

Assessing the Empirical Evidence of the National Adult Treatment Court Best Practice Standards (2nd Edition)

Criminal Justice Policy Review

1–24

© The Author(s) 2026



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/08874034261461473

journals.sagepub.com/home/cjp

Jamie C. Vaske¹ , Julie Marie Baldwin¹,
and John H. Boman, IV¹

Abstract

The Adult Treatment Court Best Practice Standards (the “Standards”) were intended to serve as “definitive guidance” for implementing evidence-based practices within the treatment court field. The tone and content of the second edition, published in 2023, differed significantly from those found in the first edition, especially for the section titled “Incentives, Sanctions, and Service Adjustments” (ISSA). Of the 72 claims evaluated from the ISSA section, only 8.33% were supported by methodologically rigorous research, and 20.83% had moderate empirical support. The remaining claims had either limited empirical support (30.56%) or none (40.28%). The authors found evidence of miscommunicated findings (or spin) in the ISSA section of the Standards, including the inappropriate extrapolation of primary study results. The current study’s findings suggest that treatment courts should be cautious when implementing new guidance from the ISSA section of the second edition of the Standards.

Keywords

treatment courts, spin, data interpretation, evidence-based practice

Introduction

The War on Drugs in the 1980s led to a sharp increase in substance-using offenders in the United States coming into contact with the criminal justice system. Between 1980

¹American University, Washington, DC, USA

Corresponding Author:

Jamie C. Vaske, Center for Research and Collaborative Partnerships, American University, 4400 Massachusetts Avenue NW, Washington, DC 20016, USA.

Email: jborin@american.edu

and 1995, the incarcerated population increased by 239% as drug offenses comprised a rising proportion of admissions. The surge of individuals entering carceral settings led to higher correctional expenditures, with the U.S. spending \$35 billion for state prisons and local jails in 1996 (Belenko et al., 1998). The straining of detention resources spurred criminal justice actors to explore ways to keep certain individuals in the community rather than incarcerating them. Drug treatment courts emerged as one of these innovations, with the first established in 1989 in Miami, Florida (Goldkamp, 1994). The treatment court model provides intensive supervision and mandated treatment to address the underlying causes of their criminal behavior, while promoting public safety and accountability outside the carceral setting. The model gained popularity, leading to the proliferation of various types of treatment courts, including community courts, mental health courts, veterans treatment courts, driving while intoxicated (DWI) courts, domestic violence courts, and human trafficking courts.¹

Despite the popularity and significant investment in treatment courts, early evaluations of the model revealed no significant differences in rearrest, reconviction, or reincarceration rates for individuals in treatment courts compared to those in the system as usual (O'Connell et al., 1999; Shaffer, 2006). Over time, studies began to show that *some* courts produced *some* positive outcomes for *some* defendants (Cissner et al., 2013; Rossman et al., 2011). Meta-analyses revealed modest (7%–14%) reductions in recidivism (Latimer et al., 2006; Lowder et al., 2018; Lowenkamp et al., 2005), with some programs being more effective than others (Mitchell et al., 2012; Shaffer, 2011; Stein et al., 2013).

Variations in treatment court functioning and effectiveness paved the way for the development of best-practice standards. In 2013 and 2015, the National Association of Drug Court Professionals convened a group of researchers, practitioners, and subject-matter experts to develop the Adult Drug Court Best Practice Standards (the “Standards”). These evidence-based standards were intended to guide the field in developing and implementing the treatment court model within a local context. While initially developed for drug courts, these standards are intended to guide all variations of the treatment court model. The Standards served as the model for best practices until 2023, when they were revised.

The new Standards seemed to deviate from the first edition in both the process of developing the guidance and the content. For instance, the opening sections of the first set of Standards identified the committee chairs, committee members, the National Association of Drug Court Professionals (NADCP) Board of Directors Research & Standards Committee, observers, and over 40 peer reviewers. In contrast, the initial Acknowledgments section of the second edition acknowledged a small core group of individuals, 35 peer reviewers, and approximately 15 contributors to the first edition. Thus, the process for developing the new Standards seemed to draw on a narrower range of expertise, which may have shaped the Standards’ content and foundational philosophy.

One section that underwent a significant revision in content was the “Incentives, Sanctions, and Service Adjustments” (ISSA) section (formerly known as Standard 4), which governs several day-to-day elements of the treatment court model, including program requirements, elements that should be included in the participant handbook

(e.g., “advanced notice”), and procedures for program termination. The new ISSA section included partial guidance from the first edition of the Standards and introduced several new concepts and approaches for implementing the treatment court model (e.g., clinical stability). An initial review of the ISSA section suggested that some guidelines lacked empirical support, had weak empirical support, or contradicted peer-reviewed research. Furthermore, the guidance in the second edition contradicted long-standing guidance from the first edition. For instance, the first edition emphasized the need for standardized phase requirements across participants (Shaffer, 2006), whereas training based on the second edition promotes flexible or different phase requirements across different participants. While the second edition contradicted some of the prior guidance and the research supporting that guidance, it did not provide research citations: (1) to support the new guidance found only in the second edition, and (2) to show why the guidance in the first set of Standards should be discarded (e.g., the former guidance did not produce positive outcomes). The new guidance led to confusion not only among treatment court researchers but also among practitioners and program participants.

Missteps in communicating translational research can be troubling, given the growing skepticism toward research among both academics and laypeople. Recent research has investigated the prevalence of miscommunication or “spin” in systematic reviews, meta-analyses, and randomized controlled trials in psychiatric clinical literature (Perlis, 2025), biomedical research (Chiu et al., 2017), and evaluations of health care interventions (Boutron et al., 2010). Spin has been described as “specific reporting that could distort the interpretation of results and mislead readers” (Boutron et al., 2010, p. 2058) and can lead research consumers, especially laypeople, to overestimate an intervention’s efficacy and safety (Boutron et al., 2019). This can have profound effects in the criminal justice system, where the consequences of misguided policies can be far-reaching and life-changing. Within treatment courts, the presence of spin in the national standards can lead programs to adopt policies that may be harmful and undermine the effectiveness of the program, leading to high substance use rates, rearrest, probation revocation, and reincarceration. This can have implications not only for the program but also for treatment court participants (who may experience collateral consequences of the programmatic choices), their communities, and the model as a whole.

This study aims to investigate the strength of the evidence cited in the second edition of the Standards. In particular, it evaluates the literature base cited in the ISSA section, which addresses incentives, sanctions, and service adjustments. It evaluates the literature base by reviewing the studies that are cited in reference to the claim and determining: (1) whether the cited studies directly test the claim, and (2) the methodological and statistical rigor of the studies. The study focuses solely on the ISSA section’s literature due to the scope of changes in the section and its relevance to day-to-day program operations. In addition, because the second edition integrates newly introduced guidance alongside content preserved from the first edition, the study evaluates whether the evidentiary strength varies between claims introduced in the second edition and those retained from the first.

Literature Review

The treatment court model integrates behavioral health services and criminal justice processing to provide intensive supervision and mandated treatment to individuals at risk of reoffending. Although eligibility criteria vary not only across different types of courts (e.g., drug, mental health, and veteran courts) but also within programs of the same type, the Standards advocate that courts direct services to high-risk, high-need individuals, given the model's intensity. Eligible individuals should be promptly placed into the court and connected to appropriate services to help them stabilize and effectively engage with both the criminal justice and behavioral health systems. Once admitted, participants receive wraparound services from a multidisciplinary team, including supports for substance use, mental health, housing, transportation, family needs, employment, and physical health. Participants are also subjected to intensive supervision, which may include random drug testing at least twice weekly, meeting with probation or supervision officers at least once a week, and attending status hearings in court at least every 2 weeks. Courts should employ a combination of incentives, sanctions, and therapeutic interventions to shape behavior and promote desistance (National Association of Drug Court Professionals, Drug Court Standards Committee, & United States. Drug Courts Program Office, 1997).

