrest, Probation violation	≤12	Report
Olson and Stemen (2014)	Basic Match[b]	344	332	2	Incarceration	25+	Report
NC Dept Public Safety (2014)[f]	Basic Match[b]	368	368	2	Probation violation	25+	Report
Edwards and Pettigrew (2016)	Propensity Score Matching	1198	1198	3	Incarceration, Probation violation	≤12	Report
MI Sup. Court (2020)[f]	Basic Match	1003	1003	2	Arrest	25+	Report
Frailing et al. (2018)	Basic Match[b]	69	84	2	Arrest, Probation violation	13–24	Report
Schlueter et al. (2011)	Basic Match[b]	89	102	2	Arrest, Probation violation	≤12	Report

[a]Lattimore et al. (2018) report included five evaluations with same design; Massachusetts, Texas, Arkansas, Oregon, and combined, noted as 1–5.

[b]Publication compared two or more groups but did not control for group differences or use matching techniques in evaluation.

[c]State abbreviations, VA = Virginia, NC = North Carolina, MI = Michigan.

*Experimental signifies random assignment.

***Note*. Sample sizes varied by outcome in some studies. The largest treatment group comparison is presented in this table.

***Rankings based on Sherman et al. (1998) (1 = low, 5 = high): Level 1. Correlation between a crime prevention program and a measure of crime or crime risk factors at a single point in time. 2. Temporal sequence between the program and outcome clearly observed, or the presence of a comparison group without demonstrated comparability to the treatment group. Level 3. A comparison between two or more comparable units of analysis, one with and one without the program. Level 4. Comparison between multiple units with and without the program, controlling for other factors, or using comparison units that evidence only minor differences. Level 5. Random assignment and analysis of comparable units to program and comparison groups.

the 15 designated quasi-experimental studies used state-of-the-art propensity score matching techniques, while the remaining either used basic matching techniques with a handful of covariates or did not use matching techniques at all. We think it is safe to assume that the results of many quasi-experimental studies could be challenged on methodological grounds. In quasi-experimental research, it is essential to create treatment and comparison groups that have similar distributions on as many relevant confounders as possible. Without doing so, the two groups will likely differ in important ways and any estimates of the effects of the treatment program will thus be biased (Guo & Fraser, 2015; Rosenbaum & Rubin, 1983).

Table 2 provides details on the characteristics of the actual programs examined in the primary studies included in the current review. Based on the criteria laid out by the developers of the original HOPE model, as well as the criteria employed in the multisite replication, programs were evaluated on the extent to which sanctions could be classified as swift, certain, and fair, using criteria that we believe reflects the intentions of the original program developers.[4] Sanctions in 71% of the 24 evaluation studies were classified as being "Swift" on the basis of infractions generally being dealt with within 7 days. The sanctions in each of the 24 evaluations were classified as "Certain," given that all detected infractions were reportedly met with a sanction. Finally, programs in 16% of studies were clearly classified as "Fair, with sanctions of 2 days or less in jail." In the remaining studies, the length of sanctions could include more than 2 days in jail or was unreported and therefore cannot be assumed to meet the criteria for fairness. We recognize that this is a limited operationalization of fairness. However, advocates have argued that a core component of the fairness principle is the "*parsimonious* use of punishment" so as to enhance "the legitimacy of the sanction package" (Hawken, 2010b, p, 41, emphasis in original). The goal is to provide certain, not severe, punishment that is seen as just. Studies that reported more than 2 jail days varied in length of stay for first sanction and some would increase the number of days for repeated violations making it difficult to provide a consistent measure of what may be considered fair across studies. Under these circumstances, we applied the 2-day recommendation by Kleiman et al. (2014) in this review which may be considered conservative among those advocating jail stays as a sanction.

Table 2 also illustrates considerable between-study variation in both the target populations and the types of offenses excluded from participation in HOPE/SCF programming. For example, in 42% of the evaluations, individuals classified based on risk level comprised the target population, while in 50% of studies substance abusing probationers were the target population. The remaining 8% included one program targeting domestic violence probationers and one targeting individuals placed on felony probation. Within each of these categories of inclusion, there are otherwise eligible persons who were be excluded based on offense type. As highlighted in Table 2, 46% of the programs excluded all sex offenses, 33% excluded all violent offenses, and 21% of the programs did not mention any offense exclusions.

Each study was also coded for characteristics of the populations included. Programs in 87% of the separate studies included solely targeted adult populations, while the remaining 13% appeared to include both juveniles and adults. Samples in all but one study (95%) comprised both male and female individuals.

2.3 | Meta-analytic approach

We used a variety of formulae (see, Lipsey & Wilson, 2001; Petrich et al., 2021) to standardize all effect sizes to the Pearson's correlation coefficient (i.e., *r*). These were further converted

TABLE 2 Program characteristics.

	Type	Swift[a]	Certain[b]	Fair[c]	Target population	Exclusions	Age	Gender
Hawken and Kleiman (2009)[1]	HOPE*	Yes	Yes	Yes	Substance abusing probationers	Violent	Adult	Mixed
Hawken and Kleiman (2009)[2]	HOPE*	Yes	Yes	No‡	Substance abusing probationers	Violent	Adult	Mixed
Loudenburg et al. (2010)	SCF*	No	Yes	No‡	DUI Probationers	Violent	Mixed	Mixed
Grommon et al. (2012)	SCF*	Yes	Yes	No†	Intensive supervision parolees	None	Adult	Male
Shannon (2013)	HOPE*	Yes	Yes	No‡	Intensive supervision probationers	Violent	Adult	Mixed
Lattimore et al. (2018)[1-5d]	HOPE*	Yes	Yes	No†	Substance abusing probationers	Sex Offenses	Adult	Mixed
Hawken et al. (2016)[1]	HOPE*	Yes	Yes	Yes	Substance abusing probationers	Violent	Adult	Mixed
Hawken et al. (2016)[2]	HOPE*	Yes	Yes	No‡	Substance abusing probationers	Violent	Adult	Mixed
O'Connell et al. (2016)	HOPE*	No	Yes	No†	Substance abusing probationers	Sex Offenses	Adult	Mixed
Hamilton et al. (2015)	SCF*	Yes	Yes	No†	High-risk probationers and parolees	None	Adult	Mixed
DeVall et al. (2015)	HOPE*	Yes	Yes	No†	Intensive supervision probationers	Sex Offenses	Adult	Mixed
Midgette and Kilmer (2021)	SCF	Yes	Yes	Yes	DUI arrestees	Sex Offenses	Adult	Mixed
Snell (2007)	SCF*	No	Yes	No	High- or medium-risk probationers	Sex Offenses	Adult	Mixed
VA Sentencing Commission (2016)	HOPE	Yes	Yes	No	Intensive supervision probationers	Violent	Adult	Mixed
Olson and Stemen (2014)	HOPE	No	Yes	No	Intensive supervision probationers	Violent	Adult	Mixed
NC Dept.Public Safety (2014)	SCF	No	Yes	No†	Intensive supervision probationers	None	Mixed	Mixed
Edwards and Pettigrew (2016)	SCF	Yes	Yes	No†	High Risk Probationers	None	Mixed	Mixed
MI Supreme Court (2020)	HOPE*	No	Yes	Yes	High Risk Probationers	Sex Offenses	Adult	Mixed
Frailing et al. (2018)	HOPE	No	Yes	No‡	Felony Probationers	None	Adult	Mixed
Schlueter et al. (2011)	SCF	Yes	Yes	No‡	DV Probationers	Sex Offenses	Adult	Mixed

*Cited by Hawken et al. (2016) as an example of HOPE or SCF-comparable program.

[a]A program is defined as "Swift" if detected infractions are sanctioned within 7 days.

[b]A program is defined as "Certain" if every infraction receives a sanction.

[c]A program is defined as "Fair" if infractions are sanctioned with no more than 2 days in jail.

[d]Lattimore et al. (2018) included five evaluations in one report, four from separate states and one combined, we noted as 1–5.

†Jail sanction was more than 2 days on average.

‡Number of jail sanction days were not reported, and therefore, cannot be assumed to meet the criteria for fairness.

to $z(r)$ scores using Fisher's r to z transformation. This transformation is useful given that the distribution of $z(r)$ scores more closely resembles a normal distribution, while the distribution of untransformed r scores does not. The models employed in this study assume a normal distribution of the outcome variable for accurate estimates of standard errors and null hypothesis significance testing (Pratt et al., 2014). All effect size conversions were computed using the Comprehensive Meta-Analysis software application.

