

List of Exhibits

Ex. 11-Haney

Ex. 12-Specialty Guidelines for Forensic Psychology

Ex. 13-Kupers Report

Ex. 14-Heilburn, et al., (1990) – An MMPI-based empirical model of malingering
a

Ex. 15-Melton, et al.,

Ex. 16-BCCLA v. Canada (2018)

Ex. 17-Colorado Study

Ex. 18-Smith (2011)

Ex. 19-Grassian & Kupers

Ex. 20-Zinger, et al., [NEW] – (2001) The psychological effects of 60 days in A

EXHIBIT 11

**Craig Haney,
"The Psychological Effects of
Solitary Confinement"**

Craig Haney

The Psychological Effects of Solitary Confinement: A Systematic Critique

ABSTRACT

Research findings on the psychological effects of solitary confinement have been strikingly consistent since the early nineteenth century. Studies have identified a wide range of frequently occurring adverse psychological reactions that commonly affect prisoners in isolation units. The prevalence of psychological distress is extremely high. Nonetheless, use of solitary confinement in the United States vastly increased in recent decades. Advocates defend its use, often citing two recent studies to support claims that isolation has no significant adverse psychological effects, including even on mentally ill people. Those studies, however, are fundamentally flawed, their results are not credible, and they should be disregarded. Critically and comprehensively analyzing the numerous flaws that compromise this recent scholarship underscores the distinction between methodological form and substance, the danger of privileging quantitative data irrespective of their quality, and the importance of considering the fraught nature of the prison context in which research results are actually generated. Solitary confinement has well-documented adverse effects. Its use should be eliminated entirely for some groups of prisoners and greatly reduced for others.

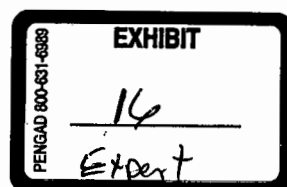
Doing prison research, Alison Liebling has long reminded us, is deeply emotional and intellectually challenging, with different methodological approaches “competing for epistemological prominence—often from different sides of the prison wall” (1999, p. 148). It takes place in “an in-

Electronically published March 9, 2018

Craig Haney is Distinguished Professor of Psychology, University of California, Santa Cruz.

© 2018 by The University of Chicago. All rights reserved.
0192-3234/2018/0047-0004\$10.00

365



VERMILLION 004821

tense, risk-laden, emotionally fraught environment” (p. 163) and within a closed environment in which prison administrators tightly control access to data and most prisoners manifest an entirely legitimate and understandable skepticism toward data gatherers.

This helps explain why, in Liebling’s words, “the pains of imprisonment are tragically underestimated by conventional methodological approaches to prison life” (p. 165). The more these conventional approaches encourage us to conceive of prisons as more or less traditional research settings and prisoners as mere specimens to be “objectively assessed,” the less likely we are to gain useful insights into prison life or accurately represent the experience of those living inside.

These cautions are doubly applicable to research on solitary confinement.¹ It involves involuntary isolation of prisoners nearly around the clock in sparse cells located in remote or inaccessible units. Solitary confinement denies prisoners any meaningful social contact and access to positive environmental stimulation.

These prisons within prisons are nearly impenetrable to outside researchers (or anyone else). Prison officials tightly control access to solitary confinement units and to the prisoners inside them. They typically rebuff attempts by researchers to observe conditions and practices, let alone to carefully assess their potentially harmful effects. Prisoners in solitary confinement tend to be even more self-protective than other prisoners are (as part of their accommodation to harsh and frequently abusive conditions) and reluctant to have their “measure” taken by persons whom they have no reason to trust. They generally subscribe strongly to prisoner norms against displaying or acknowledging vulnerabilities that could be interpreted as weakness. The inapt pejorative designation of them as collectively “the worst of the worst” does not inspire confidence in or candor toward outsiders, and certainly not toward anyone remotely associated with the prison administration.

These realities pose a host of methodological challenges for anyone interested in understanding the nature and effects of prison isolation. This is in part why studies of the effects of solitary confinement on prisoners

¹ I use “solitary confinement” to refer to forms of prison isolation in which prisoners are housed involuntarily in their cells for upward of 23 hours per day and denied the opportunity to engage in normal and meaningful social interaction and congregate activities, including correctional programming. The term subsumes a range of prison nomenclature including “administrative segregation,” “security housing units,” “high security,” and “close management,” among others.

have rarely, if ever, approximated experimental research designs (including quasi- or natural experimental designs).

Solitary confinement units not only are largely impenetrable to outsiders but also, of course, are subject to legal and ethical restrictions that preclude random assignment of prisoners into them. The rigid prison rules and operating procedures that govern these places can easily frustrate the use of the kind of meticulous controls over conditions and participants that are needed to carry out anything remotely resembling an experiment. The distinctiveness of solitary confinement units and the nonnegotiable staff mandates under which they operate make it difficult, if not impossible, to implement rigorous conventional research designs (e.g., representative samples, control groups, repeated measures). Efforts to conduct randomized or truly controlled studies inevitably face significant risks that the data collected will be so confounded by inevitable methodological compromises as to be uninterpretable and, therefore, meaningless.

Nonetheless, scholars and researchers know a great deal about the negative effects of solitary confinement. We have firsthand or autobiographical accounts by former prisoners (e.g., Burney 1961) and staff members (e.g., Rundle 1973; Slater 1986); ethnographic, interview, and observational research (e.g., Benjamin and Lux 1975; Toch 1975; Hilliard 1976; Jackson 1983; Rhodes 2004; Reiter 2016); and cross-sectional studies that assess prisoners' psychological reactions at particular times (e.g., Grassian 1983; Brodsky and Scogin 1988; Haney 2003).

Much of the important research is qualitative, but there is a substantial amount of it and the findings are robust. They can also be "triangulated," that is, studied through a range of methods and in settings sometimes similar but not necessarily identical to solitary confinement (e.g., Turner, Cardinal, and Burton 2017). Numerous literature reviews have noted that scientists from diverse disciplinary backgrounds, working independently and across several continents, and over many decades, have reached almost identical conclusions about the negative effects of isolation in general and solitary confinement in particular (e.g., Haney and Lynch 1997; Haney 2003; Grassian 2006; Smith 2006; Arrigo and Bullcock 2008). Those robust findings are also theoretically coherent. That is, they are consistent with and explained by a rapidly growing literature on the importance of meaningful social contact for maintenance of mental and physical health.

Largely because of the robustness and theoretical underpinnings of the data, numerous scientific and professional organizations have reached

a broad consensus about the damaging effects of solitary confinement. Several years ago, for example, a National Academies of Science committee reviewed the existing research and concluded that solitary confinement can precipitate such “serious psychological change” in prisoners that the practice “is best minimized” (National Research Council 2014, p. 201). The American Psychological Association (2016, p. 1), the world’s largest professional association of psychologists, asserted that “solitary confinement is associated with severe harm to physical and mental health among both youth and adults, including: increased risk of self-mutilation, and suicidal ideation; greater anxiety, depression, sleep disturbance, paranoia, and aggression; exacerbation of the onset of pre-existing mental illness and trauma symptoms; [and] increased risk of cardiovascular problems.”

Similarly, the National Commission on Correctional Health Care (2016), a highly respected organization of correctional medical and mental health professionals, promulgated a series of “principles” with respect to solitary confinement. They are intended to guide the ethical conduct of its members, including that placement in solitary confinement for longer than 15 days represents “cruel, inhumane, and degrading treatment” that is “harmful to an individual’s health” (p. 260) and that “health care staff must advocate” to remove persons from solitary confinement whenever “their medical or mental health deteriorates” (p. 261).

Summarizing this growing consensus, a joint 2016 statement of the Association of State Correctional Administrators (the largest professional association of American prison administrators) and Yale Law School’s Liman Public Interest Program observed that demands for change in use of solitary confinement are being made around the world. More specifically,

Commitments to reform and efforts to limit or abolish the use of isolating confinement come from stakeholders and actors in and out of government. Documentation of the harms of isolation, coupled with its costs and the dearth of evidence suggesting that it enhances security, has prompted prison directors, legislatures, executive branch officials, and advocacy groups to try to limit reliance on restricted housing. Instead of being cast as the solution to a problem, restricted housing has come to be understood by many as a problem in need of a solution. (Association of State Correctional Administrators and the Arthur Liman Public Interest Program 2016, p. 15)

Even more recently, the director of the Colorado Department of Corrections, Rick Raemisch, announced that Colorado has ended use of long-term solitary confinement, so that even prisoners “who commit serious violations like assault will now spend at most 15 days in solitary” (2017, p. A25). This development in Colorado is especially notable, for reasons that become clear in the pages that follow.