The earliest evaluations of treatment courts suggested these intensive interventions had little to no effect on recidivism (Belenko et al., 1994). Deschenes and Greenwood (1994) found that individuals randomly assigned to drug court (16.95%) were equally as likely to be rearrested as those assigned to traditional probation (15.37%). Granfield and colleagues' (1998) analysis of Denver's drug treatment court also found no significant difference in the prevalence of revocations and rearrests among those in drug court and two historical comparison groups. Similar results were found in early evaluations of other types of adult treatment courts, such as mental health courts (Christy et al., 2005), DWI courts (Cissner, 2009), and veterans treatment courts (Hartley & Baldwin, 2019).

As evaluations of these treatment court models continued, some studies began to show significant differences in recidivism rates between treatment court participants and comparison groups. The results, however, were often mixed, with some outcomes showing a positive impact of the intervention, while others showed no impact or had an iatrogenic effect (Anestis & Carbonell, 2014; Wright & Clymer, 2000). For instance, Rossman et al.'s (2011) multi-site evaluation of 23 drug courts found that drug court participants had greater reductions in criminal behavior and substance use than individuals not in treatment courts. However, moderation analyses revealed that drug treatment programs were more effective for some participants (i.e., those with prior violent convictions) and less effective for other participants (i.e., African Americans and those with narcissistic personality traits). Studies also showed that specific program characteristics enhanced program effectiveness (i.e., the judge demonstrates procedural justice), while other practices reduced program impact (Carey et al., 2008; Rossman et al., 2011). These results, along with those from meta-analyses, began to suggest that the impact of treatment courts was nuanced and context-dependent (Lowenkamp et al., 2005; Sarteschi et al., 2011; Stein et al., 2013).

Researchers began developing theoretical frameworks to describe the differences in treatment court operations and outcomes (Goldkamp, 1999; Senjo & Leip, 2001; Taxman & Bouffard, 2002). Longshore et al. (2001) proposed that the effectiveness of drug courts may vary based on program operations and the population served. That is, programs could be categorized based upon population severity (e.g., substance use severity), leverage (e.g., legal incentives and consequences), program intensity, predictability of program expectations and consequences, and rehabilitation focus. They argued that programs would be more effective when they had greater legal leverage over participants, served individuals with fewer needs, were more intensive, and adopted a more therapeutic approach rather than a legalistic one. This framework, along with others, laid the foundation for developing best practices in treatment courts.

While researchers were creating frameworks to measure variation in treatment court operations and effectiveness, a growing movement aimed to standardize and professionalize treatment courts (Belenko, 2002). The NADCP was established in 1994 as the “premier national membership, training, and advocacy organization for the Drug Court model” (NADCP, 2015, p. vi). NADCP became the leading professional organization to deliver training and technical assistance to the treatment court field. One of the key tasks of NADCP was to develop a set of best practice standards to promote core elements of the model and the use of evidence-based practices in treatment courts.

Adult Treatment Court Best Practice Standards

Treatment courts follow the Adult Treatment Court Best Practice Standards (“Standards”), which are consistently promoted by All Rise (NADCP rebranded to All Rise in 2023). Published by NADCP in 2013 and finalized in 2015, the first edition contained 10 standards that outline the “best practices” for treatment courts, covering important topics such as the target population, incentives and sanctions, treatment, drug testing, and evaluation. Courts are strongly encouraged to center their operations around these benchmarks. Indeed, federal and state funding opportunities often require applicants to specify the specific standards they adhere to and which standards they aim to align their programs with through funding. Training and technical assistance providers, with federal funding support, host workshops and conferences for thousands of treatment court professionals each year to help them gain compliance with the Standards.

The Standards remained essentially unchanged until 2023. In late 2023, All Rise released updated versions of select sections, with the remaining updates issued in 2024.² As noted previously, the second edition contains both new material and retains some guidance from the first edition. The most extensive set of changes appeared to be in the ISSA section, which discusses participant expectations, phase structure, incentives, and sanctions. The tone of the Standards moved from one grounded in a criminal justice philosophy (for the first edition) to one espousing a more social work paradigm (in the second edition). This shift in philosophy was reflected not only in the Standards but also in the accompanying training and technical assistance from All Rise that federal grantees are required to complete. For instance, practitioners were directed to shift from applying uniform phase criteria governing program progression to adopting

individualized requirements for participant progression and program completion. The new ISSA standard also included “fuzzy” requirements, such as participants demonstrating clinical stability (a term not found in the *Diagnostic and Statistical Manual of Mental Disorders* [5th ed.; *DSM-5*; American Psychiatric Association, 2013]) and developing a therapeutic alliance with at least one team member. Finally, the second edition allows for continued or intermittent substance use 210 to 270 days into the program, while the first edition required abstinence even during the first 60 to 90 days of the program.

While criminal justice is becoming a more integrated, multidisciplinary system, the shift to a social work, client-centered system appears to be a difficult transition for several jurisdictions and stakeholders. Treatment court practitioners have publicly expressed difficulty adopting a social work mentality within the criminal justice system. This friction between the Standards and local criminal justice actors motivated the authors to explore empirical evidence supporting the changes in the Standards, particularly within the standard related to ISSA.

Current Study

The second edition of the Standards states that its definition for best practices remained the same. Specifically, the second edition states the following:

As before, *best practices* were defined as services or interventions that have been proven through at least two high-quality experimental or quasi-experimental studies, meta-analyses, or quantitative systematic reviews to improve outcomes in treatment courts, other correctional rehabilitation programs, and/or substance use, mental health, or trauma treatment. (All Rise, 2023, p. 2)

With this foundation, the current study seeks to assess:

1. What is the strength of empirical evidence supporting the claims and guidance provided in the second edition’s ISSA standard?
2. Does the strength of the empirical evidence cited in the ISSA standard differ between claims newly introduced in the second edition and those retained from the first edition?

Methods

The claims presented in the ISSA section of the second edition of the Adult Treatment Court Best Practice Standards (“Standards”) comprise the target population for this study. The ISSA section was chosen for analysis for two reasons. First, the section underwent significant changes between the first and second editions. Second, the ISSA section addresses multiple key elements of a treatment court, including program requirements, sanctions and incentives, and graduation and termination.

Claim Identification

Claims were included for analysis if they (1) contained a citation for readings other than court opinions or case law, *and* (2) were statements about either: (a) practices or outcomes in the treatment court setting or an adjacent criminal justice setting (e.g., pretrial, probation), or (b) individuals' decision-making or psychological processes that were explicitly hypothesized to be related to treatment court practices or outcomes. Claims were excluded if they included statements about (1) the predictors of maladaptive behavior (e.g., severity of substance use is impacted by one's genetics), (2) individuals' decision-making or psychological processes without a connection to criminal justice practices or outcomes, and (3) court decisions or case law.

Assessment of the Empirical Evidentiary Strength. The strength of the evidence for each claim was evaluated by reviewing the general rigor of the empirical studies cited for the claim. The current study evaluates the strength of both qualitative and quantitative empirical studies that include a sample of observations (e.g., individuals, studies) and applies an analytical method to those observations to develop findings. Readings were not evaluated for empirical rigor if they were book chapters, law reviews, technical reports, or theoretical papers that did not directly analyze a series of observations.³

The strength of the evidence was evaluated using the decision tree approach illustrated in Figure 1. The authors originally intended to use a well-established method for evaluating the strength of the literature such as the Maryland Scientific Methods Scale (Farrington et al., 2003), the Quality Assessment Tool for Quantitative Studies (Thomas et al., 2004), or the Cochrane Handbook for Systematic Reviews (Higgins et al., 2024). However, an initial application of this approach revealed that these methods were not suitable for the current review, as some claims included citations that did not directly test the claim. For instance, there were six studies cited in reference to the following claim: "For example, studies find that many drug courts and probation programs deliver higher-magnitude sanctions for positive drug tests than for missing counseling sessions." Only one of the six cited studies tested whether individuals who tested positive for drugs or alcohol were more likely to receive harsher sanctions (i.e., jail, probation revocation) than individuals who missed treatment (Zettler & Martin, 2020). The other five studies tested the prevalence of sanctions in general (without tying them to a specific act of noncompliance), the association between substance use and jail sanctions (without including contrasts for treatment absences), or the association between receiving a jail sanction and treatment court failure. In light of these considerations, the first author reviewed various evidence-based scoring systems that are used to evaluate medical and social science research (Baker et al., 2010; West et al., 2002). At the highest level of inquiry, these systems assess the strength of the study's design and analytical methods, aligning with the second edition's benchmark for what constitutes best practice.⁴ These key features are built into the latter part of the decision tree.