Multilevel linear models (MLMs) were employed in the current study to account for the interdependency of effect size estimates and the resultant clustering of error (Hox et al., 2018; see, more generally, Raudenbush & Bryk, 2002). The MLMs presented in the subsequent section considered variance in effect size estimates at three levels. Level 1 of the equations accounted for the known variance of participants using dispersion estimates presented in primary studies (e.g., standard errors); Level 2 accounted for variance between effect sizes within each primary study; and Level 3 of the equations accounted for variance between primary studies (Hox et al., 2018; for applications, see Petrich et al., 2021; Pyrooz et al., 2016). All multilevel analyses were conducted in the R software environment using the metafor package (Viechtbauer, 2010).

2.4 | Moderators

We were interested in estimating the overall effect of SCF programs on recidivism as well as exploring the extent to which the effect size is moderated by study characteristics. The moderator variables resulting from the coding process represent study characteristics at both Levels 2 and 3 of the equation described above. In terms of the Level 2 moderators—those that capture within-study variability—the first moderator was *follow-up period*. The follow-up period was 12 months or less in 24% of effect sizes, 14% had follow-up periods between 13 and 24 months, and 62% had follow-up periods longer than 24 months. This variable was coded as categorical with 12 months or less as the reference. The moderator variable *recidivism* captured the type of reoffending measured in primary studies. Seven percent of effect sizes pertained to negative drug outcomes (i.e., failed or missed drug tests), 52% were for new arrests or convictions, 28% for technical violations and revocations, and 13% for incarcerations. This variable was coded as categorical with drug use outcome as the reference. *Sample size* for each effect size ranged from 30 to 30,449 and we included a logged version for the analysis. *Type of statistical model* describes effects estimated in either a multivariate (14%) or bivariate context (86%) and was coded as 0 = bivariate and 1 = multivariate.

Level 3 moderators capture between-study characteristics that might explain variability in effect sizes. The moderators we included and their distributions were described in the previous section and were coded as follows. Since most studies did not clearly meet the "fair" aspects of the program, we created a dichotomous measure—*Swift and Certain*—to capture programs that adhered to swift and certain features which was coded as 0 for no and 1 for yes. *Publication type* covers journal articles and was coded as 0 for report and 1 for journal article. *Research design* categories, include basic matching, propensity score matching, and RCT designs, and were coded as categorical with RCT as the reference category.

Offense exclusions include sex offenses only, violent and sex offense exclusions, and no exclusions and was coded as categorical with no exclusions as the reference. *Special Populations* identifies studies that focus on specific types of offenses and coded as 0 for special offense targeted and 1 for no special target). *Risk level* of participants includes all risk levels, medium to high risk, and not reported and was coded as categorical with all levels as the reference. *Age* was coded

as 0 for mixed age and 1 for adults. *Gender* categories cover mixed gender, males only and not reported and was coded as categorical with males as the reference category. The *number of effects* ranged from 1 to 51 with an average of 17. This variable was centered on the mean for the analysis.

Finally, the analysis did not include a measure of "experimenter allegiance" (EA), which is defined as a researcher's "personal belief both in the superiority and efficacy of a particular treatment" (Dragioti et al., 2015, p. 1) and has the potential for "results to be contaminated or distorted by the investigators' theoretical or treatment preferences" (Wilson et al., 2011, p. 119). Research has reported conflicting results on this bias including that it may not pertain to evaluations of cognitive behavioral therapy (Dragioti et al., 2015; Gaffan, 1995; Wilson et al., 2011). If bias exists, it is usually to inflate the effect size of the intervention being assessed—in this case the HOPE/SCF program. To facilitate EA analysis, Dragioti et al. (2015) have urged that theory or modality allegiance be disclosed in authors' conflict of interest statement, but this was not a practice undertaken by evaluation authors included in this study. While we may have confidence in coding evaluations by Hawken and Kleiman as EA, determining the allegiance of other evaluators we included was problematic since many of the studies did not provide interpretation of HOPE context and background sufficiently enough to gauge allegiance. Thus, a moderator analysis of EA was not undertaken.

3 | META-ANALYTIC FINDINGS

The first step in the analyses was to estimate the overall effect of HOPE/SCF programming on recidivism outcome across the entire sample of effect size estimates. The mean overall effect from this initial model was −.058 ($SE = .020$, $p < .01$), with a 95% confidence interval ranging from −.098 to −.019. The direction of this effect indicates that HOPE/SCF programs reduce recidivism. However, the effect is substantively small according to the seminal guidance of Cohen (1988, 1992), who suggested that correlation coefficients of .10, .30, and .50 should be considered small, medium, and large effect sizes, respectively. We elaborate on the implications of the magnitude of this effect in greater detail in Section 4 of this article.

The next step of the analysis involved testing the significance of the Levels 2 and 3 variance components. This was assessed through two separate log-likelihood ratio tests, where the fit of an intercept-only model was compared to (1) a model in which the variance at Level 2 was fixed and (2) another model in which both Levels 2 and 3 variances were fixed. These analyses indicated that there was significant variability of effect sizes both within studies (Level 2) and between studies (Level 3). For a visual depiction of variability in the within-study mean effect sizes, see the forest plot in Figure 2. Following the work of Cheung (2014), we estimated that the Level 1 sampling variance accounted for 5% of the variation in effect sizes. Eighty percent of the variance in effect sizes is attributed to variation within studies at Level 2 and the remaining 15% of the variance occurred between studies at Level 3. These results suggest that moderator analyses were an appropriate next step.

The first set of moderator analyses was conducted by entering each individual moderator variable into a separate MLM to test its bivariate association with effect size estimates. The findings from these models are presented in Table 3. The top panel of the table reports estimates of the effects of Level 2 moderator variables. The analyses indicated that sample size does not have a significant effect on recidivism. Results varied by type of recidivism outcome. Those evaluations examining outcomes other than drug use also tended to find stronger effects. For example, the mean effect in studies with drug-use outcomes was 0.112. A coefficient of −.207 ($SE = .036$, $p < .001$) for arrest or conviction indicates that the mean effect of HOPE/SCF programs on these

TABLE 3 Bivariate moderator analysis results.

	Coefficient	*SE*	*t*-value	Intercept
Level 2 Moderators				
Log sample size	.015	.012	1.268	−.164
Recidivism outcome				
Drug				.112**
Arrest/conviction	−.207***	.036	−5.751	
Probation violation	−.169**	.037	−4.518	
Incarceration	−.166***	.039	−4.211	
Length of follow-up				
12 months or less				−.112***
13 to 24 months	.034	.032	1.068	
25 or more months	.109***	.031	3.461	
Multivariate				
Bivariate effect				−.056**
Multivariate effect	−0.018	.027	−0.690	
Level 3 Moderators				
Research design				
RCT				.013
Basic match	−.139***	.024	−5.769	
PSM	−.140***	.037	−3.748	
Offense exclusions				
No exclusions				−.063*
Violent offenses	−.086*	.037	−2.334	
Sex offenses	.075*	.031	2.422	
Age				
Mixed				−.168**
Adult	.128*	.058	2.205	
Number of effect sizes	.003**	.001	3.004	−.081***
Swift and Certain				
Yes				−.059*
No	−.002	.044	−.042	
Publication type				
Technical report				−.067**
Journal article	.054	.059	.905	
Gender				
Males				−.007
Mixed	−.056	.085	−.659	
Special populations				
Yes				−.116*
No	.069	.059	1.165	

(Continues)

TABLE 3 (Continued)

	Coefficient	*SE*	*t*-value	Intercept
Risk				
All levels				−.069*
Medium-to-high	.038	.042	.913	
Not reported	−.045	.070	−.644	

Abbreviation. SE, standard error.
***$p < .001$; **$p < .01$; *$p < .05$.

outcomes was −.095. For length of follow-up, the bivariate moderator analyses suggest that the effects of HOPE/SCF programs on recidivism may dwindle over time. The intercept of −.112 indicates a small recidivism-reducing effect when follow-ups were 12 months or less. The coefficient of .034 in the next row, while not statistically significant, indicates that these effects start to weaken as follow-ups extend to between 13 and 24 months. This trend continues as follow-ups extend further to 25 or more months. This time, a coefficient of .109 ($SE = .031$, $p < .001$) points toward a null effect on recidivism. Whether the effect size was based on multivariate versus bivariate statistical model had no impact on recidivism.