Against this backdrop, in 2009 and 2010 word began to circulate among prison researchers and policy makers that a new, supposedly unassailable scientific study—the “Colorado study”—had produced results that contravened many decades of empirical findings on the harmful effects of prison isolation. Lovell and Toch (2011, p. 3) characterized a number of its findings as “flabbergasting,” and indeed they were. Among the most startling were that a year-long stay in solitary confinement resulted in no “significant decline in psychological well-being over time”; that on most measures, including cognitive performance, “there was improved functioning over time”; and most remarkably that many more mentally ill prisoners benefited from isolation than were damaged by it (O’Keefe et al. 2010, pp. 54, 78). The Colorado researchers thus reported data indicating that solitary confinement made prisoners feel and think better, especially if they were mentally ill.

In fact, however, the Colorado study was riddled with serious methodological problems that limited its value and made the meaning of the results impossible to decipher. Notwithstanding its authors’ frank, albeit at times opaque and oblique, acknowledgments of some of its fundamental weaknesses, defenders of solitary confinement have seized on it. It has become a last bastion of resistance against a widespread and growing consensus that use of solitary confinement should be eliminated or drastically limited.

The Colorado study’s influence has been amplified by an equally flawed meta-analysis that relied very heavily on it and significantly mischaracterized the prior literature on the effects of isolated confinement (Morgan et al. 2016). Of course, the influence of a fundamentally flawed study can grow if it and the data it produced are included in literature reviews that overlook glaring weaknesses. This risk is greater in meta-analytic than in narrative literature reviews that focus on decontextualized “effect sizes” irrespective of methodological shortcomings of individual studies. Unlike narrative reviews, meta-analyses include only quantitative outcomes or effects. This elevates the importance of numerical outcomes and often

scants nuanced assessments of data quality. This is particularly a problem for prison research, an enterprise that is fraught with emotional and methodological challenges, in which aspects of the institutional context or setting can fundamentally alter the nature of the research and the meaning of its results. That is precisely what happened in the Morgan et al. (2016) meta-analysis.

In the following pages, I first discuss the scientific basis for the broad consensus that solitary confinement has substantial negative psychological effects on prisoners. I then discuss the Colorado study and the Morgan et al. (2016) meta-analysis based largely on it. Both are textbook examples of how things can go terribly wrong when researchers fail to take account of the unique nature of the prison environment, the special emotional and methodological challenges of prison research in general, and the contingent and unpredictable conditions and practices that affect solitary confinement units in particular.

I. Solitary Confinement Research and Practice

Documentation of the damaging nature and psychological effects of solitary confinement has a very long history, dating at least to the early nineteenth century, when solitary confinement was the modal form of imprisonment. The notion that prisoners could be reformed—made “penitent”—by time spent in isolation dominated American correctional thinking and practice and eventually spread throughout Europe. Yet the practice was recognized as a dangerous failure not long after its inception. Haney and Lynch (1997), Toch (2003), Grassian (2006), and Smith (2006) reviewed much of the early historical literature. Reports on solitary confinement at Pentonville Prison in England described “twenty times more cases of mental disease than in any other prison in the country” (Hibbert 1963, p. 160). Accounts of solitary confinement in the Netherlands documented “again and again, reports of insanity, suicide, and the complete alienation of prisoners from social life” (Franke 1992, p. 128). Newspaper reports from Philadelphia observed that prisoners in solitary confinement at the Walnut Street Jail “beg, with the greatest earnestness, that they may be hanged out of their misery” (Masur 1989, p. 83). Charles Dickens concluded that a prisoner kept in that “melancholy house” was like “a man buried alive . . . dead to everything but torturing anxieties and horrible despair” (Dickens 1842, p. 116). A similar regime in Auburn, New York, was described as “a hopeless failure that led to a

marked prevalence of sickness and insanity on the part of convicts in solitary confinement” (Barnes 1921, p. 53). Stuart Grassian (2006, pp. 342–43) reported that “between 1854 and 1909, thirty-seven articles appeared in German scientific journals on the subject of psychotic disturbances among prisoners.” The “most consistent factor” accounting for prison psychoses, “reported in over half the total literature, was solitary confinement.”

Systematic early studies of solitary confinement in the United States used what is now seen as a somewhat outmoded theoretical framework, focusing narrowly on sensory rather than social deprivation (e.g., Scott and Gendreau 1969; Gendreau et al. 1972). Even so, the authors of one early study concluded that “excessive deprivation of liberty, here defined as near complete confinement to the cell, results in deep emotional disturbances” (Cormier and Williams 1966, p. 484). In a review of the sensory deprivation literature, Haney and Lynch (1997) noted that “the dissimilarities between conditions created in these studies and those in solitary confinement or punitive segregation in correctional institutions are obvious.” They also observed that, nonetheless, the early research did “emphasize the importance of sensory stimulation in human experience and the dramatic effects that can be produced when such stimulation is significantly curtailed” (p. 502).

More recent research focuses on the psychological damage that results from social deprivation. Hans Toch’s large-scale psychological study of prisoners in crisis in New York State correctional facilities included important observations about the effects of isolation. After conducting numerous in-depth interviews, Toch (1975, p. 54) concluded that “isolation panic” was a serious problem in solitary confinement. The symptoms Toch described included rage, panic, loss of control and breakdowns, psychological regression, and build-ups of physiological and psychic tension that led to incidents of self-mutilation. He noted that isolation panic could occur under other conditions of confinement but that it was “most sharply prevalent in segregation.” Moreover, it marked an important dichotomy for prisoners: the “distinction between imprisonment, which is tolerable, and isolation, which is not.”

Empirical studies have identified a wide range of frequently occurring adverse psychological reactions to solitary confinement.² These include

² For reviews of the literature documenting these adverse reactions, see Haney and Lynch (1997), Haney (2003), Cloyes et al. (2006), Grassian (2006), Smith (2006), and Arrigo and Bullock (2008).

stress-related reactions (such as decreased appetite, trembling hands, sweating palms, heart palpitations, and a sense of impending emotional breakdown); sleep disturbances (including nightmares and sleeplessness); heightened levels of anxiety and panic; irritability, aggression, and rage; paranoia, ruminations, and violent fantasies; cognitive dysfunction, hypersensitivity to stimuli, and hallucinations; loss of emotional control, mood swings, lethargy, flattened affect, and depression; increased suicidality and instances of self-harm; and, finally, paradoxical tendencies to further social withdrawal.

The prevalence of psychological distress, at least as suffered in certain solitary confinement settings, appears to be extremely high. A study conducted at the Security Housing Unit (SHU) at Pelican Bay State Prison in California (Haney 1993; Reiter 2016), an especially severe solitary confinement facility, is illustrative. Structured interviews were used to assess a randomly selected, representative sample of 100 prisoners to determine the prevalence of symptoms of psychological stress, trauma, and isolation-related psychopathology (Haney 2003). The interviews included demographic questions, brief social and institutional histories, and systematic assessments of 25 items, based in part on the Omnibus Stress Index (Jones 1976) and on other instruments similar to those used in Brodsky and Scogin (1988). Every symptom of psychological stress and trauma but one (fainting) was experienced by more than half of the assessed prisoners; many were reported by two-thirds or more and some by nearly everyone. Well over half of the prisoners reported distress-related symptoms—headaches, trembling, sweaty palms, and heart palpitations.

High numbers of the Pelican Bay SHU prisoners also reported suffering from isolation-related symptoms of pathology. Nearly all reported ruminations or intrusive thoughts, oversensitivity to external stimuli, irrational anger and irritability, difficulties with attention and often with memory, and a tendency to withdraw socially. Almost as many reported symptoms indicative of mood or emotional disorders: concerns over emotional flatness or losing the ability to feel, swings in emotional response, and feelings of depression or sadness that did not go away. Finally, sizable minorities reported symptoms that are typically associated only with more extreme forms of psychopathology—hallucinations, perceptual distortions, and thoughts of suicide.

Social withdrawal, a common reaction to solitary, is related to a broader set of social pathologies that prisoners often experience as they attempt to

adapt to an environment devoid of normal, meaningful social contact. In order to exist and function in solitary confinement, where day-to-day life lacks meaningful interaction and closeness with others, prisoners have little choice but to adapt in ways that are asocial and, ultimately, psychologically harmful.

A large international literature has reached similar conclusions on the adverse psychological effects of solitary confinement. Solitary confinement not only is a common form of mistreatment to which prisoners of war have been subjected and been adversely affected (e.g., Hinkle and Wolff 1956) but also is associated with “higher levels of later life disability” among returnees (Hunt et al. 2008, p. 616). It is frequently used as a component of torture (e.g., Foster, Davis, and Sandler 1987; Nowak 2006; Reyes 2007). Solitary confinement has been studied in more traditional international criminal justice contexts as well. For example, Barte (1989, p. 52) concluded that solitary confinement in French prisons had such “psychopathogenic” effects that prisoners placed there for extended periods could become schizophrenic, making the practice unjustifiable, counterproductive, and “a denial of the bonds that unite humankind.”