As shown in Figure 1, the first step in the tree is to evaluate whether the cited work(s) directly tested the claim. If none of the readings directly evaluated the claim presented, the authors recorded the strength of the cited empirical research as "None."

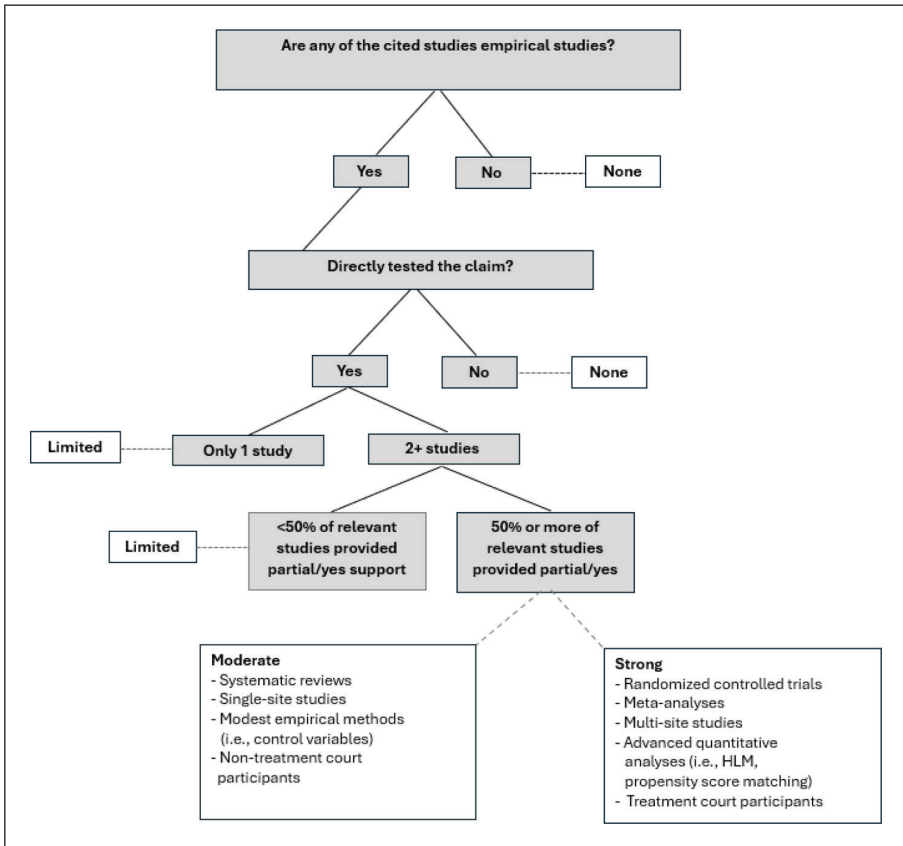


Figure 1. Decision tree for evaluating the strength of empirical evidence for ISSA claims.

The strength of the empirical research was also recorded as “None” if all citations for a claim are not empirical studies but were theoretical papers, books, technical reports, law reviews, literature reviews, or other readings that did not have a clear methodology or analytical plan. If the cited article(s) were empirical articles that explicitly evaluated the claim, the second layer of evaluation focused on the number of studies cited in reference to the claim. If the empirical article(s) directly tested the claim but none of the articles found evidence for the claim, the strength of support was rated as “None.” If there was only one study cited and the results supported the claim, then the strength of the evidence was recorded as “Limited.” If there were two or more studies that directly tested the claim, then the first author evaluated the strength of the research based on its research methodology and statistical methods. The strength of the evidence was rated as “Moderate” if most of the cited studies were systematic reviews, single-site treatment court studies, used basic statistical methods to address confounding or unobserved variables, or included non-treatment court samples (e.g., inmates,

Table 1. Summary of the Rating System for Empirical Support.

Rating	Definition	Evidence criteria
None	No empirical studies support the claim	No evidence identified
Limited	Only one study provides empirical support for the claim	Single supporting study of any methodological quality
Moderate	At least 50% of studies provide partial or full support for the claim, but rely on moderately rigorous research designs	Systematic reviews, single-site treatment court studies, studies of non-treatment court participants, or modest empirical methods ^a
Strong	At least 50% of studies provide partial or full support for the claim and employ rigorous research designs	Randomized controlled trials, meta-analyses, multi-site studies, studies sampling treatment court participants, or advanced quantitative methods ^b

Note. Rating categories reflect both the proportion of supporting studies and the methodological rigor of the evidence base.

^aSuch as multivariate rather than bivariate analyses. ^bSuch as difference-in-differences, propensity score matching, hierarchical linear modeling.

probationers, inpatient treatment participants) with moderate to strong quantitative empirical methods. Evidence was rated as “Strong” if the results supported the claim, predominantly cited studies of treatment court participants, and were founded on meta-analyses, randomized controlled trials, multi-site studies, and advanced quantitative analyses (e.g., differences-in-differences, propensity score matching, instrumental variables). Table 1 below summarizes the rating system of empirical support.

Please note, in the latter part of the analysis, the authors combine moderate and strong, given the small number of claims rated as “Strong” ($n = 6$) in the original classification.

Use of “e.g.” In addition to recording the strength of the evidence for each claim, the dataset included indicators of whether the claim was flagged with an “e.g.” The use of “e.g.” indicates that the study is used as one example of a study that finds what a claim purports; that is, there may be several studies that find a specific result, and the author is merely using “e.g.” to identify one or two studies. However, authors may inconsistently use “e.g.” and use it merely to refer to an idea or a concept within a statement. For instance, the ISSA section included the following statement that is marked with an “e.g.”: “Providing advance notice of behavioral expectations and responses also enhances participants’ perceptions of procedural fairness in the program, which produces significantly better and more rapid improvement” (All Rise, 2023, p. 78). One of the two empirical studies cited in support of this claim (Stutts & Cohen, 2022) used data from the Pathways to Desistance study to examine how changes in youths’ perceived procedural justice were associated with changes in depression trajectories. The study did not explicitly test whether providing advanced written notice of expectations

led to higher levels of procedural justice and better outcomes, nor did it focus on adults. Thus, the author may have used “e.g.” merely to reference the concept of procedural fairness.

ISSA Claim Found in First Edition. The study also included an indicator of whether an ISSA claim was also found in the first edition of the Standards. ISSA claims may be partially included in the first edition if the exact statement was not found, but the same ideas were conveyed in the first edition. For instance, the first edition does not include the following ISSA claim:

Studies find that behavior improves most rapidly and efficiently when (1) participants receive clear advance notice of what behaviors are expected of them or prohibited, (2) participants are informed of the range of responses that will result from meeting or not meeting these expectations, and (3) responses are delivered as described. (All Rise, 2023, p. 78)

However, the key ideas are found in the first edition’s section on Advance Notice (Standard 4.A) which begins with the sentence: “Numerous studies reported significantly better outcomes when Drug Courts developed a coordinated sanctioning strategy that was communicated in advance to team members and participants” (NADCP, 2015, p. 29). At other times, the exact statement can be found in both the first and second editions. For analytical purposes, claims are coded as being solely found in the second edition of the Standards if the same statement or idea is not reflected in the first edition (0 = Claim is in both the first and second edition, 1 = Claim is absent from the first edition and only found in the second edition). This variable is included in the analyses as the tone of the first edition was more nuanced and highlighted the weaknesses or inconsistencies in the literature base. In contrast, the second edition is more declarative in tone.