Turning to the Level 3 moderators, several of these variables had no significant association with effect size estimates. In particular, the extent to which programs adhered to the swift, certain, and/or fair principles had no impact on effect sizes, nor did the type of publication in which the results were reported, the gender of participants, their risk level, and the targeting of special populations. Several Level 3 models did, however, reveal significant associations with effect size estimates. In terms of research design, the nonsignificant intercept for this model (.013, $p > .05$) illustrates that HOPE/SCF programs had no discernible effect on recidivism outcomes when programs were evaluated using the strongest research designs, RCTs. The other two rows for this variable show that stronger program effects are reported when the evaluations used weaker designs such as basic matching (−.139, $p < .001$) or propensity score matching (−.140, $p < .001$). Offense exclusion also had a moderating influence on the recidivism effect. Programs that excluded all violent offenses appear to be more effective at reducing recidivism (−.086, $p < .05$) than those that had no exclusions, but those with only sex offense exclusions had weaker effects than those with no exclusions (.075, $p < .05$). Evaluations of programs that included only adults had weaker effects than those that included mixed ages and those reporting larger numbers of effect sizes had weaker effects on recidivism.

The next step of our moderator analyses was to enter all significant variables from the previous bivariate analyses into a multivariate, meta-regression model. Variance inflation statistics were assessed for potential multicollinearity. Most variance inflation statistics had values lower than 5 with exception of age (5.96). This value is slightly higher than the cut-off of 4.00 suggested by Fox (1991), but much lower than others who recommended a cut-off of 10 at which multicollinearity is considered a serious problem (Menard, 2001). Age was thus retained in the metaregression model. The results of this model are presented in Table 4. The intercept of this model (.060, SE = .088, $p > .05$) suggests a nonsignificant impact of HOPE/SCF programs when all included moderators are set to their reference categories. After controlling for other variables in the model, age, research design, offense-type exclusions and number of effect sizes had no significant association with effect size estimates. Recidivism measures with a follow-up of 25 months or more showed weaker effects than those with a follow-up of 12 months or less. For outcome type —effects remain significant and follow the same patterns as in the bivariate analyses. In sum, HOPE/SCF

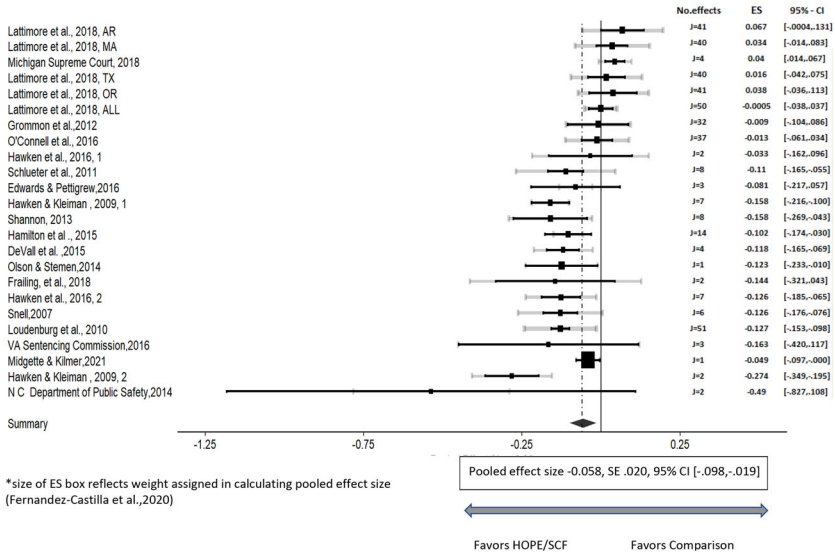

FIGURE 2 Forest plot of HOPE/SCF program evaluations on recidivism. [Color figure can be viewed at wileyonlinelibrary.com]

programs are less effective at reducing recidivism for drug use than other recidivism outcomes and any impact is likely to be short term.

3.1 | Supplementary analyses

As discussed above, our main meta-analysis considering the entire sample of effect size estimates suggested that participation in HOPE/SCF has a small beneficial effect on reoffending outcomes ($r = -.058$, $p < .01$). Moderator analyses revealed some potentially important factors that account for variations in effect sizes. Perhaps most important among these was the type of research design employed to evaluate programs, although several other moderators emerged as significant as well. To probe these findings further, we conducted another set of bivariate moderator analyses in which we split the sample into two groups: (1) effect sizes from non-RCT studies and (2) effect sizes only from RCT studies. The results of these analyses are reported in Appendix A, Tables A1 and A2, respectively. For the non-RCT subsample, an intercept-only model suggested a small beneficial effect of participation in HOPE/SCF, with an effect size of −.117 ($SE = .017$, $p < .001$). Few of the moderating variables had a significant effect on effect size estimates within this sample, meaning that these studies find that HOPE/SCF reduces offending outcomes regardless of other methodological characteristics of the studies. However, effects did tend to grow weaker with longer follow-up periods and with larger sample sizes.

For the RCT-only sample, an intercept-only model indicated HOPE/SCF group status had an overall null effect on reoffending outcomes ($r = .014$, $SE = .012$, $p > .05$). When the type of recidivism outcome was entered as a moderating variable, analyses indicated that participation in the program resulted in an increased number of drug test failures. Effect sizes for arrest or conviction, probation violations, and incarceration all differed significantly from drug test-related effect sizes, each landing close to or slightly below zero (i.e., a null effect). The program also tended to

TABLE 4 Multivariate meta-regression results.

	Coefficient	*SE*	*t*-value
Intercept	.060	.088	.679
Age			
Adult (ref)			
Mixed	.021	.063	.343
Research design			
RCT (ref)			
Basic match	−.001	.045	−.029
PSM	−.008	.057	−.145
Offense Exclusions			
No exclusions (ref)			
Violent offenses	−.072	.047	−1.520
Sex offenses	.033	.040	.822
Number of effect sizes	.001	.001	1.232
Outcome			
Drug (ref)			
Arrest/conviction	−.211***	.035	−5.877
Probation violation	−.175***	.037	−4.707
Jail	−.162***	.039	−4.107
Follow-up			
12 months or less (ref)			
13–24 months	.001	.030	.025
25+ months	.063*	.030	2.114
Variance components	LR		
Level 2	2869***		
Level 3	.344		

Abbreviations. LR, likelihood ratio test; SE, standard error.
***$p < .001$; **$p < .01$; *$p < .05$.

have weaker effects in evaluations with a larger number of statistical models and stronger effects when violent offenses were excluded from participation. Analyses with the RCT-only sample also revealed that effect sizes did not vary significantly on the basis of sample size, follow-up period, adherence to swift, certain and fair principles and the risk level of participants (see Appendix A, Table A2).

For other supplementary analyses, we evaluated the possibility for publication bias by conducting an Egger's test, which is a commonly used quantitative method that tests for asymmetry in the funnel plot. Akin to a visual inspection of a funnel plot, the test can only identify small study effects and does not directly indicate whether publication bias exists. The test is based on a simple linear regression model where the effect sizes are divided by their standard error, the result of which is regressed on its precision measured as the inverse of the standard error. The intercept for the model was not significant, indicating that there were no small study effects.

We also conducted several sensitivity tests. First, we applied a robust variance estimation procedure as an additional way to account for dependent data (i.e., multiple outcomes from single

studies) (Harrer et al., 2022). Standard errors were corrected using the bias-reduced linearization estimator described in Tipton and Pustejovsky (2015). Results in this model with robust variance estimation were unchanged. Second, we ran the initial model after removing the outlier study by the North Carolina Department of Public Safety. The overall effect was slightly smaller than the one estimated with the full dataset (−.050), but still significant and thus we did not exclude it from our reported results.

4 | DISCUSSION

The Project Hope/SCF model had great appeal when initially formulated in Hawaii, and now, two decades later, this intervention continues to receive government funding and to be implemented in the United States and beyond. Despite the common-sense appeal of the program—how can consistently holding someone accountable for misconduct not reduce negative outcomes —the empirical evaluations of this model have been unfavorable (Duriez et al., 2014). Negative findings have been reported in experimental studies (Lattimore, Dawes, et al., 2016; Lattimore, MacKenzie, et al., 2016), in a small-sample meta-analysis (Cullen et al., 2016), and in a review of the literature (Cullen et al., 2018). One rebuttal, however, is that these assessments ignore the many programs that are localized and embody the principles of SCF punishment (Hawken, 2016; Kleiman et al., 2014). The current meta-analysis was undertaken to address this concern by including evaluations of SCF and replications of HOPE programs in the sample, as well as documenting which of the underlying principles of deterrence are measured in the SCF/HOPE programs. The latter is important because it gives meaning to the label and provides the ability to state with confidence what are the features associated with swift, certain, and/or fairness.