Koch (1986, pp. 124–25) studied “acute isolation syndrome” among detainees in Denmark that occurred after only a few days in isolation and included “problems of concentration, restlessness, failure of memory, sleeping problems and impaired sense of time and ability to follow the rhythm of day and night.” If isolation persisted for a few weeks or more, it could lead to “chronic isolation syndrome,” including intensified difficulties with memory and concentration, “inexplicable fatigue,” a “distinct emotional liability” that included fits of rage, hallucinations, and the “extremely common” belief among prisoners that “they have gone or are going mad.”

Volkart, Dittrich, et al. (1983) studied penal isolation in Switzerland. They concluded that, compared with prisoners in normal confinement, those in solitary displayed considerably more psychopathological symptoms, including heightened feelings of anxiety, emotional hypersensitivity, ideas of persecution, and thought disorders (see also Waligora 1974; Volkart, Rothenfluh, et al. 1983; Bauer et al. 1993).

The major reviews of the literature reach the same conclusions as the seminal studies. Haney and Lynch (1997, pp. 530, 537) noted that “distinctive patterns of negative effects have emerged clearly, consistently, and unequivocally from personal accounts, descriptive studies, and sys-

tematic research on solitary and punitive segregation.” The “psychologically destructive treatment” to which prisoners are exposed in solitary confinement is so severe that it likely “would not be countenanced for any other group in our society.”

Grassian’s extensive survey of solitary confinement research concluded that “the restriction of environmental stimulation and social isolation associated with confinement in solitary are strikingly toxic to mental functioning, including, in some prisoners, a stuporous condition associated with perceptual and cognitive impairment and affective disturbances” (2006, p. 354).

That same year, Smith’s comprehensive review concluded that “the vast majority” of studies on the effects of solitary confinement “document significant negative health effects” (2006, p. 456). He observed that “research on effects of solitary confinement has produced a massive body of data documenting serious adverse health effects” (p. 475) including “anger, hatred, bitterness, boredom, stress, loss of the sense of reality, suicidal thoughts, trouble sleeping, impaired concentration, confusion, depression, and hallucinations” (p. 488).

Similarly, Arrigo and Bullock (2008) concluded that “nearly all investigators acknowledge that long-term segregation, mistreatment by correctional staff, and preexisting psychological vulnerability are all apt to result in negative mental health consequences for convicts” and that “the extreme isolation and harsh conditions of confinement in [solitary confinement] typically exacerbate the symptoms of mental illness” (p. 632).

There is an important, theoretically coherent framework that helps explain the consistency of these conclusions. A burgeoning literature in social psychology and related disciplines shows that solitary confinement is a potentially harmful form of sensory deprivation but also, and more destructively, exposes prisoners to pathological levels of social deprivation. Numerous studies have established the critical psychological significance of social contact, connectedness, and belonging (e.g., Fiorillo and Sabatini 2011; Hafner et al. 2011; Cacioppo and Cacioppo 2012). Meaningful social interactions and social connectedness can have a positive effect on people’s physical and mental health in settings outside of prison and, conversely, social isolation in general can undermine health and psychological well-being. Thus, it makes sound psychological sense that exposure to especially severe forms of material, sensory, and social deprivation harms prisoners’ mental health.

Indeed, researchers have concluded that human brains are “wired to connect” to others (Lieberman 2013). Thwarting the need to establish and maintain connections to others undermines psychological well-being and increases physical morbidity and mortality. Because “social connection is crucial to human development, health, and survival,” experts have called for it to be recognized as a national public health priority (Holt-Lunstad, Robles, and Sbarra 2017, p. 527). The involuntary, coercive, hostile, and demeaning aspects of solitary confinement are likely to exacerbate the negative effects of social isolation that have repeatedly been documented in more benign contexts.

Given these long-standing and theoretically informed findings, a study purporting to show that psychological effects of solitary confinement range from harmless to beneficial would normally not be taken seriously. Sometimes, however, the appearance of seemingly objective scientific findings provides legitimacy to doubtful conclusions, especially when they support contested policy or political agendas. That is precisely what happened in the case of the Colorado study. Its authors described it as a scientific advance over all previous studies, and some commentators prematurely lauded its methodological rigor. It appeared on the surface to be an ambitious and well-designed longitudinal study, with appropriate comparison groups and a host of dependent variables that were to be examined. Data were collected through the repeated administration of instruments said to be validated, and an unusually large number of prisoners were to be assessed over a 1-year period.

The reality was very different. The project could not be, and was not, carried out as planned, partly because of powerful demands and correctional contingencies inherent in prison settings in general and solitary confinement in particular. The problems proved insurmountable: comparison groups were not comparable, and the integrity of the “treatments” each group received was quickly corrupted. I discuss these and numerous other problems in the next section. The fundamental methodological flaws that plagued the study prevented collection of any meaningful data and ensured that no meaningful conclusions could be drawn.

The Colorado study nonetheless has continued to play an outsized role in contentious policy debates in which proponents of solitary confinement draw on it to support positions that are becoming indefensible. Defenders have characterized the study as “an outstanding example of applied correctional research” that was “planned with great care,” em-

ployed a “rigorous” design, and produced results that “were about as conclusive as possible” showing that solitary confinement has few or no adverse effects (Gendreau and Labrecque 2016, p. 9).

A year after the study’s release, the National Institute of Corrections devoted an entire issue of *Corrections and Mental Health* to discussion of it. One writer (other than the Colorado researchers themselves) who endorsed its results and defended its methodology was Paul Gendreau, a well-known Canadian researcher and long-time prison system employee. Despite not having published primary research data on isolation since the early 1970s, he had defended its use over many decades, for example, in a 1984 article entitled “Solitary Confinement Is Not Cruel and Unusual: People Sometimes Are!” (Gendreau and Bonta 1984). In *Corrections and Mental Health*, Gendreau hailed the Colorado study as a “truly significant contribution to our knowledge base about the effects of prison life for one of the most severe forms of incarceration” and asserted that “in terms of its methodological rigor” no other study “comes close” (Gendreau and Theriault 2011, p. 1). Moreover, despite the deep skepticism voiced by all of the other contributors to the special issue except Gendreau and the study’s authors, the journal’s editor described the Colorado study as “an important report” because it showed that “administrative segregation is not terribly harmful” (Immarigeon 2011, p. 1).

Similarly, when a brief summary of the study appeared in a scholarly journal (O’Keefe et al. 2013), it was accompanied by commentary written by several prominent clinicians who claimed to have witnessed as much as or more psychological improvement among isolated prisoners than decompensation. They praised the study as “groundbreaking” and described its methodology as “solid” (Berger, Chaplin, and Trestman 2013, pp. 61–63). The authors averred that “the extremes of solitary confinement have been misunderstood” and that “people are resilient and are able to thrive under even difficult environmental conditions.”

The respected Irish prison researcher Ian O’Donnell, though more circumspect, offered similar observations. Although O’Donnell acknowledged some limitations, he praised the study’s methodology and invoked its results to support some of his own views. “However unpalatable they might appear to some parties,” he asserted, the study’s findings “must be taken seriously” (2014, p. 120). O’Donnell characterized the study as “valuable” because, he said, it “highlights the individual’s capacity to adapt” (p. 122). He defended the Colorado researchers against criticism, noting that it is ethically impossible to study solitary confinement with “suffi-

cient scientific rigour to satisfy everyone” (p. 122). The study’s results suggest, he wrote, “that segregation was not highly detrimental to those forced to endure it” (p. 120) and that the harmfulness of this form of penal confinement “may have been over-emphasized” (p. 123).³

The Colorado study also figures prominently in correctional policy reviews by recalcitrant prison officials who do not want to modify segregation practices and in litigation over the harmful effects of solitary confinement, where those defending it are eager to find support.⁴ For example, the US Government Accountability Office conducted a review of segregated housing practices in the federal Bureau of Prisons (BOP): “BOP HQ officials cited the 2010 DOJ-funded study of the psychological impact of solitary confinement in the Colorado state prison system. This study showed that segregated housing of up to 1 year may not have greater negative psychological impacts than non-segregated housing on inmates. While the DOJ-funded study did not assess inmates in BOP facilities, BOP management told us this study shows that segregation has

³ O’Donnell indicated that the study documented the “benefits” of solitary, ones he suggested derived from “the many hours spent in quiet contemplation” in solitary confinement units. He also suggested that the results buttressed his own belief that “severe forms of trauma are sometimes accompanied by an improvement in functioning” (p. 123).