Analytical Sample, Evidence, and Interrater Reliability. After excluding claims that did not meet the inclusion criteria, the final analytical sample consisted of 72 claims, which cited 174 unique readings. The raw data file for each claim’s ratings, the summary of each article, and the section summaries can be found at: <https://github.com/AU-CRCP/ISSA.git>. Forty-seven (27.01%) of the 174 readings were deemed non-empirical studies as they were literature reviews, theoretical papers, textbooks, or book chapters that did not include a description of the research methodology or analytical plan. The remaining 127 readings were empirical articles that described how observations were selected for the respective studies. The 127 articles were used to assess 72 ISSA claims. The first author evaluated all 72 claims, and the third author independently evaluated 9 (of the 72) claims (across 33 articles) to assess interrater reliability in the latter portion of the ISSA section. The latter portion of the ISSA section was chosen because the third author is well-versed in that literature base (i.e., risk, need, and responsivity) and employed his expertise to independently evaluate the claims articulated in those sections. There was 78% concordance between the ratings of the first and third authors for the nine claims, indicating substantial agreement (Altman, 1999; Landis & Koch, 1977). Cohen’s Kappa was .64, indicating moderate agreement between the two raters (McHugh, 2012).

Results

All 72 unique claims were evaluated from the ISSA section of the Standards. Of the 72 claims, 47.22% ($n = 34$) were flagged with an “e.g.” citation style, and 59.72% ($n = 43$) were not found in the first edition of the Standards. The data showed that 8.33% were rated as having strong empirical evidence, and 20.83% had moderate evidence. Notably, 30.56% had limited empirical evidence, and 40.28% ($n = 29$) had no empirical evidence to support the claim. Of the 29 claims that were rated as having no empirical support, 55.17% ($n = 16$) had studies that did not directly test the claim put forth, while 44.83% ($n = 13$) were not empirical articles (e.g., were law reviews, theoretical papers, or literature reviews with no selection criteria).

A sizable percentage of claims (55.17% of those without empirical support) cited research that did not directly test the specific hypothesis. For example, the “Responding to Proximal Goal Infractions” section stated: “Outcomes are consistently better when staff express their belief, convincingly, that participants can get better, and that responses are being imposed to help them reach their rehabilitative goals” (All Rise, 2023, p. 91). One study was cited in reference to this claim (Connor, 2019). This qualitative study focused on participants’ perceptions of procedural justice in community courts but did not empirically evaluate whether procedural justice practices improved outcomes in the community courts.

As another example, the “Jail Sanctions” subsection within the ISSA section included the following claim:

Although specific types of in-custody programs such as therapeutic communities (TCs) have been shown to improve outcomes, the benefits from these programs were attributable to the fact that they increased the likelihood that persons would enter and complete treatment after release from custody. (All Rise, 2023, p. 96)

Focusing on the latter part of the claim, one would assume that the cited articles test whether therapeutic communities decrease recidivism by increasing the likelihood of entering aftercare treatment; thus, aftercare would be considered the mechanism or mediator by which therapeutic communities reduce recidivism. However, the three cited empirical articles showed that aftercare treatment *moderated* the effects of therapeutic communities on recidivism (Bahr et al., 2012; Martin et al., 1999; Wexler et al., 1999), with therapeutic communities’ effectiveness further enhanced by aftercare. *None* of the articles tested whether aftercare engagement mediated the association between therapeutic communities and recidivism. This important nuance helps stakeholders understand whether therapeutic communities alone reduce recidivism because they encourage continued treatment engagement (i.e., mechanism) or whether one can enhance therapeutic communities’ effectiveness by supplementing them with aftercare treatment (i.e., moderator). If aftercare is a mechanism (rather than a moderator), this could suggest that entities should invest in aftercare because it directly impacts recidivism (rather than the therapeutic community). Instead, if aftercare is a moderator, it could suggest that therapeutic communities alone may not significantly impact recidivism and that it is only the combination of aftercare and therapeutic communities that

Table 2. Association Between Strength of Evidence and Claim Features.

Strength of Evidence	No “e.g.”	Used “e.g.”	χ^2
No empirical support	36.84% (14)	44.12% (15)	10.802**
Limited support	18.42% (7)	44.12% (15)	
Moderate or strong support	44.74% (17)	11.76% (4)	
	In both editions	Second edition only	χ^2
No empirical support	13.79% (4)	58.14% (25)	17.172***
Limited support	34.48% (10)	27.91% (12)	
Moderate or strong support	51.72% (15)	13.95% (6)	

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

reduces recidivism. While a technical statistical point, it does have practical implications for funding and programming.

It was also expected that the strength of the evidence may vary based on whether the claim was flagged with an “e.g.” citation style. Indeed, Table 2 shows that only 11.76% of claims with this citation style had moderate or strong support compared to 44.74% that did not use this citation style ($\chi^2 = 10.802, p = .005$). An examination of these results suggests that claims flagged with “e.g.” would include articles that generally discussed the topic within the claim but did not directly test the claim itself (see the example of Stutts and Cohen’s (2022) test of procedural justice within this article’s “Methods” section).

Finally, it was hypothesized that claims found solely in the second edition would have less empirical support than claims found in both the first and second editions. Table 2 shows that 58.14% of claims found solely in the second edition were *not* supported by empirical evidence compared to 13.79% of claims that were found in both the first and second editions. In contrast, 51.72% of claims found in both editions had moderate or strong support, compared to only 13.95% of claims found solely in the second edition ($\chi^2 = 17.122, p < .001$).

Discussion

The current study assessed the evidence cited supporting the ISSA guidance in the second edition of the Adult Treatment Court Best Practice Standards, which serves as the national guidelines for the thousands of treatment courts operating across the United States. These standards serve as the practice guidelines for treatment courts and

were intended to be based on rigorous research designs, such as randomized controlled trials and meta-analyses. This evaluation revealed that the majority (70.84%) of the 72 ISSA claims had *little to no empirical support*. Specifically, 40.28% were *not* empirically supported by the cited evidence, and 30.56% had only *limited* empirical support. The two most common issues with claims lacking empirical support were that the studies did not directly test the claims made in the ISSA standard (55.17%) or that the readings were not empirical studies (44.83%), such as law reviews, literature reviews, textbooks, or theoretical papers. While these types of sources contribute to the overall literature base and discourse in the field, they can selectively choose research—whether consciously or unconsciously—that supports the authors’ perspectives and does not constitute rigorous empirical research.

In addition, over half of the claims lacking empirical support cited literature testing hypotheses that differed from those explicitly stated in the ISSA Standard. Instead, it appears that the author(s) of the ISSA Standard inappropriately extrapolated from the available research base. Yavchitz and colleagues (2016, p. 59) define inappropriate extrapolation as “inappropriate generalization of study results” where researchers may overgeneralize the effects: (1) of an intervention on one specific outcome to a global disorder (e.g., cravings to substance use disorder diagnosis), (2) of a specific practice to a larger intervention (e.g., use of sanction grids to deterrence oriented programs), (3) from one intervention to a different intervention (e.g., delayed discounting studies to evidence on contingency management or the use of incentives), or (4) from one population or setting to a different population or setting. For instance, the ISSA Standard included the claim “Even brief intervals of detention-induced abstinence without MAT can cause a substantial decline in opioid tolerance, which increases a person’s overdose risk 10-to-40 fold if the person resumes opioid use upon release” (p. 96). To support this claim, All Rise cited two references (Binswanger et al., 2013; Ranapurwala et al., 2018). While both studies examined the mortality and overdose rates among inmates released from state-level Department of Corrections facilities, neither study discussed the inaccessibility of medication-assisted therapy within the facilities, and further, detention within a prison setting may not be considered brief relative to jail incarcerations.