Consistent with prior research, our meta-analysis does little to reverse the existing assessment that HOPE/SCF programs have minimal effects on recidivism, and that the core concepts of swift, certain and fair are not well articulated. The main analysis indicated that these programs have a beneficial impact on reoffending. However, although this effect is statistically significant, its size is substantively limited ($r = -.058$, $p < .01$) according to conventional standards for judging the strength of effect sizes. These standards converge on the suggestion that correlations of .10 should be considered small in magnitude (e.g., Cohen, 1992; see also Brydges, 2019). Further, both bivariate and multivariate moderator analyses illustrated that participation in HOPE/SCF programs has a nonsignificant statistical effect on reoffending outcomes when evaluation studies use the strongest research designs—randomized controlled trials. Supplementary analyses strengthened these conclusions, with moderator analyses restricted to RCT-derived effect sizes showing that the program generally has a null effect on arrests or convictions, probation violations, and incarceration and across most other methodological conditions (see Appendix A, Table A2). Participation in HOPE/SCF did, however, result in more drug test failures, which is not surprising given the nature of the program. To the contrary, analyses of effect sizes derived from studies with less rigorous research designs generally conclude that the HOPE/SCF program model is beneficial across a variety of conditions, though effects do tend to be weaker in studies with longer follow-up times and larger sample sizes (see Appendix A, Table A1).

The results of the current study do not mean that HOPE/SCF sanctions have no effects or, for example, that graduated sanctions should not be integrated into community supervision. What they do suggest, however, is that the HOPE/SCF model currently has weak empirical support in terms of its capacity to substantially reduce recidivism. From a policy perspective, this fact should be disclosed to all funding sources and to every agency that is considering implementing an SCF

program. It is the case that the SCF programs do not have iatrogenic effects and, as will be noted, may be justified on grounds other than recidivism.

Still, the program should come with a "warning label" that it has not been shown to be effective or, in the least, to have only small effects (Cullen et al., 2014). In fact, it may be time to follow the practice in the pharmaceutical industry where drugs cannot be marketed until shown to reduce their targeted problem. In the same vein, correctional products should not be marketed unless experimental research has clearly demonstrated that they work to reduce recidivism and that there are known active ingredients. Otherwise, we end up, as we do today, with limited understanding of program effects, leaving the field with partial understanding of mechanisms that affect better outcomes (Taxman, 2018). This responsibility to disclose program effectiveness should not pertain exclusively to SCF interventions but should extend to all programs including rehabilitation modalities (Cullen, 2012; MacDonald, 2023). Recent meta-analyses, for example, have challenged the efficacy of psychological interventions including whether the RNR model outperforms treatment programs in general (Beaudry et al., 2021; Duan et al., 2023). These claims must be vetted but, if sustained, must be acknowledged by sponsors of these modalities.

The findings regarding the HOPE/SCF model should not be surprising. Research studies on deterrence-oriented programs, interventions imposing discipline, and over-reliance on sanctions have reported limited effects, if not criminogenic impacts (Cullen et al., 2002; Jonson & Manchak, 2018; Lipsey, 2009; Lipsey & Cullen, 2007; Petrich et al., 2021). Advocates of the HOPE/SCF model claim that it is based on a longstanding theory extending back to the classical school of Beccaria. But being founded in the 1700s is not evidence of a theory's merit (Pratt & Turanovic, 2018). In fact, rational-choice theory has been roundly criticized for ignoring developments in cognitive psychology and in behavioral economics (Kahneman, 2011; Thaler, 2015). More consequential, rational-choice theory ignores virtually all of criminology and the wealth of studies showing risk factors for recidivism (Bonta & Andrews, 2017), the virtues of incentivizing behavior (Mowen et al., 2018; Taxman et al., 2020), and the importance of using cognitive restructuring techniques (Labrecque et al., 2022). Because the HOPE/SCF approach focuses exclusively on the misconduct committed (violating a condition of supervision), any effects are likely to be short term and pay no attention to individual differences in criminogenic needs that underly the choice of crime. It also pays little attention to the empirical evidence surrounding interventions linked to dynamic risk factors that affect positive supervision outcomes (Taxman et al., 2020). Because it leaves these needs untreated, its capacity to change behavior is limited (Duriez et al., 2014).

An exception to this pattern is the effectiveness of "focused deterrence" interventions that seek to reduce crime by making the threat of punishment salient (for systematic reviews, see Braga et al., 2018; Braga & Kennedy, 2020). Focused deterrence involves identifying higher-risk individuals or groups, intervening with certainty and swiftness when a rule violation occurs (usually a violent act), and responding fairly by use of procedural justice (Braga & Kennedy, 2020). But this approach differs in three ways from SCF supervision: It is conducted in the community typically by prosecutors before an arrest occurs, it seeks to change the punishment perceptions not just of a person but of their associates, and enforcement is accompanied by communication to the individual of the consequences of not complying with the law as well as an offer of intervention services (Braga & Kennedy, 2020). Focused deterrence thus moves beyond individualized deterrence to a model that also includes enhancing informal social control and social support. These are perhaps ingredients that existing SCF programs might incorporate.

One option for HOPE/SCF advocates is to argue that even if this model does not achieve large reductions in recidivism, it still achieves other valuable goals that make its use worthwhile—especially since the results of the meta-analysis show positive, not iatrogenic, effects. This is a fair

rebuttal. One purpose might be to administer probation supervision that demands and instills probationer accountability. The SCF model ensures accountability by officers and judges to apply sanctions equitably and within a context of procedural justice. In a survey of participants in the Lattimore, MacKenzie, et al. (2016) DFE, Zajac and Dawes (2020) found that HOPE respondents were "universally aware" of what was required of them and of the sanction that would attach to noncompliance. Another compelling consideration is if the program reduces incarceration and proves cost effective—an issue on which the results are mixed (Campbell et al., 2021; Cowell et al., 2018; Hamilton et al., 2016;). Furthermore, although not advocating for a formal merger of treatment and deterrence goals, both Hawken (2016) and Kleiman (2016) favor agencies individualizing program ingredients to the local environment. As noted, with arguably good results, both Hawaii and the state of Washington introduced SCF principles to a system that trained officers in treatment skills. It is possible that SCF might be incorporated into probation agencies that retain a treatment mission as a means of increasing the dosage and consistency of supervision. If more research was undertaken demonstrating an array of positive effects beyond recidivism, then the HOPE/SCF model would have added legitimacy.

One limit to this line of argument, however, is the possible outcomes beneficial to supervisees that these programs do not seek to accomplish. Thus, by focusing on behavioral compliance to rules, HOPE/SCF programs make no effort to use probation as a means to address probationers' personal deficits, to build on their personal strengths, to incentivize them to change their behaviors (Mowen et al., 2018), or to change their personal identity to promote desistance (Bushway, 2021, 2022). There is no attempt, for example, to try to limit risk factors or to increase human capital.

Finally, the continued advocacy for and implementation of HOPE/SCF programs carries an opportunity cost. Other approaches to supervision exist, most notably a therapeutic approach based on the risk-needs-responsivity model (Bonta, 2023; Bonta & Andrews, 2017; Lowenkamp et al., 2014; Smith et al., 2012; Taxman, 2008; Taxman & Pattavina, 2013; Taxman et al., 2020). As another example, research by Skeem et al. (2007) shows that the quality of the relationship between supervisee and officer can reduce criminal involvement (see also Blasko, et al., 2015; Chamberlain et al., 2018; Okonofua et al., 2021; Taxman & Smith, 2020.) Researchers have found that individuals' perception of fair and just systems can reduce recidivism (Blasko & Taxman, 2018). There is also evidence that supervision focusing on reducing opportunities to offend can have positive effects (Schaefer et al., 2016; Schaefer and Little, 2020).

Other scholars reject the SCF logic more fully (Doherty, 2016). Thus, Lovins et al. (2018) propose that supervising officers should act more like coaches that guide and improve the skills of probationers than referees who judge and impose penalties for infractions. Klingele (2013) favors dramatically reducing the use of supervision conditions and imposing only those "that bear directly on an offender's criminal rehabilitation and risk of harm to the community" (p. 1061). Similarly, Corbett (2015, p. 1729) has proposed "zero-based condition setting" in which judges, officers, and probationers would start with a "blank slate" and then "work collaboratively" to determine conditions essential for "appropriate sanctioning and treatment" (p. 1729). Such conditions would be individualized, few in number, and never lead to revocation. A return to prison would be reserved for the commission of a new criminal offense. And Taxman et al. (2020) promote the use of incentives to identify those prosocial behaviors that will help individuals pursue a crime and/or drug-free lifestyle and to allow supervision officers to serve as facilitators of change (Sloas et al., 2019).