⁴ For example, consider the “Expert Report by Robert Morgan, PhD, Ashker, et al. v. Governor, et al., Case No.:C09-05796 CW (N.D. Cal.)” submitted under oath to a federal district court. Morgan opined that being housed in extremely harsh solitary confinement (the SHU in California’s Pelican Bay State Prison) for “*ten or more continuous years* does not place inmates at substantial risk of serious mental harm” (p. 1; emphasis added), a position that he supported in part by citing the Colorado study. He described the study as “the most sophisticated study to date on the topic” of the effects of solitary confinement, claimed it showed “an absence of adverse effects for segregated inmates” (p. 1), and cited the results of his own meta-analysis (which was incorporated into Morgan et al. [2016], which I discuss later in this essay) to buttress his defense of long-term solitary confinement. Similarly, see the “Expert Report Provided in the Matter of BCCLA and JHS v. AGC, Court No.:S150415” by Jeremy Mills, PhD, filed in support of the continued use of solitary confinement in Canadian prisons. The Colorado study is described by Mills as “quite likely the most sophisticated longitudinal study to date examining the effects of segregation on mentally ill and non-mentally ill offenders” (p. 13). He also characterized meta-analyses like the Morgan et al. meta-analysis, of which he was a coauthor, as “a hallmark of the scientific process” (p. 12). Mills embraced the Colorado study’s conclusions as supportive of his own, which were gleaned from his “clinical experience” working in segregation units on behalf of the Canadian Correctional Service. These included his view that both mentally ill and non-mentally ill prisoners usually need only “a few days” of “a period of adjustment” to get used to solitary confinement. He suggested that prisoners placed in solitary confinement “more frequently” forgo the adjustment period entirely because “they are familiar with the environment” (p. 14). Neither Morgan nor Mills acknowledged the Colorado study’s numerous fundamental methodological flaws or indicated that the Morgan et al. meta-analysis on which they relied was based primarily on it.

little or no adverse long-term impact on inmates” (Government Accountability Office 2013, p. 39).

The Colorado study’s continuing cachet in prison policy making and important legal circles means that its scientific bona fides bear especially careful analysis. Examining and deconstructing its methodology is a tedious but worthwhile exercise because it illustrates the difficulty of honoring norms of scientific rigor in a setting in which conventional research designs are nearly impossible to implement and necessary trade-offs are especially costly to the quality of the data collected. I turn to that exercise in Section II and to a deconstruction of the Morgan et al. (2016) meta-analysis in Section III.

II. Interrogating the Colorado Study

Results of the Colorado study appeared in two versions: a lengthy final report to the National Institute of Justice (O’Keefe et al. 2010) and a short article in the *Journal of the American Academy of Psychiatry and Law* (O’Keefe et al. 2013). I mostly discuss the more detailed National Institute of Justice report.⁵ I also draw on two depositions, under oath, of Maureen O’Keefe, the lead researcher, in connection with prisoner litigation concerning Colorado’s “supermax” facility (where much of the study was conducted). In response to detailed questions, O’Keefe discussed numerous issues not raised in the report or fully addressed in published exchanges following its release.⁶

Why the study was undertaken is unclear. Neither of the primary researchers had prior experience with solitary confinement. Maureen O’Keefe had a master’s degree in clinical psychology but no prior involvement in research on the effects of isolation. Kelli Klebe was a psychometrician who also had no direct experience with solitary confinement (O’Keefe 2010, pp. 13–14). Yet they designed the study (pp. 77–79).

The study’s impetus may have come from Larry Reid, warden of the Colorado supermax prison that housed prisoners assigned to administra-

⁵ A number of brief but highly critical commentaries by prison researchers also questioned aspects of the methodology: Grassian and Kupers (2011), Rhodes and Lovell (2011), Shalev and Lloyd (2011), and Smith (2011). See also the response to at least some of these criticisms by Metzner and O’Keefe (2011).

⁶ The two depositions are Deposition of Maureen O’Keefe, *Dunlap v. Zavaras*, Civil Action no. 09-CV-01196-CMA-MEH, October 5, 2010; and Deposition of Maureen O’Keefe at 96, 101 *Sardakowski v. Clements*, Civil Action no. 12-CV-01326-RBJ-KLM, October 25, 2013.

tive segregation. O'Keefe indicated that Reid "kept pushing for the study to be done" and served as a member of the study's advisory board (2010, p. 51). A few years before the Colorado study was planned, administrators at a Wisconsin supermax had lost a lawsuit over their use of solitary confinement (*Jones 'El v. Berge*, 164 F.Supp. 2d 1097 [W.D. Wis. 2001]), and Reid apparently wanted to avoid a similar decision. As O'Keefe (2013, p. 44) observed, "I believe [Reid's] concern was that Wisconsin had lost the case and it had severely restricted their ability to use administrative segregation."

The Colorado researchers said that they expected to find that administrative segregation had negative psychological effects: "We hypothesized that inmates in segregation would experience greater psychological deterioration over time than comparison inmates, who were comprised of similar offenders confined in non-segregation prisons" (O'Keefe et al. 2010, p. viii). If so, Warden Reid did not appear to share that view. The Colorado Department of Corrections then housed "three times as many people in solitary confinement as the average state prison system" (*Correctional News* 2012, p. 1). Moreover, O'Keefe (2013, p. 46) acknowledged that Reid "was very pro administrative segregation and all of us on the project felt that way."

Psychologist John Stoner, the mental health coordinator at the Colorado supermax prison, also strongly supported administrative segregation and served as a member of the study's advisory board. He had testified in the Wisconsin case that administrative segregation was not "as detrimental to mental health as others have found it to be" (*Jones 'El v. Berge*, p. 1104). Among other things, Stoner said that he was not troubled by Wisconsin's use of "boxcar" cells with solid metal doors that closed off visual contact and muffled sound because he thought they were "necessary for the protection of staff and other inmates" (p. 1104). He also observed in written testimony that prisoners in isolation who appeared to be seriously mentally ill were likely not as sick as other experts indicated; he speculated that they might be malingering. Although Stoner told the court in *Jones 'El v. Berge* that the isolated housing conditions at the prison were entirely appropriate, the judge disagreed. She held that the Wisconsin facility was unconstitutionally harsh for mentally ill prisoners and ordered them removed.

In any event, the Colorado researchers started out with a seemingly good idea and what appeared to be a reasonable research design. They would identify groups of prisoners housed in administrative segregation

(AS) and in the general population (GP), subdivided into those suffering from serious mental illness (MI) and not (NMI). Their psychological status would be tracked for 1 year to determine whether and how the different groups were affected by different conditions of confinement.⁷ The characteristics of the AS and GP prisoners were not matched at the outset but were expected to be more or less comparable because all had committed rules violations for which they might have received an AS placement.

Assignments to AS were thus not random. The researchers reported that “placement into AS or GP conditions occurred as a function of routine prison operations, pending the outcome of their AS hearing, without involvement of the researchers. . . . Inmates who returned to GP following an AS hearing were assumed to be as similar as possible to AS inmates and, therefore, comprised the comparison groups” (O’Keefe et al. 2010, p. 17). The prisoners whom prison authorities chose to send to administrative segregation became the treatment group and those returned to the general population became the comparison group (again, with each group subdivided into those identified by the prison system as mentally ill and those not).

Unfortunately, the plan fell apart almost immediately. The prison context and “routine prison operations” fundamentally undermined the research design.

A. Contamination of Treatment and Comparison Groups

The study’s implementation was compromised in two fundamental ways. It is important at this juncture to acknowledge the distinction between mere methodological “limitations”—respects in which a study is not perfect—and problems that are so fundamental that they make the resulting data uninterpretable. The two flaws from which the Colorado study suffered were fatal—separately and in combination.

1. *All Participants Were Exposed to the Treatment.* All participants in the study, including those in the comparison group, were initially placed

⁷ Data for one group of participants—prisoners “with the most acute psychiatric symptoms” housed at a psychiatric treatment facility where they lived and interacted with one another “on their living unit” (O’Keefe et al. 2010, pp. 14–15)—did not bear directly on the issue of whether and how much prisoners were affected by AS. The researchers included them separately “to study inmates with serious mental illness and behavioral problems who were managed in a psychiatric prison setting” (p. 17). The prisoners in this group were not living in conditions remotely comparable to prisoners housed in conventional GP or AS units.

in “punitive segregation,” a severe form of solitary confinement, for unspecified but not insignificant periods, before being assigned to administrative segregation or the general population. “At the time leading up to and during their AS hearing,” the researchers acknowledged, “inmates have typically been in segregation” (O’Keefe et al. 2010, p. 8).⁸ The reason was that Colorado prison officials were required to hold hearings to determine whether prisoners were guilty of infractions and if so whether AS punishment was warranted. Prisoners in Colorado as elsewhere are placed in special housing while they await the outcomes of their disciplinary hearings, often for days or weeks before the process is complete. Thus, the researchers also noted that “offenders reclassified to AS *remain* in a punitive segregation bed until an AS bed becomes available” (O’Keefe et al. 2013, p. 50; emphasis added).