These results reveal that the ISSA Standard was not created following the articulated approach for development of the Standards:

As before [in the first edition], best practices [in the second edition] were defined as services or interventions that have been proven through at least two high-quality experimental or quasi-experimental studies, meta-analyses, or quantitative systematic reviews to improve outcomes in treatment courts, other correctional rehabilitation programs, and/or substance use, mental health, or trauma treatment. (All Rise, 2023, p. 2)

This discovery is concerning and raises questions about how the guidelines were developed.

There were other places in the ISSA Standard where a strong research base supported the practice, but the evidence was poorly cited in the text. In several cases, the Standard cites only a limited number of studies, or even just one, to support statements

that have a broader, more robust research base. For instance, the Advance Notice section contains the claim “Too much flexibility is associated with ineffective outcomes because staff may not deliver responses predictably or as described” (All Rise, 2023, p. 80). Only one study (Cissner et al., 2013) was cited in support of this claim, and it focused on team members’ use of a formal response system to determine sanctions, with little to no discussion of flexibility in the schedule. In contrast, Rossman et al.’s (2011) multi-site evaluation of drug treatment courts found that courts with low predictability in their use of sanctions (i.e., too much flexibility) prevented fewer crimes per month than courts with medium predictability. In addition, the ISSA Standard states that supervision officers should model appropriate behavior, provide positive reinforcement, and maintain professionalism in their interactions with those under supervision. While one study showed that officers use of case planning and evidence-based practices was related to reductions in recidivism (Abt Associates, 2018), there are several other studies which highlight the importance of developing a working alliance with probationers (Sturm et al., 2021) and employing social learning modeling in supervision, such as Effective Practices in Community Supervision (EPICS), Staff Training Aimed at Reducing Rearrest (STARR), and Strategic Training Initiative in Community Supervision (STICS; Bonta, 2023; Labrecque et al., 2023; Smith et al., 2012).

The key concern here is the inconsistent and incomplete use of evidence to justify specific claims within the ISSA Standard. This selective or insufficient citation practice can inadvertently weaken the credibility of otherwise well-supported practices. As just shown, the claim about flexibility and ineffective outcomes relies on a single study that does not fully address the issue, despite the availability of stronger and directly relevant evidence from a rigorous multi-site evaluation (i.e., Rossman et al., 2011). Similarly, in areas related to officer behavior and the therapeutic alliance, the standard under-references well-established literature on working alliance and skills-based supervision models (e.g., EPICS, STARR, STICS). When the cited evidence does not fully reflect the depth of the empirical literature, it risks overstating or understating the strength of support for particular practices, which may lead practitioners to question the rigor, transparency, and reliability of the recommendations.

While a substantial proportion of claims in the ISSA Standard lacked empirical support, it is important to note that 29.17% received moderate to strong empirical support (20.83% moderate, 8.33% strong). Table 3 identifies the specific claims that had moderate or strong support, as well as any additional information to help readers understand the nuance of findings supporting that claim.

In addition, 30.56% of claims had limited support, and several of those claims were categorized as “limited” because only one study provided support for the claim. Stakeholders may choose to adhere to claims that have moderate or strong support and cautiously implement practices with limited support. Careful monitoring of outcomes can inform practitioners if their implementation of evidence-based practices is working as intended in their jurisdiction.

Aside from confirming the evidence base for specific guidance, the current study also highlights gaps in the treatment court literature. The practices flagged as lacking empirical support offer an opportunity for researchers to examine whether they improve

Table 3. Claims From the Second Edition of the Adult Treatment Court Best Practice Standards With Moderate or Strong Support.

Claim	Additional explanation
Contingency management can be especially effective for high-risk and high-need persons who may lack intrinsic motivation for change when they first enter treatment court or whose motivation may fluctuate when they confront stressors in their social environment such as family discord or interpersonal conflict.	The effectiveness of contingency management seems to depend upon other conditions being present (e.g., diagnosed with antisocial personality disorder, not currently being involved with the criminal justice system, having 100+ self-reported prior criminal acts).
Numerous studies have reported significantly better outcomes when drug courts developed a written strategy for delivering incentives and sanctions that was distributed to all team members, participants, and other interested parties.	Having a written strategy was effective when guidelines had a medium level of predictability (i.e., menu-based versus prescriptive) and when teams prioritized their use. Mixed and limited effects were found for merely having guidelines as part of policy.
Studies also find that outcomes are significantly better when staff periodically remind participants about their obligations in the program and the responses for meeting or not meeting those obligations.	None
Most treatment court teams reserve discretion to modify their responses in light of participants' individualized needs, and studies in drug courts have found that employing reasonable discretion in incentive and sanctioning practices was associated with significantly better outcomes.	Courts performed best when there was: (1) a larger number of available incentives and sanctions (i.e., do not consistently use a limited number of responses), and (2) there was a medium predictability to sanctions (i.e., menu-based system vs. prescriptive).
Treatment courts that include community supervision officers or law enforcement officers on their teams have significantly better outcomes.	Those roles are important to participant success, but team members should employ evidence-based practices such as core correctional practices.
Studies in drug courts and probation have also found that frequent drug and alcohol testing was associated with significantly higher program completion rates and lower rates of positive drug tests and criminal recidivism.	Treatment courts are more effective than traditional responses like probation because participants are tested more frequently. Research indicates testing should be at least twice weekly during Phase I, but it's unclear if this should continue beyond that, as no research exists yet.
Outcomes are also better when drug courts and other criminal justice programs employ substance use monitoring tests or practices that extend the time window for detection, such as sweat patches, continuous alcohol monitoring devices, or EtG/EtS testing.	There are a limited number of studies on sweat patches in treatment courts, but the research on continuous alcohol monitoring and EtG/EtS testing indicates improved outcomes with their use.
Numerous studies in adult drug courts have reported significantly better outcomes when participants attended status hearings on a biweekly basis (every 2 weeks) during the first phase of the program.	None
Studies in drug courts and other community corrections programs confirm that outcomes are significantly better when participants have more opportunities to earn incentives for their accomplishments than to receive sanctions for infractions, ideally at a 4:1 ratio of incentives to sanctions.	The 4:1 ratio has not been directly tested in the treatment court literature. However, available research indicates that a greater number of compliments/verbal praise is associated with higher graduation rates, lower substance use, and lower criminal activity.
Payment of treatment fees does not improve treatment outcomes.	The payment of treatment fees on treatment outcomes has not been directly tested in the treatment court literature.

(continued)

Table 3. (continued)

Claim	Additional explanation
In the criminal justice system, high-magnitude sanctions, including jail detention lasting up to a few weeks, have been shown to improve outcomes for high-risk (but not high-need) individuals on probation or pretrial supervision when the sanctions were delivered for avoidable infractions with certainty, celerity, and procedural fairness.	Swift, certain, and fair/graduated sanction practices appear to be more effective than sanctioning practices that do not follow that model. However, research suggests the most effective components are the certainty of detection and response, with a portion of that response requiring the addition of evidence-based therapeutic responses (i.e., targeting criminogenic needs).
Brief intervals of jail detention have been associated with better outcomes in drug courts, but only when they were no longer than 3 to 6 days in length and were delivered in the later phases of the program when participants could satisfy more demanding requirements.	The available research suggests that brief jail stays are associated with higher graduation rates when they are administered at least 31 days after a participant is admitted to the program.
A substantial body of research on procedural fairness or procedural justice has determined that sanctions are most effective when participants are given a fair opportunity to voice their perspective concerning factual controversies and the appropriateness of the sanction before it is imposed, and when they receive a clear rationale for the judge's decision.	None
Studies find that jail sanctions in the first 30 to 60 days of treatment court are associated with lower program completion rates and higher criminal recidivism.	Program completion rates are lower when jail sanctions are imposed during the first 30 days of the program.
In later months or phases of treatment court, when participants can satisfy more demanding requirements, jail sanctions for repeated proximal infractions have been associated with improved outcomes.	Jail sanctions early in the program decrease the likelihood of program completion, but research has not identified what is generally considered proximal and distal goals.
Most studies have reported minimal gains from providing substance use treatment in jails or prisons.	The impact of in-custody treatment depends on the use of evidence-based practices that are implemented with fidelity.
Although specific types of in-custody programs such as therapeutic communities (TCs) have been shown to improve outcomes, the benefits from these programs were attributable to the fact that they increased the likelihood that persons would enter and complete treatment after release from custody.	Therapeutic communities have been shown to be effective in reducing substance use and recidivism, especially when combined with aftercare programs.
Studies find that marijuana use . . . reduces the likelihood that participants will successfully complete drug court. . . and increases traffic accidents and fatalities.	Individuals who were terminated from the treatment court were more likely to have tested positive for marijuana.
Studies consistently find that most participants enter drug court or mental health court primarily to avoid a criminal conviction or incarceration and outcomes are consistently better when participants avoid a felony conviction or incarceration if they complete the program.	Many individuals enter treatment courts to avoid incarceration (30–80%), but the literature is mixed on whether avoiding convictions or incarceration impacts recidivism. One meta-analysis did show that dismissing or expunging charges reduced drug recidivism.
Criminal recidivism is significantly higher, cost-effectiveness is significantly lower, and racial and other cultural disparities are significantly greater in drug courts that discharge participants unsuccessfully for continued substance use.	Automatic termination policies for new drug possession charges appear to lead to negative and disparate outcomes.