More importantly, recent reviews of contemporary supervision question its overall effectiveness and highlight that the "less is more" philosophy may be misguided given the current state

of empirical evidence (Doleac & LaForest, 2022; Lopoo et al., 2023). As shown here, it is clear that there is limited empirical evidence in support of SCF/HOPE, and for the field of supervision more generally (Mackey et al., 2022). A consistent observation in the current body of literature is that studies often fail to disentangle the active ingredients and mechanism of action that define effective supervision practices, and the tendency is to dismiss the value associated with effective strategies involving incentives, cognitive restructuring, and identity transformation. Promulgating that SCF sanctions for substance use behaviors has "beneficial effects" (Doleac & LaForest, 2022, p. 9) appears to be premature. Research needs to first address the question of which ingredients—sanctions or incentives, including the ratio of incentives to sanctions which Mowen et al. (2018) found in their study, and/or the treatments (cognitive behavioral therapy, contingency management, etc.) that are included —produce the more demonstrative impacts.

5 | CONCLUSIONS

The meta-analysis presented in this study does little to reverse the existing assessment that HOPE/SCF programs have minimal effects on recidivism. Our review further reveals that the core concepts of swift, certain and fair are not well articulated and highlights the importance of clearly operationalizing and measuring the effects of both treatment and control components in the next wave of community corrections programs. On a positive note, the Office of Justice Programs is now encouraging applicants that apply for SCF funding based on their jurisdiction's view of the appropriate target population and key design elements, which now can include treatment.[5] This change by funding providers can and should provide a unique opportunity for evaluators to assess both the treatment and control components of these programs. Unfortunately, the SCF evaluations we reviewed focused almost exclusively on measuring the deterrence components of these programs; as a result, only three of these studies, Hamilton et al. (2015), DeVall et al. (2015), and Grommon et al. (2012) provided details needed to begin to examine the comparative effects of treatment and control components. Even here, the research is quite limited: two of the studies included comparison of treatment referrals in experimental and control groups and one provided assessment of treatment availability more generally.

There is much more work to be done before we can offer full assessment of these programs. It is clear from our initial review of SCF/HOPE program evaluations that researchers have failed to measure a key—and underappreciated—component of both HOPE and SCF programs: treatment (modality, availability, quality, duration). We know from research on intensive supervision programs implemented in the eighties and early nineties that prospects for desistance among participants improved when they included the appropriate balance of treatment and control (see, e.g., Byrne, 1990; Byrne et al.,1992; Byrne & Pattavina, 1992; Byrne and Taxman, 2005; Petersilia & Turner, 1990; Taxman, et al. 2021). HOPE and SCF program developers are burying the lead by not recognizing and emphasizing this point and citing the existing evidence-based reviews that support this strategy (e.g., Drake, 2011). It is our suspicion that in the coming years, SCF evaluators will find that treatment—in all its myriad forms—is, once again, the *secret sauce* needed for these community supervision programs to be effective. But we will not know if this prediction is accurate unless we conduct the necessary evaluation research examining—and disentangling—the effects of various types of treatment (Tartafe et al., 2021).

To date, criminology does not typically embrace comparative effectiveness trials where two competing approaches are compared to identify which approach is better for what type of individual. Comparative effectiveness trials would focus on examining the factors that contribute to

positive (or negative) outcomes, and they would provide the needed coverage to ensure that promotion of SCF (or intensive supervision, electronic monitoring, cognitive restructuring programs, or any named approach) documents their critical ingredients have occurred and do in fact reduce recidivism.

Organized skepticism, a core norm of science (Merton, 1942/1973), should greet all claims to be the guiding model of supervision—including those just mentioned. In the end, an evidence-based approach to correctional practices should insist that science be the arbiter of how best to supervise those on probation and parole. Although other considerations are relevant (e.g., cost, equity), a key consideration—which this meta-analysis speaks to—is the impact of community supervision on reoffending. In this regard, the scientific evidence is not strong enough to support the use of SCF principles as the primary guide of probation supervision.

CONFLICT OF INTEREST

The authors confirm that they have no conflict of interest to declare.

ORCID

April Pattavina https://orcid.org/0000-0003-2632-388X
Joshua S. Long https://orcid.org/0000-0003-1476-6974
Francis T. Cullen https://orcid.org/0000-0003-4975-915X

ENDNOTES

[1] BJA's Resource Center actually links to an early assessment of the HOPE program by CrimeSolutions.gov as promising, but this rating was before the multisite replication was completed. The current rating is no effect but a visitor to the BJA funded site would not see the revised rating, only the original one. https://crimesolutions.ojp.gov/ratedprograms/675#:~:text=In%20January%202020%2C%20Honest%20Opportunity,only%20on%20measures%20of%20recidivism.

[2] We cross-checked our evaluations with a list maintained by researchers at the Marron Institute of Urban Management and found that we had not overlooked any that can reasonably be considered "HOPE" or "HOPE-inspired" programs (personal correspondence 7/30/2021).

[3] See Welsh & Farrington (2001) and Kennedy et al. (2019) for an overview and critical review.

[4] We have operationalized these three program components using specific review criteria that was not found in the original HOPE model. The original Hawken & Kleiman (2009) study does not offer advice on the number of days before a sanction is imposed or the number of days in jail. As long as the punishment is "very-brief… typically only a few days" they considered it to be "parsimonious." The original HOPE script said "If you test positive, you will be arrested on the spot, held in custody, and we will have a hearing 2 days later. If you used drugs, you will go to jail." For a full discussion of the difficulties related to replicating the key components of the original HOPE model, see the replication study by Lattimore and colleagues.

[5] https://bja.ojp.gov/program/swift-certain-and-fair-supervision-program/overview

REFERENCES

Alm, S. S. (2016). HOPE probation: Fair sanctions, evidence-based principles, and therapeutic alliances. *Criminology & Public Policy*, *15*(4), 1195–1214.

Bartels, L. (2016). Looking at Hawaii's Opportunity with Probation Enforcement (HOPE) program through a therapeutic jurisprudence lens. *QUT Law Review*, *16*(3), 30–49.

Beaudry, G., Yu, R., Perry, A. E., & Fazel, S. (2021). Effectiveness of psychological interventions in prison to reduce recidivism: A systematic review and meta-analysis of randomized controlled trials. *Lancet Psychiatry*, *8*(9), 759–773.

Blair, C. (2016). Judge Steven Alm is taking his message of "HOPE" to DC. *Civil Beat*. https://www.civilbeat.org/2016/09/hawaiis-steve-alm-is-taking-his-message-of-hope-to-dc/

Blasko, B., Friedmann, P. F., Rhodes, A., & Taxman, F. S. (2015). The parolee–parole officer relationship as a mediator of criminal justice outcomes. *Criminal Justice and Behavior*, *42*(7), 722–740.

Blasko, B., & Taxman, F. S. (2018). Are supervision practices procedurally fair?: Development and predictive utility of a procedural justice measure for use in community corrections settings. *Criminal Justice and Behavior*, *45*(3), 402–420.

Bonta, J. (2023). Training community supervision officers in the risk-need-responsivity model of offender rehabilitation: A review and implications. *Journal of Offender Rehabilitation*, *62*(1), 39–58.

Bonta, J., & Andrews, D. A. (2017). *The psychology of criminal conduct* (6th ed.). Routledge.

Braga, A. A., & Kennedy, D. M. (2020). *A framework for addressing violence and serious crime: Focused deterrence, legitimacy, and prevention (Elements in Criminology)*. Cambridge University Press.

Braga, A. A., Weisburd, D., & Turchan, B. (2018). Focused deterrence strategies and crime control: An updated systematic review and meta-analysis of the empirical evidence. *Criminology & Public Policy*, *17*(1), 205–250.

Brydges, C. R. (2019). Effect SizeGuidelines, Sample Size Calculations, and Statistical Power in Gerontology. *InnovAging*, *3*, PMCID: PMC6736231.

Bureau of Justice Administration. (2018). Swift, Certain, and Fair supervision programs. https://bja.ojp.gov/program/swift-certain-and-fair-supervision-program/overview

Bushway, S. (2021). *Reentry, desistance, and identity achievement*. American Enterprise Institute. https://www.aei.org/research-products/report/reentry-desistance-and-identity-achievement/

Bushway, S. D. (2022). Job-related programs for people on supervision: Reframing the problem. *Annals of the American Academy of Political and Social Science*, *701*(1), 98–113.

Byrne, J. (1990). The future of intensive probation supervision and the new intermediate sanctions. *Crime & Delinquency*, *36*(1), 6–41.

Byrne, J., Lurigio, A., & Petersilia, J. (Eds.). (1992). *Smart sentencing: The emergence of intermediate sanctions*. Sage.