Although this is routine correctional practice, its methodological implications were disastrous. It meant that all members of the comparison group were exposed to a severe dose of the isolation “treatment” before the study began. O’Keefe et al. (2010, p. 9) indicated that the punitive segregation conditions where prisoners were kept while disciplinary proceedings unfolded were so harsh that they were “only intended to be used for a short period of time.” This severity distinguished it from AS, which was intended to be used for much longer periods. Here is how they described punitive segregation:

Punitive segregation offenders remain in their cell for 23 to 24 hours a day, only coming out for recreation and showers, both of which are located in the living unit. Therefore, most do not leave the unit during their segregation time. Services including meals, library, laundry, and even medical and mental health appointments occur at the cell door. If a situation warrants an offender to be out of cell, the offender is placed in full restraints and escorted to a room within the unit

⁸ Why “typically” is unclear. The report indicates that all prisoners (including the GP comparison groups) were placed in some form of isolation before, during, and shortly after their AS hearings. It is hard to imagine a procedure in which a prisoner would be taken directly out of GP, immediately given an AS hearing, and immediately returned to GP, without having spent time in some form of isolated housing. In fact, the authors reported that AS participants “on average completed their initial test 7 days (SD = 7.3) after their AS hearing,” that GP participants on average “were tested 16 days (SD = 18.9) after their hearing,” and that “on average, 43 percent of inmates . . . [had] been confined in segregation (40 percent in AS groups and 3 percent in GP groups) for an average of 18.2 days (SD = 18.1)” (p. 30). These figures are mathematically impossible. Moreover, they are at odds with O’Keefe’s deposition testimony and with a statement in a more recent published “reflection” on the study (O’Keefe 2017).

where he or she can meet privately. Many offenders do not like being taken out of their cells because of the use of full restraints. Additionally, they may not like leaving their cell because officers may take the opportunity to search the cell for contraband.

Due to the disciplinary nature of punitive segregation, offenders are stripped of most privileges during their stay. Punitive segregation inmates are neither allowed to work nor permitted to participate in programs or education. Furthermore, their televisions are removed, and they cannot order canteen beyond essential hygiene items. (O'Keefe et al. 2010, p. 8)

Punitive segregation prisoners were denied visits, which were considered too labor intensive for prison staff to administer.

In contrast to AS, prisoners in punitive segregation also were denied the opportunity to engage in programming or education and were "unable to begin working their way toward leaving segregation" (O'Keefe et al. 2010, p. 9). Thus, even study participants who wound up in AS likely experienced punitive segregation as a much worse form of treatment.

This initial exposure of all participants to an especially harsh form of solitary confinement in punitive segregation made it impossible to draw meaningful inferences about any separate, subsequent effects of GP versus AS. There can be no comparison group in a study in which all of its participants are subjected to a harsh form of the treatment whose effects are being measured.

It is impossible to know whether or how control group prisoners were damaged by the time spent in punitive segregation and whether those effects continued throughout the study. Nor could anyone know whether the AS prisoners were actually relieved to enter the "treatment" because it was less harsh than punitive segregation. These imponderables could account for participants' psychological reactions, including the reported lack of differences between the AS and GP groups and the reported "improvement" or lack of deterioration of many members of the AS group. This was thus no longer a study of administrative segregation compared with no administrative segregation, but of varying and unspecified amounts of segregation experienced by everyone.

A different kind of analysis might have salvaged something by using the exact periods of overall exposure to administrative segregation-like conditions (including time in punitive segregation) as a continuous variable to estimate whether duration had an effect. However, the amount of time in segregation each prisoner experienced is not reported, so this

kind of analysis was apparently not conducted. O'Keefe et al. (2010) treated their data as if they had done a classic treatment versus no treatment study, even though they had not.

The likelihood that initial exposure to punitive segregation conditions had significant negative psychological effects on most participants is more than just speculation. The National Institute of Justice report acknowledged that three of the four groups "showed symptoms that were associated with the SHU syndrome" from the outset (O'Keefe et al. 2010, p. viii), which seems a clear indication that the initial period of segregation adversely affected participants before their AS terms began. High levels of psychological distress measured during or after the prisoners' initial exposure to punitive segregation continued throughout the study. O'Keefe emphasized in a deposition that prisoners in all groups reported "pretty high elevations" of psychological distress (2010, p. 171) and that "clearly, very clearly, the offenders responded with very high elevations. They reported high levels of psychological distress" (p. 201).

Symptoms of distress were so elevated that the researchers wondered, and tried to test, whether the prisoners were malingering: "We had this huge rate of offenders who looked like they could be malingering" (O'Keefe 2013, p. 89). O'Keefe recognized, however, that high scores on a malingering scale "could indicate a lot of psychological problems." In the end, the researchers "didn't really believe that [the prisoners] were malingering" and discarded the results of the malingering scale without analyzing them (p. 89).

Thus, although the researchers acknowledged that most of the participants began the study very much affected by emotional and behavioral trauma, they seem not to have considered that much of that trauma resulted from time spent in the punitive segregation units. Nor did they consider that, when participants "naturally got better as time went on" (O'Keefe 2013, p. 91), it was likely because the conditions of punitive segregation that all of them had experienced were now alleviated, even for those who ended up in AS.

The amount of time that the study participants spent in punitive segregation was problematic, especially because even very brief periods of isolation can have damaging psychological effects. The United Nations Special Rapporteur on Torture, Juan Mendez, has noted that "it is clear short-term solitary confinement can amount to torture or cruel, inhuman, or degrading treatment" and recommended that solitary confinement "in excess of 15 days should be subject to an absolute prohibition"

(2011, p. 23). The United Nations adopted that recommendation in the “Mandela Rules,” which defined “prolonged solitary confinement” as lasting “for a time period in excess of 15 consecutive days,” and mandated prohibition of such prolonged confinement (Commission on Crime Prevention and Criminal Justice 2015, rules 43.1, 44). The National Commission on Correctional Health Care (2016) also characterized “prolonged solitary confinement” lasting for more than 15 days as “cruel, inhumane, and degrading treatment” because it is “harmful to an individual’s health” (p. 260). Yet all of the prisoners in GP and AS experienced a nontrivial duration or dose of isolation that lasted well beyond this potentially damaging threshold. A key table in the National Institute of Justice report indicated that, at the time of their first test interval, participants had spent considerable average times in “Other seg”: GP MI prisoners 12.4 days, GP NMI 39.8 days, AS MI 88.9 days, and AS NMI 90.3 days (O’Keefe et al. 2010, table 5).

In her deposition testimony, O’Keefe could not remember exactly how long study participants remained in punitive segregation before their charged disciplinary infractions were resolved. At one point, she said, “When an offender acted out, they were put in punitive seg and generally given notice of a hearing pretty quickly, and then the hearing happened, again pretty quickly after that” (2013, p. 93). Later she “guessed” the time was around “the two week mark” (p. 94). That was not remotely accurate, according to table 5 in the report, except for the GP MI group. O’Keefe later offered another estimate, this time that prisoners were kept in various punitive segregation units “an average of 30 days” before their initial testing session (2017, p. 2). This, too, is much less time than the National Institute of Justice report showed. In any event, it appears that all study participants were subjected at the outset to harsh conditions of punitive segregation for at least twice as long as the Mandela Rules would prohibit, even before the study officially began.

2. *Uncontrolled Cross Contamination.* The second fundamental flaw was as important as the first. It, too, occurred because placement and retention in AS were correctional rather than methodological decisions. The researchers admitted that they “lack[ed] control over the independent variable, which in this case is the conditions of confinement” (O’Keefe et al. 2010, p. 35). There was, in their words, “contamination across groups,” because some AS participants “were not confined in segregation for their entire period of participation in the study” and because some GP participants “may have at some time during their study partic-

ipation been placed in punitive segregation or even AS” (p. 35). The researchers also acknowledged that prisoners in the various subgroups “may have [been in] multiple locations within a study period” (p. 35).⁹ In fact, not only did participants move between AS and GP, but a number of them were housed in other conditions during the study, including the hospital and “community placement” (p. 36).

Transferring prisoners back and forth between locations and custody statuses is routine correctional practice, but it had disastrous methodological consequences. It meant that some AS prisoners in the study were released into GP for good behavior, some GP prisoners were placed in AS (or punitive segregation) for rule violations, and some members of both groups were transferred to other settings. Having both control and experimental group members move back and forth between treatment and control conditions (and other unspecified places) destroyed the integrity of the two groups and made it impossible to compare their experiences meaningfully.

The contamination occurred differently between groups. By the end of the study, only small and very different numbers of “uncontaminated” participants were left in each group.¹⁰ Methodologically speaking, a true, a natural, or even a quasi experiment cannot be completed if researchers lose control of the integrity of their treatment and comparison groups. The researchers, however, simply aggregated the contaminated prisoners’ data into the groups in which they were originally placed.

O’Keefe et al. (2010, p. 35) acknowledged that “one of the challenges of applied research is the researchers’ lack of control over the independent variables,” but that admission does not ameliorate the problem. They

⁹ They wrote that “participants remained in their assigned group regardless of their placements throughout the prison system” (O’Keefe et al. 2010, p. 35), but mean by this that individual prisoners were considered to be in those groups for purposes of data analyses even though they did not actually remain housed there.