outcomes in treatment courts. Beyond these claims, empirical questions remain about whether practices such as the use of incentives (e.g., the 10:1 principle), sanction grids, contingency management, and core correctional practices improve outcomes for treatment court participants. The literature is also limited to the appropriate sequencing of interventions, phase length/requirements, the use of peer support specialists and mentors, the impact of treatment quality, and the treatment of moderate-risk offenders in these courts. Thus, the current review highlights several research opportunities for academics to explore, “How do we improve treatment courts so we can reduce recidivism beyond the typical 10% to 15% reduction?”

The current study has several limitations. First, though we used a high-level approach to evaluate empirical evidence, other rating systems exist that assess various aspects of research, such as selection, attrition, reporting, and performance bias, in different ways. While the current review did not account for this level of detail within each cited study, further exploration of these elements may serve to illustrate the strengths or weaknesses of the cited. In addition, there are (inherently) subjective interpretations regarding the strength of the evidence. There was 78% concordance between two reviewers (one a subject-matter expert and one not) on a sample of claims. However, it is possible that one reviewer could rate a claim as having limited evidence while another rates it as moderate. Second, the current study only reviewed the cited literature within the ISSA Standard. As discussed, there may be a well-developed literature base supporting the use of specific practices, but that literature was not cited in the ISSA Standard and therefore was not included in the current review. The next step in evaluating the ISSA Standard may be to assess the strength of claims beyond the research cited in the ISSA Standard to determine the broader empirical support for the policies and practices. Third, a newer set of standards may be warranted that exhaustively and carefully spell out the benefits and shortfalls of adopting individual practices. While that is also outside the scope of this study, this is potentially a resolvable limitation.

Policy Implications and Next Steps

The findings from this study raise significant concerns regarding the evidentiary foundation of the ISSA section in the second edition of the Adult Treatment Court Best Practice Standards (the “Standards”). Although All Rise claims that the Standards reflect practices supported by rigorous, replicated research, this study reveals that the majority of ISSA claims do not meet this benchmark. More than 70% of the 72 claims were supported by either no empirical evidence or only limited evidence, and the most substantial gaps were found among claims introduced for the first time in the second edition. These results call into question the transparency of the revision process, the extent to which the second edition adheres to its own definitional criteria for best practices, and, thus, the legitimacy of the ISSA standard itself.

The implications of this study’s findings are far-reaching and both short- and long-term in nature. Treatment courts rely on the Standards to shape policy, staff training, and program design; all of which directly affect participant outcomes, public safety,

and the use of significant public funds.⁵ When guidance is not firmly rooted in methodologically rigorous research, courts may unknowingly implement practices that reduce program effectiveness or, worse, contribute to unintended harms such as continued substance misuse, increased surveillance, or incarceration. These program and participant consequences ripple outward, affecting families, communities, and confidence in the treatment court model itself.

The long-term damage to the evidence-based practice movement may exceed the short-term harm to individual programs and the people they serve. Selective or incomplete citation practices can undermine the credibility of the Standards overall and erode trust among practitioners, researchers, and policymakers. When practices represented as “evidence-based” are implemented and fail to produce results, not only are investments of time, money, and capital lost, but faith in evidence-based practices generally may wane among both investors and observers. This “poisons the well” for any legitimate evidence-based practices going forward. Essentially, when goodwill efforts to implement best practices are expended on standards falsely sold as evidence-based, eventually people will no longer be willing to try evidence-based practices at all, which will further harm programs, participants, and the field at large.

At the same time, the study identifies opportunities for constructive advancement. Nearly 30% of the ISSA claims are supported by moderate or strong empirical evidence, and several practices with limited support have at least one promising evaluation. As such, stakeholders may confidently adopt claims that have moderate or strong support. For claims supported by limited evidence, stakeholders should proceed cautiously, pairing implementation with monitoring to assess local impact and not requiring program adherence for funding. Beyond informing practice, the evaluation results raised unanswered empirical questions that warrant focused research attention, including the effectiveness of incentives, sanctioning frameworks, phase requirements, supervision practices, and treatment sequencing. Addressing these areas is critical for strengthening the scientific foundation of treatment court operations and enhancing outcomes for participants and the public.

Moving forward, the authors of this study deem several steps essential. First, the findings demonstrate the need for All Rise to reassess the evidentiary foundation of the ISSA standard and to conduct a systematic assessment of the claims contained in the remaining standards, through transparent engagement with researchers and practitioners, to ensure future revisions are grounded in rigorous evidence and clear methodological justification. The first edition of the Standards seemed to embrace a multidisciplinary approach to developing the guidance, drawing on the expertise of over 80 individuals rather than a small core. Strengthening the Standards will require an open and collaborative process that involves many individuals and clearly articulates how evidence is evaluated, how claims are vetted, and how new guidance aligns with (or departs from) established research. The second author has already notified several relevant parties of these findings and the concerns they have raised, initiating dialogue about the need for clarification, corrective action, and a more rigorous and accountable revision process moving forward. In addition, All Rise should either (1) remove claims that are not supported by the research or (2) recategorize practices lacking empirical support as “Innovative or Emerging Practices” and convey to stakeholders that these

practices have not been empirically vetted and should be cautiously implemented. Programs that adopt such practices should have strong researcher-practitioner partnerships where unanticipated problems and adverse events are carefully monitored.

Second, organizations that specialize in training, technical assistance, and research on the treatment court model are encouraged to help practitioners understand which practices are evidence-based and which are innovative or emerging. For instance, the authors will support the translation and dissemination of this study's findings to the diverse groups associated with the treatment court field. Funding will be provided to permit this article to be published open source, removing this important information from living behind a pay wall. In addition, the authors will develop and disseminate translational research products to convey this study's findings and implications to practitioners, policymakers, and other treatment court stakeholders. These materials will be designed to support informed decision-making and facilitate evidence-based implementation at the local level. In addition, the authors will host, broadcast, and house on their website a live panel session to present and discuss the study's findings, implications, recommendations, and updates on progress toward the essential steps identified by the study authors. These efforts are intended to assist practitioners in understanding the strength of the evidence for various treatment court practices.

Third, treatment court researchers should prioritize rigorous, context-sensitive studies to test the ISSA claims that lack empirical support. These efforts can be further supported by state and federal funding that enables researchers to use methodologically and/or statistically rigorous designs (e.g., randomized controlled trials, differences-in-differences, Bayesian models) to refine our knowledge of ISSA. The National Institute of Justice (NIJ) prioritizes funding for rigorous, applied research that advances knowledge about criminal justice practices, and thus, NIJ funding could support a large-scale evaluation of incentives and sanctions in treatment courts. In addition, some Bureau of Justice Assistance (BJA) grants (e.g., the Second Chance Act) give priority consideration to proposals that include random assignment of services. The BJA treatment court solicitations could include a similar priority statement and ground the solicitations' objectives in the application and refinement of behavior modification principles. By strengthening the research base and ensuring that future revisions to the Standards accurately reflect the best available evidence, the field can continue advancing the treatment court model in ways that promote participant well-being, improve public safety, and preserve its integrity.