Byrne, J., & Pattavina, A. (1992). The effectiveness issue: Assessing what works in the adult community corrections system. In J. Byrne, A. Lurigio, & J. Petersilia (Eds.), *Smart sentencing: The emergence of intermediate sanctions* (pp. 281–306). Sage.

Byrne, J., & Taxman, F. (2005). Crime (control) is a choice: Divergent perspectives on the role of treatment in the adult corrections system. *Criminology & Public Policy*, *4*(2), 291–310.

Campbell, C. M., van Wormer, J., & Hamilton, Z. K. (2021). A supervision policy with scope: Revisiting Washington States' swift-and-certain initiative. In P. K. Lattimore, B. M. Huebner, & F. S. Taxman (Eds.), *Handbook on moving corrections and sentencing forward: Building on the record* (pp. 124–142). Routledge.

Chamberlain, A. W., Gricius, M., Wallace, D. M., Borjas, D., & Ware, V. M. (2018). Parolee-parole officer rapport: Does it impact recidivism? *International Journal of Offender Therapy and Comparative Criminology*, *62*(11), 3581–3602.

Cheung, M. W. L. (2014). Modeling dependent effect sizes with three-level met-analyses: A structural equation modeling approach. *Psychological Methods*, *19*, 211–229.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). L. Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Corbett, R. P. Jr. (2015). The burdens of leniency: The changing face of probation. *Minnesota Law Review*, *99*(5), 1698–1733.

Cowell, A. J., Barnosky, A., Lattimore, P. K., Cartwright, J. K., & DeMichele, M. (2018). Economic evaluation of the HOPE demonstration field experiment. *Criminology & Public Policy*, *17*(4), 875–899.

Cullen, F. T. (2012). Taking rehabilitation seriously: Creativity, science, and the challenge of offender change. *Punishment & Society*, *14*(1), 94–114.

Cullen, F. T., & Gendreau, P. (2001). From nothing works to what works: Changing professional ideology in the 21st century. *Prison Journal*, *81*(3), 313–338.

Cullen, F. T., & Gilbert, K. E. (1982). *Reaffirming rehabilitation*. Anderson.

Cullen, F. T., Manchak, S. M., & Duriez, S. A. (2014). Before adopting Project HOPE, read the warning label: A rejoinder to Kleiman, Kilmer, and Fisher's comment. *Federal Probation*, *78*(2), 75–77.

Cullen, F. T., Pratt, T. C., & Turanovic (2016). It's hopeless: Beyond zero-tolerance supervision. *Criminology & Public Policy*, *15*(4), 1215–1227.

Cullen, F. T., Pratt, T. C., Micelli, S. L., & Moon, M. M. (2002). Dangerous liaison? Rational choice theory as the basis for correctional intervention. In A. R. Piquero & S. G. Tibbetts (Eds.), *Rational choice and criminal behavior: Recent research and future challenges* (pp. 279–296). Routledge.

Cullen, F. T., Pratt, T. C., Turanovic, J. J., & Butler, L. C. (2018). When bad news arrives: Project HOPE in a post-factual world. *Journal of Contemporary Criminal Justice*, *34*(1), 13–34.

*Devall, K., Lanier, C., & Hartman, D. (2015). Evaluation of Michigan's Swift and Sure Sanctions Probation Program. http://courts.mi.gov/Administration/admin/op/problem-solving-courts/Documents/SS-Eval.pdf

Doherty, F. (2016). Obey all laws and be good: Probation and the meaning of recidivism. *Georgetown Law Journal*, *104*(2), 291–354.

Doleac, J., & LaForest, M. (2022). Community supervision and public safety. Arnold Ventures. https://craftmediabucket.s3.amazonaws.com/uploads/AVCJIReport_CommunitySupervisionPublicSafety_DoleacLaForest_v2.pdf

Dragioti, E., Dimoliatis, I., Fountoulakis, K. N., & Evangelou, E. (2015). A systematic appraisal of allegiance effect in randomized controlled trials of psychotherapy. *Annals of General Psychiatry*, *14*(25), 1–9.

Drake, E. (2011). "What works" in community supervision: Interim report (Document No. 11-12–1201). Washington State Institute for Public Policy.

Duriez, S. A., Cullen, F. T., & Manchak, S. J. (2014). Is Project HOPE creating a false sense of hope? A case study in correctional popularity. *Federal Probation*, *78*(2), 57–70.

Duan, W., Wang, Z., Yang, C., & Ke, S. (2023). Are risk-need-responsivity principles golden? A meta-analysis of randomized controlled trials of community correction program. *Journal of Experimental Criminology*, Advance online publication. https://doi.org/10.1007/s11292-022-09550-w

*Edwards, D., & Pettigrew, G. (2016). Implementing swift & certain sanctions: Evidence from North Carolina's "Quick Dip" program. *Justice Center: The Council of State Governments.*

Fernandez-Castilla, B., Declercq, L., Jamshidi, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2020). Visual representation of meta-analyses of multiple outcomes:Extensions to forest plots, funnel plots, and caterpillar plots. *Methodology*, *16*(4), 12–22.

Fleming, R. (2021). The influential HOPE probation program is badly misnamed. *Filtermag.org*. https://filtermag.org/hope-probation-program/

Fox, J. (1991) *Regression diagnostics*. Sage.

*Frailing, K., Alfonso, B., Rapp, V., & Taylor, R. (2018). 24th JDC Swift and Certain probation two-year report (3/14/16-3/14/18).

Gaffan, E. A., Tsaousis, I., & Kemp-Wheeler, S. M. (1995). Researcher allegiance and meta-analysis: The case of cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*, *63*(6), 966–980.

*Grommon, E., Cox, S. M., Davidson, W. S., & Bynum, T. S. (2012). Alternative models of instant drug testing: Evidence from an experimental trial. *Journal of Experimental Criminology*, *9*(2), 145–168.

Guo, S., & Fraser, M. W. (2015). *Propensity score analysis: Statistical methods and applications* (2nd ed.). Sage.

Hamilton, Z., Campbell, C. M., van Wormer, J., Kigerl, A., & Posey, B. (2016). The impact of swift and certain sanctions: An evaluation of Washington State's policy for offenders on community supervision. *Criminology & Public Policy*, *15*(4), 1009–1072.

*Hamilton, Z., van Womer, J. G., Kigerl, A. C., Campbell, C. M., & Posey, B. (2015). Evaluation of Washington State Department of Corrections (WADOC) Swift and Certain (SAC) Policy Process, Outcome, and Cost-Benefit Analysis. Washington State University, Institute for Criminal Justice.

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2022). *Doing meta-analysis with R: A hands on guide*. CRC Press.

Hawken, A. (2010a). HOPE for probation: How Hawaii improved behavior with high-probability, low-severity sanctions. *Journal of Global Drug Policy and Practice*, *4*(3), 1–5. http://hopehawaii.net/assets/the-journal-of-global-drug-policy-practice.pdf

Hawken, A. (2010b). The message from Hawaii: HOPE for probation. *Perspectives*, *34*(3), 36–49.

Hawken, A. (2016). All implementation is local. *Criminology & Public Policy*, *15*(4), 1229–1239.

*Hawken, A., & Kleiman, M. (2009). *Managing drug involved probationers with swift and certain sanctions: Evaluating Hawaii's HOPE*. National Institute of Justice, U.S. Department of Justice.

*Hawken, A., Kulick, J., Smith, K., Mei, J., Zhang, Y., Jarman, S., & Vial, T. (2016). *HOPE II: A follow-up to Hawaii's HOPE evaluation*. National Institute of Justice, U.S. Department of Justice.

Hox, J., Moerbeek, M., & van de Schoot, R. (2018). *Multilevel analysis: Techniques and applications* (3rd ed.). Routledge.

Jonson, C. L., & Manchak, S. M. (2018). Corrections and deterrence. In D. S. Nagin, F. T. Cullen, & C. L. Jonson (Eds.), *Deterrence, choice, and crime: Contemporary perspectives* (Advances in Criminological Theory)(Vol, 23, pp. 309–337). Routledge.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kennedy, C., Fonner, V., Armstrong, K., Denison, J., Ping, T., O'Reilly, K., & Sweat, M. (2019). The Evidence project risk of bias tool: Assessing study rigor for both randomized and non-randomized intervention studies. *Systematic Reviews*, *8*(3), 1–10.

Kleiman, M. A. R. (2016). Swift-certain-fair: What do we know now, and where do we need to know? *Criminology & Public Policy*, *15*(4), 1185–1193.

Kleiman, M. A. R., Kilmer, B., & Fisher, D. T. (2014). Response to Stephanie A. Duriez, Francis T. Cullen, and Sarah M. Manchak: Theory and evidence on the swift-certain-fair approach to enforcing conditions of community supervision. *Federal Probation*, *78*(2), 71–73.