¹⁰ There were only 26 “pure” cases in the AS MI group (of the original 64), 39 in AS NMI (of 63), 13 in GP MI (of 33), and only 11 in GP MI (of 43) (O’Keefe et al. 2010, p. 35). All the others moved back and forth between treatment, control, and miscellaneous other conditions on an unspecified number of occasions. Thus two-thirds (52 of 76) of the GP control participants spent time in segregation or other non-GP settings during the study period, and their self-reports were used to contrast their prison experiences and reactions with those of the AS prisoners, half of whom (62 of 127) spent unspecified amounts of time in GP or elsewhere. The “pure” cases were pure only in the sense that they were not contaminated by moving back and forth between treatment, control, and other conditions during the study. They were still “contaminated” by being exposed to punitive segregation before the study officially began.

nonetheless asserted that “a significant advantage of this study is the use of comparison groups to determine if [persons in AS] change over time differentially compared to similar groups who are not placed in AS” (p. 59). However, they did not compare similar groups and thus can reach no conclusions about differences in the groups’ experiences.

In fact, it is impossible to conclude anything meaningful from the Colorado results. Lovell and Toch (2011, p. 4) in their initial commentary on it correctly concluded that “despite the volume of the data, no systematic interpretation of the findings is possible.”

B. Additional Serious Flaws

The researchers’ inability to maintain control of key aspects of their research created numerous additional methodological problems. These problems further negated the possibility that any credible or meaningful findings would emerge from the study.

The additional problems pertained to how the participants were selected and how the various groups were composed, what the researchers recorded (or failed to record) about the experiences of members of the different groups, and questionable data collection procedures. Most stemmed from unyielding correctional realities and some from unwise methodological choices.

1. *Sampling and Group Composition.* The initial sample was drawn from among prisoners deemed eligible for the study by virtue of having received a disciplinary write-up and scheduled hearing to determine whether they would be placed in AS or returned to GP. The initial group of eligible prisoners was much larger than the number selected to participate. The decision about whom to approach was made single-handedly and, as she would characterize it, “haphazardly” by O’Keefe: “I would determine who we used, who we included in our study” (2010, p. 116).

The major consideration for inclusion was proximity to the field researcher: “We had one researcher, so we had to be able to manage her workload” (O’Keefe 2010, p. 116). She described the process as “haphazard selection. . . . We didn’t do it in a random fashion, but we didn’t necessarily do it in a very targeted fashion either” (p. 116). Participants were drawn from only 10 of Colorado’s 26 men’s GP prisons (O’Keefe et al. 2013, p. 51). A disproportionate number came from Limon Correctional Facility “[because] it’s fairly close” (O’Keefe 2013, p. 66). This was not mentioned in either the National Institute of Justice report or

the briefer published version of the study. If there was anything significantly different about that prison, for example, if its punitive segregation unit (where participants were housed before the study began) was especially harsh or its GP units (to which many participants were returned) were particularly dangerous, troubled, or inhumane, then a disproportionate number of prisoners would have been affected by being held there.¹¹ There is no way to tell.

There was also unexplained and unnecessary imprecision in the composition of the groups. In addition to being composed of persons subjected to punitive segregation immediately before they entered GP, the GP group began as an amalgam of prisoners who subsequently lived under different conditions of confinement. Thus, “thirteen participants in the GP groups were selected from the diversion program (for being at risk of AS placement)” (O’Keefe 2010, p. 30). The report elsewhere implied that all of the prisoners were at risk of AS placement because all had AS hearings; apparently that was not true, and some were “diverted” out of the process entirely.

A potentially more serious problem concerned the composition of the AS group. O’Keefe et al. (2010, p. 8) asserted that “Colorado does not house protective custody; therefore, no AS placements occur at the request of inmates.” This is a correctional non sequitur. Colorado may not officially house protective custody inmates, but they exist in every American prison system. Protective custody inmates often end up housed in AS, whether or not they formally request it. In the Colorado study, an unusually large group of AS participants were identified as having sex offender needs: 30 percent of the AS NMI prisoners and 44 percent in the full AS group (p. 45). In other prison systems, many, possibly all, such prisoners would be protective custody cases. To be sure, protective custody prisoners are subject to the painful and potentially harmful effects of social and sensory deprivation. However, they are in a very different situation psychologically than prisoners placed in AS for punishment. Protective custody prisoners typically prefer to be housed in AS-type conditions instead of what they regard as more dangerous GP environments. As a result, they are likely to be reluctant to voice complaints about living

¹¹ O’Keefe understood the implications of the sampling methods. Concerning work by others on the effects of administrative segregation, she wrote, “Of particular concern is that sampling procedures are often not discussed, and thus it is impossible to know if the findings were based on a representative sample” (2008, p. 127).

conditions or adverse emotional reactions, lest they be moved. That a third of the AS NMI prisoners and nearly half of the AS group overall in the Colorado study were probably protective custody cases undermined any straightforward interpretation of the data.

Gang members presented a similar problem. Thirty percent of AS MI prisoners and 43 percent of those in the AS NMI group were identified as gang members (O’Keefe et al. 2010, table 9). Being a gang member would ordinarily reduce a prisoner’s willingness to report psychological distress because that would be a sign of vulnerability that might be interpreted as weakness.

Thus, nearly three-quarters of both the mentally ill and non-mentally ill AS prisoners were likely protective custody cases or gang members. Yet the researchers ignored the implications of this entirely.

2. *Uncontrolled Differences in GP Conditions.* The control condition—GP—referred to placement in one of 10 different prisons. However, none of the specific conditions of confinement at any of those prisons is described.¹² Variations in GP environments matter because, obviously, unless all GP prisoners experienced the same environment, they were not really in the same condition. If some of the GP environments were so troubled, dangerous, and harsh that they approximated or were worse than conditions in AS, it would be impossible to make meaningful comparisons.

A disproportionate number of study participants were housed in the Limon Correctional Facility (O’Keefe 2013, p. 66). This appears to have been an especially troubled prison when the study was conducted. In 2010, a journalist wrote about “Limon’s long history of inmate violence, including two fatal stabbings in five years and the beating death of a correctional officer” (Mitchell 2010).¹³ The prison’s 5-year violent history encompassed the entire period of the Colorado study from July 2007 through March 2010 (O’Keefe et al. 2010, p. vii). This meant that many study participants came from (and GP comparison group prisoners remained in) an especially harsh and dangerous GP environment, perhaps one as psychologically stressful as an AS unit. In fact, Limon’s vi-

¹² The published article indicated only that “GP inmates have access to significant out-of-cell time (e.g., >10 hours/day), jobs, and programming” (O’Keefe et al. 2013, p. 51). No additional information about the GP environments was provided.

¹³ There were also allegations that in 2008 sex offenders at the prison were targeted by gang members who extorted them to pay “rent” and repeatedly threatened and assaulted them (*Davis v. Zavaras*, 2010 WL 625043 [D. Colorado 2010]).

olent history may have been serious enough to have precipitated recurring violence-related lockdowns (e.g., Associated Press 2007), including in the GP units where some of the control inmates were housed. None of this was commented on or taken into account.

3. *Uncontrolled Differences in AS Conditions.* Colorado study AS participants were ostensibly in the same study condition but were nonetheless exposed to very different conditions of confinement. These differences were not recorded or quantified and thus could not be taken into account. First, as I noted, all study participants experienced varying amounts of a harsh form of prison isolation, punitive segregation, before the study began. For a significant number (apparently, the majority) of the AS prisoners, that continued for a quarter or more of the length of the study. Thus, “When the study began, there was a 3-month average wait for inmates to be transferred to [AS],” which was “due to a shortage of beds. While on the waitlist, AS inmates were held in a punitive segregation bed at their originating facility” (O’Keefe et al. 2010, p. 19).

The median stay in punitive segregation for AS participants was reported as 99 days (which means that half were longer), although a very small group of prisoners were moved “quickly” into AS. Despite these very different periods in prestudy punitive isolation, all AS participants were lumped together for purposes of analysis.¹⁴

There was additional imprecision about how much and what kind of isolation any one AS participant experienced. Some “were not confined in segregation for their entire period of participation in the study” but were released into GP or other less onerous settings (O’Keefe et al. 2010, p. 19).

However, even beyond this, it is impossible to know exactly what conditions of confinement were experienced by participants who remained in AS throughout the study. The reason is that Colorado’s AS program operated a “level” system in which a prisoner’s “quality of life” (QOL) varied as a function of behavioral compliance and programming. Changes in QOL were meant to be incentives for compliance with unit rules and eventual reassignment to GP. The average length of AS stay was said to be 2 years, with the expectation that prisoners would spend at least 1 year in AS. However, the minimum stays specified for the QOL program

¹⁴ The “distance between when they were ad-seged and when they went to CSP became longer and longer because of the wait list in DOC” (O’Keefe 2010, p. 108). An unspecified but not insignificant number of administrative segregation prisoners “were held in the punitive segregation bed but classified as ad-seg. And that’s the—for the study average to be about 90 days, but people could be there pretty short, pretty long” (p. 109).

envisioned much shorter stays: 7 days at level I, 90 at level II, and 90 at level III—187 days altogether—after which prisoners were eligible for consideration for reassignment back to GP (O’Keefe et al. 2010, p. 11).

Providing achievable incentives for good behavior and early release from AS are sensible correctional practices. However, they, too, further compromised any meaningful interpretation of the study results.