Fourth, the field should conceptually examine how standards are created and how they are applied in day-to-day work in treatment courts. A 'standard' should inform and improve real-world practice via a set of standardized (hence the term 'standard') organizational behaviors that will reliably produce a beneficial on-the-ground impact upon wide-scale adoption. Failing to correctly and exhaustively situate a set of standards within the proper research base is inappropriate and can (or perhaps even will) cause logistical problems and/or inconsistent practices that undermine programmatic effectiveness, leading to ineffectiveness or – even worse – counterproductivity. While research on specialized and treatment courts has certainly come a long way in the last two decades, the second edition of the Standards misrepresents much of this work and, as a result, may

unwittingly be contributing to decreased cross-site reliability, replicability, and effectiveness in programmatic delivery, implementation, and evaluation. Based on notable inconsistencies, practitioners should avoid adopting the standards set forth in the second edition of the Adult Treatment Court Best Practice Standards. Until a more exhaustive set of Standards is developed, we recommend that court practitioners and researchers revert to the more rigorous recommendations and practices laid out in the Standards' first edition.

Conclusion

In sum, this study contributes to ongoing efforts to evaluate and strengthen the empirical foundations of treatment court practices. By identifying areas where the evidence base is robust and areas where it remains underdeveloped, the findings offer a roadmap for future research, revision, and collaboration. Engaging in this work collaboratively allows the field not only to correct course but to elevate the scientific and practical foundations of treatment courts nationwide. Because the Standards wield significant influence over policy, funding, and daily operations, they must be held to the highest level of methodological rigor and transparency. Sustained attention to the Standards' evidentiary foundation will help ensure that treatment courts remain aligned with the best available science and continue to evolve in ways that protect participants, strengthen program effectiveness, and maintain confidence in the treatment court model.

ORCID iD

Jamie C. Vaske  <https://orcid.org/0009-0006-2247-7231>

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

The current study reviewed sources cited in the Incentives, Sanctions, and Service Adjustments section of the 2023 Second Edition of the Adult Treatment Court Best Practice Standards. The first authors summary of all sources and the evidentiary strength of each claim are openly available at <https://github.com/AU-CRCP/ISSA>.

Notes

1. This study focuses on adult treatment courts and their best practice standards and does not address juvenile treatment courts. Juvenile treatment courts operate within the juvenile justice system, which has its own set of treatment courts and standards, the Juvenile Drug Treatment Court Guidelines.

2. Although the second edition of the Standards was released in two parts across 2023 and 2024, the full publication is dated 2023.
3. Literature is also excluded if it was not directly cited in reference to a claim; thus, a literature base may exist that supports the use of a specific practice, but the claim was not well cited in the ISSA section. The authors did not include these external sources in their review, as the focus of the current study is on evaluating the cited research and evidence of “spin” in the 2024 version of the ISSA section.
4. “As before [in the first edition], *best practices* [in the second edition] were defined as services or interventions that have been proven through at least two high-quality experimental or quasi-experimental studies, meta-analyses, or quantitative systematic reviews to improve outcomes in treatment courts, other correctional rehabilitation programs, and/or substance use, mental health, or trauma treatment” (All Rise, 2023, p. 2).
5. Millions of dollars are annually devoted to treatment court programs receiving training and technical assistance based on the Standards, and much federal and state funding for the field requires and/or supports adherence to the Standards.

References

- Abt Associates. (2018). *Evaluating the impact of probation and parole home visits*. <https://www.ojp.gov/pdffiles1/nij/grants/254342.pdf>
- All Rise. (2023). *Adult treatment court best practice standards: Definitive guidance for treatment court practitioners*.
- Altman, D. G. (1999). *Practical statistics for medical research*. Chapman & Hall/CRC Press.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing.
- Anestis, J. C., & Carbonell, J. L. (2014). Stopping the revolving door: Effectiveness of mental health court in reducing recidivism by mentally ill offenders. *Psychiatric Services*, 65(9), 1105–1112.
- Bahr, S. J., Masters, A. L., & Taylor, B. M. (2012). What works in substance abuse treatment programs for offenders? *The Prison Journal*, 92(2), 155–174.
- Baker, A., Young, K., Potter, J., & Madan, I. (2010). A review of grading systems for evidence-based guidelines produced by medical specialties. *Clinical Medicine*, 10(4), 358–363.
- Belenko, S. (2002). The challenges of conducting research in drug treatment court settings. *Substance Use & Misuse*, 37(12–13), 1635–1664.
- Belenko, S., Fagan, J. A., & Dumanovsky, T. (1994). The effects of legal sanctions on recidivism in special drug courts. *Justice System Journal*, 17(1), 53–81.
- Belenko, S., Peugh, J., Califano, J. A., Usdansky, M., & Foster, S. E. (1998). Substance abuse and the prison population: A three-year study by Columbia University reveals widespread substance abuse among the offender population. *Corrections Today*, 60, 82–89.
- Binswanger, I. A., Blatchford, P. J., Mueller, S. R., & Stern, M. F. (2013). Mortality after prison release: Opioid overdose and other causes of death, risk factors, and time trends from 1999 to 2009. *Annals of Internal Medicine*, 159(9), 592–600.
- 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.
- Boutron, I., Dutton, S., Ravaud, P., & Altman, D. G. (2010). Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA*, 303(20), 2058–2064.

- Boutron, I., Haneef, R., Yavchitz, A., Baron, G., Novack, J., Oransky, I., Schwitzer, G., & Ravaud, P. (2019). Three randomized controlled trials evaluating the impact of “spin” in health news stories reporting studies of pharmacologic treatments on patients’/caregivers’ interpretation of treatment benefit. *BMC Medicine*, 17(1), 105.
- Carey, S., Finigan, M., & Pukstas, K. (2008). *Exploring the key components of drug courts: A comparative study of 18 adult drug courts on practices, outcomes and costs*. NPC Research.
- Chiu, K., Grundy, Q., & Bero, L. (2017). ‘Spin’ in published biomedical literature: A methodological systematic review. *PLoS Biology*, 15(9), e2002173.
- Christy, A., Poythress, N. G., Boothroyd, R. A., Petrila, J., & Mehra, S. (2005). Evaluating the efficiency and community safety goals of the Broward County mental health court. *Behavioral Sciences & the Law*, 23(2), 227–243.
- Cissner, A. B. (2009). *The drug court model and persistent DWI: An evaluation of the Erie and Niagara DWI/Drug courts*. Center for Court Innovation.
- Cissner, A. B., Rempel, M., Franklin, A. W., Roman, J., Bieler, S., Cohen, R. L., & Cadoret, C. R. (2013). *A statewide evaluation of New York’s adult drug courts: Identifying which policies work best*. Center for Court Innovation.
- Connor, T. A. (2019). Legitimation in action: An examination of community courts and procedural justice. *Journal of Crime and Justice*, 42(2), 161–183.
- Deschenes, E. P., & Greenwood, P. W. (1994). Maricopa County’s drug court: An innovative program for first-time drug offenders on probation. *Justice System Journal*, 17(1), 99–115.
- Farrington, D. P., Gottfredson, D. C., Sherman, L. W., & Welsh, B. C. (2003). The Maryland scientific methods scale. In L. W. Sherman, D. P. Farrington, B. C. Welsh, & D. L. MacKenzie (Eds.), *Evidence-based crime prevention* (pp. 13–21). Routledge.
- Goldkamp, J. S. (1994). Miami’s treatment drug court for felony defendants: Some implications of assessment findings. *The Prison Journal*, 74(2), 110–166.
- Goldkamp, J. S. (1999). Challenges for research and innovation: When is a drug court not a drug court. In W. Clinton Terry, III (Ed.), *The early drug courts: Case studies in judicial innovation* (pp. 166–177). Sage.
- Granfield, R., Eby, C., & Brewster, T. (1998). An examination of the Denver drug court: The impact of a treatment-oriented drug-offender system. *Law & Policy*, 20(2), 183–202.
- Hartley, R. D., & Baldwin, J. M. (2019). Waging war on recidivism among justice-involved veterans: An impact evaluation of a large urban veterans treatment court. *Criminal Justice Policy Review*, 30(1), 52–78.
- Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2024). *Cochrane handbook for systematic reviews of interventions* (Version 6.5). Cochrane. <https://www.cochrane.org/authors/handbooks-and-manuals/handbook>
- Labrecque, R. M., Viglione, J., & Caudy, M. (2023). The impact of community supervision officer training programs on officer and client outcomes: A systematic review and meta-analysis. *Justice Quarterly*, 40(4), 587–611.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- Latimer, J., Morton-Bourgon, K., & Chrétien, J. A. (2006). *A meta-analytic examination of drug treatment courts: Do they reduce recidivism?* (pp. 1–24). Research and Statistics Division, Department of Justice.
- Longshore, D., Turner, S., Wenzel, S., Morral, A., Harrell, A., McBride, D., Deschenes, E., & Iguchi, M. (2001). Drug courts: A conceptual framework. *Journal of Drug Issues*, 31(1), 7–25.