Klingele, C. (2013). Rethinking the use of community supervision. *Journal of Criminal Law and Criminology*, *103*(4), 1015–1069.

Labrecque, R. M, Viglione, J., & Caudy, M. (2022). The impact of community supervision officer training programs on officer and client outcomes: A systematic review and meta-analysis. *Justice Quarterly*, *48*, Advance online publication. https://doi.org/10.1080/07418825.2022.2120062

Lattimore, P. K. (2021). Thoughts from the multi-site evaluation of the Honest Opportunity Probation with Enforcement Demonstration Field Experiment. In P. K. Lattimore, B. M. Huebner, & F. S. Taxman (Eds.), *Handbook on moving corrections and sentencing forward: Building on the record* (pp. 99–123). Routledge.

*Lattimore, P. K., Dawes, D., MacKenzie, D. L., & Zajac, G. (2018). *Evaluation of the Honest Opportunity Probation with Enforcement Demonstration Field Experiment (HOPE DFE), Final Report*. RTI International.

Lattimore, P. K., Dawes, D., Tueller, S., MacKenzie, D. L., Zajac, G. X., & Arsenault, E. B. (2016). *Summary findings from the national evaluation of the Honest Opportunity Probation with Enforcement Demonstration Field Experiment: The HOPE DFE evaluation*. RTI International and the Penn State Justice Center for Research.

Lattimore, P. K., MacKenzie, D. L., Zajac, G. X., Dawes, D., Arsenault, E., & Tueller, S. (2016). Outcome findings from the HOPE Demonstration Field Experiment: Is swift, certain, and fair an effective supervision strategy? *Criminology & Public Policy*, *15*(4), 1103–1139.

Lipsey, M. W. (2009). The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims & Offenders*, *4*(2), 124–147.

Lipsey, M. W., & Cullen, F. T. (2007). The effectiveness of correctional rehabilitation: A review of systematic reviews. *Annual Review of Law and Social Sciences*, *3*, 297–320.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage.

*Loudenburg, R., Drube, G., & Leonardson, G. (2010). *South Dakota 24/7 Sobriety Program evaluation findings report*. Mountain Plains Evaluation LLC.

Lopoo, E., Schiraldi, V., & Ittner, T. (2023). How little supervision can we have? *Annual Review of Criminology*, *6*, 3.1–3.20.

Lovins, B. K., Cullen, F. T., Latessa, E. J., & Jonson, C. L. (2018). Probation office as a coach: Building a new professional identity. *Federal Probation*, *81*(1), 13–19.

Lowenkamp, C. T., Alexander, M., & Robinson, C. (2014). Using 20 minutes wisely: Community supervision officers as agents of change. In M. S. Crow & J. O. Smykla (Eds.), *Offender reentry: Rethinking criminology and criminal justice* (pp. 181–202). Jones and Bartlett Learning.

MacDonald, J. (2023). Criminal justice reform guided by evidence: Social control works—The Academy of Experimental Criminology 2022 Joan McCord Lecture. *Journal of Experimental Criminology*, Advance online publication. https://doi.org/10.1007/s11292-023-09558-w

MacKenzie, D. L. (2006). *What works in corrections: Reducing the criminal activities of offenders and delinquents*. Cambridge University Press.

Mackey, B., Appleton, C. J., Skidmore, S., Lee, J., & Taxman, F. S. (2022). At the intersection of research and practice: Constructing guidelines for a hybrid model of community supervision. *Aggression and Violent Behavior*, *63*, 101689. Advance online publication. https://doi.org/10.1016/j.avb.2021.101689

Menard, S. (2001). *Applied logistic regression analysis* (2nd. ed.). Sage.

Merton, R. K. (1973). The normative structure of science. In N. W. Storer (Ed.), *The sociology of science: Theoretical and empirical investigations* (pp. 267–278). University of Chicago Press. (Original work published 1942).

*Michigan Supreme Court. (2020). *FY 2021 swift and sure sanctions probation program annual report*. Michigan Supreme Court Office of Public Information.

*Midgette, G., & Kilmer, B. (2021). Can novel 'swift-certain-fair' programs work outside of pioneering jurisdictions? An analysis of 24/7 Sobriety in Montana, USA. *Addiction*, *116*(12), 3381–3387.

Mowen, T. J., Wodahl, E., Brent, J. J., & Garland, B. (2018). The role of sanctions and incentives in promoting successful reentry: Evidence from the SVORI data. *Criminal Justice and Behavior*, *45*(8), 1288–1307.

Nagin, D. S. (2016). Project HOPE: Does it work? *Criminology & Public Policy*, *15*(4), 1005–1007.

*North Carolina Department of Public Safety. (2014). Analysis Summary: Short-term jail confinement (Quick Dips) efficacy. *Rehabilitative Programs & Services Research & Design Support Analysis*.

*O'Connell, D., Brent, J. J., & Visher, C. A. (2016). Decide your time: A randomized trial of a drug testing and graduated sanctions program for probationers. *Criminology & Public Policy*, *15*, 1073–1102.

Okonofua, J. A., Saadatian, K., Ocampo, J., Ruiz, M., & Oxholm, P. D. (2021). A scalable empathetic supervision intervention to mitigate recidivism from probation and parole. *PNAS*, *118*(14), E2018036118.

*Olson, D. E., & Stemen, D. (2014). Process evaluation of the Cook County Adult Redeploy Illinois (ARI) Modified Project HOPE: Program. ICJIA.

Petersilia, J., & Turner, S. (1990). Intensive supervision for high-risk probationers: Findings from three California experiments. RAND.

Petersilia, J., & Turner, S. (1993). Intensive probation and parole. In M. Tonry (Ed.), *Crime and justice: A review of research* (Vol., 17, pp. 281–335). University of Chicago Press.

Petrich, D. M., Pratt, T. C., Jonson, C. L., & Cullen, F. T. (2021). Custodial sanctions and reoffending: A meta-analytic review. In M. Tonry (Ed.), *Crime and justice: A review of research* (Vol., 50, pp. 353–424). University of Chicago Press.

Pratt, T. C., & Turanovic, J. J. (2018). Celerity and Deterrence. In D. S. Nagin, F. T. Cullen, & C. L. Jonson (Eds.), *Deterrence, Choice, and Crime, Volume 23: Contemporary Perspectives* (pp.187–210). Routledge.

Pratt, T. C., Turanovic, J. J., Fox, K. A., & Wright, K. A. (2014). Self-control and victimization: A meta-analysis. *Criminology*, *52*(1), 87–116.

Pyrooz, D. C., Turanovic, J. J., Decker, D. H., & Wu, J. (2016). Taking stock of the relationship between gang membership and offending. *Criminal Justice and Behavior*, *43*(3), 365–397.

Raudenbush, S., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Sage.

Ritchie, S. (2020). *Science fictions: How fraud, bias, negligence, and hype undermine the search for truth*. Metropolitan Books.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41–55.

Schaefer, L., Cullen, F. T., & Eck, J. E. (2016). *Environmental corrections. A new paradigm for supervising offenders in the community*. Sage.

Schaefer, L., & Little, S. (2020). A quasi-experimental evaluation of the "environmental corrections" model of probation and parole. *Journal of Experimental Criminology*, *16*(4), 535–553.

*Schlueter, M., Wicklund, P., Adler, R., Owen, J., & Halvorsen, T. (2011). *Bennington County Integrated Domestic Violence Docket Project: Outcome evaluation*. Vermont Center for Justice Research.

*Shannon, L. (2013). *Kentucky SMART probation program: Year one report*. Kentucky Administrative Office of the Courts.

Sherman, L. W., Gottfredon, D. C., MacKenzie, D. L., Eck, J., Reuter, P., & Bushway, S. D. (1998). *Preventing crime: What works, what doesn't, what's promising*. U.S. Department of Justice: National Institute of Justice.

Skeem, J. L., Louden, E. L., Polaschek, D., & Camp, J. (2007). Assessing relationship quality in mandated community treatment: Blending care with control. *Psychological Assessment*, *19*(4), 397–410.

Sloas, L., Wooditch, A., Murphy, A., & Taxman, F. S. (2019). Assessing the use and impact of points and rewards across four federal probation districts: A contingency management approach. *Victims & Offenders*, *14*(7), 811–831. NIHMS1540092.

Smith, P., Schweitzer, M., Labrecque, R. M., & Latessa, E. J. (2012). Improving probation officers' supervision skills: An evaluation of the EPIC'S model. *Journal of Crime and Justice*, *35*(2), 189–199.