This methodological problem was significant because the differences in QOL at different levels of AS were substantial. The researchers acknowledged that “it was expected that [prisoners in AS] might experience varying amounts of isolation based on the amount of time spent at different [QOL] levels” (O’Keefe et al. 2010, p. 40). But these varying amounts of isolation were not documented or taken into account.

O’Keefe acknowledged that the researchers initially wanted information from prison staff on participants’ out-of-cell time, “to track every time they left their cell,” but could not obtain it because the data “just were not coded consistently or every time” by correctional officers (2013, p. 55). That meant that the researchers were unable to track the basic facts of whether, when, and for how long any one prisoner was at one or another AS level or incorporate these data into their analysis (p. 60). O’Keefe et al. (2010, pp. 40–41) reported that staff records yielded “conflicting information,” and “it was often difficult to decipher and/or interpret the records.” Thus, “it was not possible to code or use [them] in the study.”

4. *Failure to Control or Record Treatment Dose.* There was more to these uncontrolled and unrecorded variations than just minor differences in the amount or duration of isolation. The variations in isolation in the AS condition—including for the relatively few prisoners who stayed in AS continuously—were very significant. The QOL level III AS prisoners were given additional privileges and allowed to have jobs as orderlies or in the barbershop. This permitted significant out-of-cell time, during which the prisoners were presumably unrestrained and in contact with others.¹⁵ These opportunities are rare in prison AS units anywhere and

¹⁵ As O’Keefe et al. (2010, p. 12) noted, “Arguably one of the most important benefits of QOL level three is an offender’s ability to have more contact with friends and family. While offenders’ visits remain noncontact, they are increased to four 3-hour visits per month and four 20-minute phone sessions. . . . One additional benefit is that offenders may now be eligible to work as a porter or barber. . . . Benefits to being offered a job position include the ability to earn money, increased time out of cell, and two additional phone sessions per month.”

constitute a significant modification in the nature of the isolation experienced by an unspecified number of AS prisoners. They introduced even more heterogeneity into the “same” condition in the study than already existed.

The researchers also noted that an AS prisoner who acted out could be even more significantly locked down by being placed “on special controls in the intake unit where he can be carefully monitored” and “additional sanctions may be imposed through the disciplinary process” (O’Keefe et al. 2010, p. 13).

None of these and other variations in actual day-to-day conditions of confinement were taken into account. The researchers also did not record and were unable to estimate other basic, important variations in the experiences and treatment of the study participants. These included the number of social or family visits prisoners had, visits from attorneys (O’Keefe 2010, p. 164), and the nature or amount of mental health services the prisoners (including those who were mentally ill) received. As O’Keefe summarized, “We did not look at any facet of segregation or correctional conditions that might affect the outcome of the study. We merely looked at, based on their conditions of confinement—that is, whether they had originally been coded ‘AS’ or ‘GP’—and then noted ‘if they reported worse change over time’” (p. 207). But whether a prisoner had originally been coded AS or GP did not indicate what “conditions of confinement” he had experienced in the course of the study.

C. Miscellaneous Data Collection Problems and Issues

In addition, there were very serious problems with how the Colorado researchers initially structured and eventually implemented the data collection process as well as with the dependent measures they used. Some of these problems were the product of the challenging nature of the prison environment. Others were not.

1. *A Single, Inexperienced Field Researcher.* Almost all the data collection was done by one inexperienced research assistant who had only a bachelor’s degree, no graduate training, and no prior experience working with prisoners or in a prison setting. She was single-handedly responsible for conducting five to six separate testing sessions in which she administered between 10 and 12 separate tests with each of 247 participants in 10 different prisons.

The data collection was unusually challenging. O’Keefe noted, “Say when she was at CSP [the AS facility], she might have a whole bunch

of [participants] and she would go back and forth checking to make sure that they were all right, and administering the questionnaires when she needed to” (2010, p. 118). Yet no one oversaw her day-to-day work (p. 130). O’Keefe had no recollection of ever observing her administering the tests and indicated Klebe did not (2013, p. 85).

2. *Solicitation and Consent.* When prisoners’ participation and consent were solicited, they were told, somewhat misleadingly, that “we’re looking at how inmates across the entire DOC are adjusting to prison life” (O’Keefe 2010, p. 199). O’Keefe characterized this as “being cautious without being dishonest” (p. 200). The consent form told prisoners that the “risks of this study to you are very small in contrast with the benefits that are high. This study will help us to figure out what types of men adjust better to prison and how to help those who are struggling with prison life” (O’Keefe 2013, pp. 81–82). This, too, was misleading. The study was not about the types of men who adjust better to prison and how to help them. Moreover, no consideration was apparently given to the possibility that prisoners might want to appear to be “adjusting” rather than “struggling.” This would apply with special force to AS prisoners, hoping to advance their QOL level and with that gain additional privileges and earlier release from the unit.

3. *Prison Employee?* The field researcher had to complete “the full CDOC [Colorado Department of Corrections] training academy” and at all times was required “to wear a visible CDOC badge that permitted her unescorted access to the facilities” (O’Keefe et al. 2010, p. 28). Although O’Keefe was “not sure” how the field researcher introduced herself to prisoners, she conceded that “it could be” that prisoners thought the field researcher was a DOC employee (2010, p. 125).

Prisoners in general, and especially in AS units, are typically reluctant to confide in prison staff (including even mental health staff) because of potential adverse consequences. Those consequences can include increased surveillance, placement in degrading “suicide watch” cells, or transfer to or retention in some other form of AS. For these reasons, prisoners frequently avoid admitting that they feel suicidal, depressed, frightened, angry, panicky, out of control, or violent.

That prisoners could reasonably infer that the field researcher/prison employee was checking on their “adjustment” is likely to have dampened their willingness to disclose sensitive feelings. This possibility is nowhere discussed. Despite the fact that while the study was under way, O’Keefe acknowledged awareness of the fraught nature of prisoner-staff

relations, especially in AS units: “Administrative segregation facilities are characterized by the complete control exerted over inmates by correctional staff. The typical ‘we-they’ dynamic between inmates and staff is exacerbated in segregated settings where inmates have almost no control over their environment. Prisoner abuses have been discovered and punished in administrative segregation settings, but in other situations Human Rights Watch found that ‘management has tacitly condoned the abuse by failing to investigate and hold accountable those who engage in it’” (2008, p. 126; internal citations omitted).

4. *Undermining Trust.* Little was done to overcome what O’Keefe described as the “we-they” dynamic that she believed was likely to be exacerbated in prison AS units. Two related problems with the Colorado study likely exacerbated the effects of this dynamic. The first was an error of omission: no interviews were conducted to establish rapport with prisoners. O’Keefe indicated that “it was not part of the study to probe and ask them [the prisoners] about themselves” (2013, p. 75). Without rapport-building interactions, prisoners in the study were unlikely to have had much confidence that the field researcher was interested in their well-being or that personal revelations would be handled with sensitivity.

The second problem is more troubling. The field researcher was apparently required (or decided on her own) to challenge prisoners if she thought their answers were “questionable” or “untruthful, or if she found the pattern of their responses abnormal” (O’Keefe et al. 2010, p. 36). There was no explicit or systematic protocol by which this judgment was reached (none is described). In any event, the field researcher reviewed the prisoners’ responses on the spot, in their presence, every time they completed a questionnaire. If she was skeptical, the prisoner was asked to redo the test. Prisoners could decide to redo the test or not, but “if the participant said he was being honest and the researcher still did not believe him, she marked the test as questionable” (p. 36).

These practices potentially created very significant data quality problems. They not only jeopardized the development of rapport or trust but also increased the chances that prisoners would give situationally desirable answers. In addition, the problems likely extended to more prisoners than only those who were challenged directly, but to other prisoners who learned through word of mouth that they would be asked to redo their questionnaires if the researcher was skeptical of their answers.

5. *“Untruthful” and Other Questionable Data.* Twelve percent of participants “had a questionable response pattern on any measure at any

time period” (O’Keefe et al. 2010, p. 36). It is unclear whether that figure included all participants who were asked about their answers or only those whose answers were marked “questionable.” If challenged prisoners admitted being untruthful and redid the questionnaire, the second versions of their answers were incorporated into the study data. However, even if the field researcher was skeptical and prisoners chose not to redo their questionnaires, “we still included that in the study. . . . In order to increase our statistical power . . . we left those cases in” (O’Keefe 2010, p. 166).

In addition, 23 participants withdrew their consent and dropped out before the study was completed. However, their data were retained and used in the overall analyses (O’Keefe et al. 2010, p. 19). The dropouts constituted nearly 10 percent of the 247 participants. This meant that, in total, more than 20 percent of the participants whose data were included in the study results were adjudged to have given untruthful responses or withdrew from the study.