- Lowder, E. M., Rade, C. B., & Desmarais, S. L. (2018). Effectiveness of mental health courts in reducing recidivism: A meta-analysis. *Psychiatric Services*, 69(1), 15–22.
- Lowenkamp, C. T., Holsinger, A. M., & Latessa, E. J. (2005). Are drug courts effective? A meta-analytic review. *Journal of Drug Issues*, 31(1), 177–206.
- Martin, S. S., Butzin, C. A., Saum, C. A., & Inciardi, J. A. (1999). Three-year outcomes of therapeutic community treatment for drug-involved offenders in Delaware: From prison to work release to aftercare. *The Prison Journal*, 79(3), 294–320.
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276–282.
- Mitchell, O., Wilson, D. B., Eggers, A., & MacKenzie, D. L. (2012). Assessing the effectiveness of drug courts on recidivism: A meta-analytic review of traditional and non-traditional drug courts. *Journal of Criminal Justice*, 40(1), 60–71.
- National Association of Drug Court Professionals. (2015). *Adult drug court best practice standards: Vol. I*.
- National Association of Drug Court Professionals, Drug Court Standards Committee, & United States. Drug Courts Program Office. (1997). *Defining drug courts: The key components*. US Department of Justice, Office of Justice Programs, Drug Courts Program Office.
- O'Connell, J. P., Nestlerode, E., & Miller, M. L. (1999). *Evaluation of the Delaware juvenile drug court diversion program*. Statistical Analysis Center, State of Delaware.
- Perlis, R. H. (2025). Characterizing spin in psychiatric clinical research literature using large language models. *JAMA Network Open*, 8(2), e2459500.
- Ranapurwala, S. I., Shanahan, M. E., Alexandridis, A. A., Proescholdbell, S. K., Naumann, R. B., Edwards, D., Jr., & Marshall, S. W. (2018). Opioid overdose mortality among former North Carolina inmates: 2000–2015. *American Journal of Public Health*, 108(9), 1207–1213.
- Rossman, S. B., Rempel, M., Roman, J. K., Zweig, J. M., Lindquist, C. H., Green, M., & Farole, D. J. (2011). *The multi-site adult drug court evaluation: The impact of drug courts* (Vol. 4). Urban Institute. <https://www.ojp.gov/pdffiles1/nij/grants/237112.pdf>
- Sarteschi, C. M., Vaughn, M. G., & Kim, K. (2011). Assessing the effectiveness of mental health courts: A quantitative review. *Journal of Criminal Justice*, 39(1), 12–20.
- Senjo, S. R., & Leip, L. A. (2001). Testing and developing theory in drug court: A four-part logit model to predict program completion. *Criminal Justice Policy Review*, 12(1), 66–87.
- Shaffer, D. K. (2006). *Reconsidering drug court effectiveness: A meta-analytic review*. University of Cincinnati.
- Shaffer, D. K. (2011). Looking inside the black box of drug courts: A meta-analytic review. *Justice Quarterly*, 28(3), 493–521.
- Smith, P., Schweitzer, M., Labrecque, R. M., & Latessa, E. J. (2012). Improving probation officers' supervision skills: An evaluation of the EPICS model. *Journal of Crime and Justice*, 35(2), 189–199.
- Stein, D. M., Deberard, S., & Homan, K. (2013). Predicting success and failure in juvenile drug treatment court: A meta-analytic review. *Journal of Substance Abuse Treatment*, 44(2), 159–168.
- Sturm, A., de Vogel, V., Menger, A., & Huibers, M. J. (2021). Changes in offender-rated working alliance in probation supervision as predictors of recidivism. *Psychology, Crime & Law*, 27(2), 182–200.
- Stutts, M., & Cohen, J. R. (2022). The role of hopelessness and procedural justice on depressive outcomes in serious adolescent offenders. *Law and Human Behavior*, 46(6), 415.

- Taxman, F. S., & Bouffard, J. (2002). Treatment inside the drug treatment court: The who, what, where, and how of treatment services. *Substance Use & Misuse*, 37(12–13), 1665–1688.
- Thomas, B. H., Ciliska, D., Dobbins, M., & Micucci, S. (2004). A process for systematically reviewing the literature: Providing the research evidence for public health nursing interventions. *Worldviews Evidence Based Nursing*, 1(3), 176–184. <https://doi.org/10.1111/j.1524-475X.2004.04006.x>
- West, S., King, V., Carey, T. S., Lohr, K. N., McKoy, N., Sutton, S. F., & Lux, L. (2002). *Systems to rate the strength of scientific evidence: Summary* (AHRQ evidence report summaries). Agency for Healthcare Research and Quality (US).
- Wexler, H. K., Melnick, G., Lowe, L., & Peters, J. (1999). Three-year reincarceration outcomes for Amity in-prison therapeutic community and aftercare in California. *The Prison Journal*, 79(3), 321–336.
- Wright, D., & Clymer, B. (2000). *Evaluation of Oklahoma Drug Courts, 1997-2000* [Technical report]. Oklahoma Criminal Justice Resource Center.
- Yavchitz, A., Ravaud, P., Altman, D. G., Moher, D., Hrobjartsson, A., Lasserson, T., & Boutron, I. (2016). A new classification of spin in systematic reviews and meta-analyses was developed and ranked according to the severity. *Journal of Clinical Epidemiology*, 75, 56–65.
- Zettler, H. R., & Martin, K. D. (2020). Exploring the impact of technical violations on probation revocations in the context of drug court. *American Journal of Criminal Justice*, 45(6), 1003–1023.

Author Biographies

Jamie C. Vaske is the founding associate director of the Center for Research and Collaborative Partnerships (CRCP) at American University. Her research interests focus on treatment courts, translational research, and quantitative methods.

Julie Marie Baldwin is interim vice provost for research and innovation and the founding director of the CRCP at American University. Her extramural research focuses on treatment courts and behavioral health, employing strong researcher-practitioner partnerships.

John H. Boman, IV is a Research Professor at American University's School of Public Affairs in the Justice, Law & Criminology program and the Center for Research and Collaborative Partnerships. Dr. Boman is a quantitative criminologist who specializes in crime policy (especially surrounding the use of narcotics), measurement, construct validation, and latent variable modeling. His recent work has appeared in top-tier criminology (e.g., *Criminology*, *Journal of Criminal Justice*) and general science journals (e.g., *Nature Human Behaviour*, *PLOS One*).