*Snell, C. (2007). Fort Bend County community supervision and corrections Special Sanctions Court Program: Evaluation report.

Tartafe, R. C., Mitchell, D., Mackey, B. J., Appleton, C. J., Walters, S. T., Lee, J., & Taxman, F. S. (2021). Surfing the three waves of CBT in community supervision. *Federal Probation*, *85*(3), 21–26.

Taxman, F. S. (2008). No illusion, offender and organizational change in Maryland's proactive community supervision model. *Criminology & Public Policy*, *7*(2), 275–302.

Taxman, F. S. (2018). The partially clothed emperor: evidence-based practices. *Journal of Contemporary Criminal Justice*, *34*(1), 97–114.

Taxman, F. S., & Pattavina, A. (2013). *Simulation strategies to reduce recidivism: Risk needs responsivity (RNR) modeling for the criminal justice system*. Springer.

Taxman, F. S., & Smith, L. (2020). Risk-Need-Responsivity (RNR) classification models: Still evolving. *59*, 101459. *Aggression and Violent Behavior*, Advanced online publication https://www.sciencedirect.com/science/article/abs/pii/S1359178920301634

Taxman, F. S., Smith, L., & Rudes, D. (2021). From *mean* to meaningful probation: The legacy of intensive supervision programs. In P. Lattimore, B. Huebner, & F. Taxman (Eds.). *Handbook on moving corrections and sentencing forward: Building on the record* (pp. 56–74). Routledge.

Thaler, R. (2015). *Misbehaving: The making of behavioral economics*. Norton.

Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics*, *40*(6), 604–634.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48.

*Virginia Criminal Sentencing Commission. (2016). *Immediate sanction probation pilot program evaluation*. Supreme Court of Virginia.

Welsh, B. C., & Farrington, D. P. (2001). Toward an evidence-based approach to preventing crime. *Annals of the American Academy of Political and Social Science*, *578*(1), 158–173.

Wilson, G. T., Wilfley, D. E., Agras, W. S., & Bryson, S. W. (2011). Allegiance bias and therapist effects: Results of a randomized controlled trial of binge eating disorder. *Clinical Psychology*, *18*(2), 119–125.

Zajac, G., & Dawes, D. (2020). HOPE is a good thing? Probationer perceptions of and experiences with the Honest Opportunity Probation with Enforcement Demonstration Field Experiment. *Journal of Crime and Justice*, *44*(1), 103–118.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Pattavina, A., Long, J. S., Petrich, D. M., M. Byrne, J., Cullen, F. T., & Taxman, F. S. (2024). Revisiting the effectiveness of HOPE/swift-certain-fair supervision programs: A meta-analytic review. *Criminology & Public Policy*, *23,* 45–76. https://doi.org/10.1111/1745-9133.12635

APPENDIX A
SUPPLEMENTAL ANALYSES

TABLE A1 Bivariate moderator results with sample restricted to non-RCT effect sizes **(*N* = 116).**

	Coefficient	*SE*	*t*-value	Intercept
Level 2 Moderators				
Log sample size	.029***	.008	3.488	−.328***
Recidivism outcome				
Drug				−.188***
Arrest/conviction	.083	.057	1.458	
Probation violation	.086	.057	1.508	
Incarceration	.032	.062	.510	
Length of follow-up				
12 months or less				−.143***
13 to 24 months	.026	.029	.893	
25 or more months	.080**	.028	2.829	
Bivariate or multivariate effect				
Bivariate				−.120***
Multivariate	.021	.035	.596	
Level 3 Moderators				
Offense exclusions				
No exclusions				−.130***
Violent offenses	−.012	.041	−.291	
Sex offenses	.057	.044	1.298	
Age				
Mixed				−.151***
Adult	.045	.045	1.003	
Number of effect sizes	−.000	.001	−.358	−.119***
Swift and Certain				
Yes				−.129***
No	.036	.039	.924	
Publication type				
Technical report				−.119**
Journal article	.070	.108	.647	
Gender	N/A in non RCT-only sample			
Special populations				
Yes				−.109**
No	−.010	.044	−.227	

(Continues)

TABLE A1 (Continued)

	Coefficient	*SE*	*t*-value	Intercept
Risk				
All levels				−.126***
Medium-to-high	.021	.044	.489	
Not reported	.008	.051	.160	

Abbreviations. RCT, randomized controlled trial; SE, standard error.
*** $p < .001$; ** $p < .01$; * $p < .05$.

TABLE A2 Bivariate moderator results with sample restricted to RCT effect sizes ($N = 290$).

	Coefficient	*SE*	*t*-value	Intercept
Level 2 Moderators				
Log sample size	−.013	.020	−.665	.100
Recidivism outcome				
Drug				.207***
Arrest/conviction	−.243***	.042	−5.809	
Probation violation	−.206***	.044	−4.660	
Incarceration	−.192***	.047	−4.124	
Length of follow-up				
12 months or less				-.030
13 to 24 months	.034	.048	.736	
25 or more months	.055	.032	1.685	
Bivariate or multivariate effect				
Bivariate				.019
Multivariate	-.028	.032	-.874	
Level 3 Moderators				
Offense exclusions				
No exclusions				−.008
Violent offenses	−.160*	.074	−2.170	
Sex offenses	.031	.036	.865	
Age	N/A in RCT-only sample			
Number of effect sizes	.004*	.002	2.379	−.072
Swift and Certain				
Yes				.018
No	−.031	.035	−.892	
Publication type				
Technical report				.021
Journal article	−.032	.027	−1.203	
Gender				
Males				−.008
Mixed	.024	.038	.638	
Special populations	N/A in RCT-only sample			

(Continues)

TABLE A2 (Continued)

	Coefficient	*SE*	*t*-value	Intercept
Risk				
All levels				.003
Medium-to-high	.018	.027	.665	
Not reported	N/A in RCT-only sample			

Abbreviations. RCT = randomized controlled trial; SE = standard error.
***$p < .001$; **$p < .01$; *$p < .05$.

AUTHOR BIOGRAPHIES

April Pattavina, Ph.D., is a professor in the School of Criminology and Justice Studies at the University of Massachusetts Lowell. Her research focuses on correctional reform, program evaluation, and the criminal legal system response to gender-based violence. Her work in the corrections area includes funding from the National Science Foundation to explore ways in which emerging technologies can be used in ways to support those on community supervision. She is currently working with the Massachusetts Executive Office of Public Safety and Security on an evaluation of a peer mentoring re-entry program.

Joshua Long, Ph.D., is an assistant professor in the School of Criminology and Justice Studies at the University of Massachusetts Lowell. His research interests include rehabilitation program evaluations, deterrence, risk assessment, and prison violence.

Damon M. Petrich, Ph.D., is an assistant professor in the Department of Criminal Justice and Criminology at Loyola University Chicago. His work focuses on the causes and consequences of exposure to community violence, desistance from crime, and the effectiveness of correctional policies. His recent research has appeared in Crime and Justice, Campbell Systematic Reviews, Corrections, and the Handbook of Issues in Criminal Justice Reform in the United States.

James M. Byrne, Ph.D., is Professor Emeritus, School of Criminology and Justice Studies, University of Massachusetts Lowell, and an affiliated faculty member at George Mason University's Center for Advancing Correctional Excellence. He is currently the project director for a NIDA and BJA-funded multi-site research project examining service provision in Veterans Treatment Courts. He also serves as a consultant for the Global Corrections Group, where he is conducting a review of the research on the effectiveness of all United States Federal Bureau of Prisons programming. Dr. Byrne is the Editor-in-Chief of the journal, Victims and Offenders: An International Journal of Evidence-based Research, Policy, and Practice, and the Co-Director of the Global Community Corrections Initiative, www.globcci.org.

Francis T. Cullen is a distinguished research professor emeritus and senior research associate in the School of Criminal Justice at the University of Cincinnati. He is a past president of the American Society of Criminology and the Academy of Criminal Justice Sciences. His current research interests include redemption as a correctional policy, advancing social support theory, the impact of racial beliefs on public punitiveness, and the criminology of Donald Trump.

Faye S. Taxman, Ph.D., is a University Professor at the Schar School of Policy and Government and Director of the Center for Advancing Correctional Excellence! at George Mason University. She is a health service criminologist. She has conducted experiments to examine different processes to improve treatment access and retention, to assess new models of probation supervision consistent with RNR frameworks, and to test new interventions. She is the current Principal Investigator for the National Institute on Drug Abuse's Justice Community Opioid Innovation Network (JCOIN). She received many awards including the Rita Warren and Ted Palmer Differential Intervention Treatment award, the Joan McCord Award from the Division of Experimental Criminology, and a Fellow of the American Society of Criminology.