6. *An AS “Heisenberg Effect”?* The repeated testing procedure changed the conditions of confinement, especially for AS prisoners otherwise subject to extreme social deprivation. The six interactions of approximately an hour each between the field researcher and the prisoners, no matter how strained or superficial they might have been, increased the otherwise minimal social contact that AS prisoners had with people outside the segregated housing unit.¹⁶ In many prison systems, there are many AS prisoners who get no visits at all. The mere act of repeatedly attempting to measure the effects of severe conditions of isolated confinement can change them, if only slightly, for the better.

7. *Miscellaneous Issues.* There were other irregular, questionable, and unexplained research decisions and data anomalies. Exactly why prisoners were assigned to AS or GP was not indicated, even though this was how the treatment and control groups were created. Assignment to AS was apparently nearly automatic: no more than “approximately 10 percent of hearings do not result in AS placement” (O’Keefe et al. 2010, p. 17). This raised questions, never addressed, about what accounted for the unusual outcome in the case of the group that was returned to GP.

¹⁶ It apparently exceeded the contact AS MI prisoners had with mental health staff: “Offenders with mental illness who are stable are offered a one-on-one session at least once every 90 days,” which takes place “in a noncontact booth in the visiting room” (O’Keefe et al. 2010, p. 11).

Nor were reasons discussed for why the NMI prisoners who returned to GP had more disciplinary infractions (average 16 each) than those sent to AS (13.2 average). Nor were reasons discussed for why AS MI prisoners had 70 percent more disciplinary infractions on average than the AS NMI inmates (22 infractions compared with 13.2; O’Keefe et al. 2010, table 9). Nor was there discussion of the effects of exclusion of prisoners from the study who did not read English at an eighth-grade level on the representativeness of the final group of participants, especially with respect to ethnicity and the prevalence of cognitive impairments.

D. Troubling Dependent Measures

There were also serious problems in the handling of dependent variables in the study. Dependent measures were said to have been selected on the basis of several important criteria. However, the first two criteria the researchers identified—“(1) use of assessments with demonstrated reliability and validity, (2) use of multiple sources for providing information (e.g., self-report, clinician ratings, files)” (O’Keefe et al. 2010, p. 19)—did not apply to the dependent measures that were actually used in the analyses.

1. *Unvalidated Scales and Instruments.* Some of the study’s scientific bona fides were based on its claimed use of validated and objective assessment instruments. The researchers asserted that “the use of a reliable and valid standardized measure in the present study enabled objective assessment of psychological functioning” (O’Keefe et al. 2013, p. 57).

Indeed, O’Keefe acknowledged that “inaccurate judgments” could be made if instruments were not properly validated (2010, p. 22). However, she later conceded that only “a very low number” of the numerous scales and measures used, perhaps no more than one or two, had been normed or validated with a prisoner population (pp. 144–45).¹⁷

¹⁷ There was no evidence that even the Brief Symptom Index (BSI), on which the researchers relied exclusively in the published version of the study, O’Keefe et al. (2013), had ever been validated with a prisoner as opposed to a “forensic” population. One study that the authors cited to support its psychometric properties (Kellet et al. 2003) concerned the BSI’s reliability with persons suffering from intellectual disabilities and did not include a representative sample of prisoners (the “forensic” portion of the sample consisted of 45 “intellectually disabled” convicted persons who were “detained in a maximum security hospital” [p. 129]). The second, Boulet and Boss (1991), was a study of “psychiatric inpatients and outpatients who presented for evaluation at the forensic service of a psychiatric hospital” (p. 434). The third, Zinger, Wichmann, and Andrews (2001), focused on prisoners but did not report reliability or validity data for the BSI.

2. *“Constructs” That Could Not Be Interpreted or Compared.* The near-exclusive reliance on prisoners’ self-report assessments was problematic because the researchers chose to separate the various scales into their component parts and then recombine items into eight separate “constructs.” Instead of reporting scores on the instruments or scales themselves, only the constructs built from them were presented as standardized composite rather than numerical scores (O’Keefe et al. 2010, p. 22). This meant that the significance of reported overall trends and comparisons between groups was, as Lovell and Toch (2011, p. 4) put it, “difficult to assess because of the degree to which the data have been cooked.”

There are a number of unanswered questions concerning construction of composite scales including their basic validity (whether the instruments measured what they purported to measure), whether the various subscales were reliable for this population, and whether the distributions of scores lent themselves to the statistical manipulations and recombinations that occurred. Transformations to the data, the number of instruments, items, and constructs, and the amount of scale and subscale reconstruction that occurred make the results difficult to put in the context of any larger literature using the same self-reported assessments.

3. *Ignoring Behavioral Data.* Researchers who use many rating scales (especially ones not validated for the particular population) generally use other methods of data collection as a validity check. The most basic is a face-to-face interview to establish rapport and acquire background information. When possible, behavioral data (by records reviews or behavioral rating scales completed by others) are included. These different sources of information should be reconcilable, and the interviews provide the glue that binds them. Prison researchers typically take things prisoners say to them very seriously, in part because they contextualize other things being measured or studied. However, no interviews were conducted in the Colorado study, and little or no special effort appears to have been expended to establish rapport. Instead, the researchers engaged in context-free coding and analysis of answers on prepackaged forms associated with tests not typically used with this population. As Lovell and Toch (2011, p. 3) observed, “Readers find themselves swimming in a flood of psychometric data; every so often a clue drifts by, lacking, however, a tether to the context—to what was going on around the prisoners and staff while they carried out this study—we are left to guess what it might mean.”

Other kinds of data collection were contemplated including asking corrections officers and clinicians to complete rating scales: “The Brief Psychiatric Rating Scale was completed by clinical staff and the Prison Behavior Rating Scale was completed by correctional officers and case managers” (O’Keefe et al. 2010, p. 26). However, key details about this process were omitted (i.e., exactly who was supposed to complete scales, when, and with what kind of training). In the end, it did not matter. The rating scales were infrequently completed and the responses were too unreliable to be useful. The data were discarded. The researchers ultimately relied only on data from prepackaged, field researcher-administered rating scales.

There was one potential exception. Prison mental health staff kept official accounts of genuine psychiatric emergencies or “crisis events.” Any situation that required “immediate psychological intervention is considered a crisis event; crisis events are documented by clinicians” (O’Keefe et al. 2010, p. 42). Because these are typically extreme, clinically significant events, they tend to be reliably recorded. If the prisoners’ self-reporting was valid, the results should be more or less consistent with behavioral measures of psychological distress or crisis. In the Colorado study, they were not. Among the 33 GP MI prisoners for whom data were reported, there were only three “crisis events” (on average, one for every 11 inmates). Among the 64 AS MI prisoners, there were 37 “crisis events” (one for every two; O’Keefe et al. 2010, figs. 29, 30). This suggests that at least some mentally ill prisoners were doing much worse in AS than their counterparts were doing in GP.

The researchers dismissed the implications of this incongruity: “Because the number of participants who experienced a crisis event was so small, it was not possible to include this variable as an outcome measure in the change over time analyses” (O’Keefe et al. 2010, p. 42). Thus the significant disparity between self-reports and the behavioral measures was ignored, even though it directly contradicted the study’s main finding that AS did not adversely affect the mental health of mentally ill participants. Instead, as they put it, because the mental health crisis data “raise more questions than they provide answers,” they were deemed “outside the scope of the current research” (p. 42).

In sum, for all of the above stated reasons, the Colorado study is so methodologically flawed that literally no meaningful conclusions can be drawn from it. Drastic compromises necessitated by the complex realities of the prison setting and a series of questionable methodological decisions made

by the researchers rendered its results uninterpretable. The Colorado study was not the “most sophisticated” study done to date on the psychological effects of solitary confinement. Its results do not “need to be taken seriously,” but cannot be taken for anything at all. Commentators who have praised the study either did not read it very carefully, were unaware of available sources of information on how it was actually conducted, or did not seriously consider the implications of its fundamental flaws.

Ordinarily, a study of this sort would die a quiet death, notwithstanding an occasional prison system’s attempt to resuscitate it to defend questionable segregation practices or a scholar overlooking its flaws because its findings comport with his or her own views. However, it has recently been given a second life, figuring prominently in a recently published meta-analysis (Morgan et al. 2016). Its results threaten to live on in another form and to misrepresent the findings of the large, long-established, and frequently reconfirmed literature on the harmful effects of solitary confinement.

III. The Limits and Dangers of Meta-Analysis

Meta-analysis—“a quantitative method of synthesizing empirical research results in the form of effect sizes” (Card 2012, p. 7)—is an important methodological advance that allows researchers to estimate the overall magnitude of relationships between variables. However, it cannot substitute for careful narrative reviews of scientific literature. Meta-analysis comes with substantial limitations, especially for prison research. The prison setting rarely lends itself to collection of meaningful quantitative data capable of generating the kinds of effect sizes on which meta-analyses depend. Most classic book-length treatments of prison life have been primarily ethnographic—not quantitative at all. They contain few if any numerical data, including in the seminal American works by Cressey (1940), Sykes (1958), Toch (1975, 1977), Jacobs (1977), and Irwin (1980) and major comparable British works including Cohen and Taylor (1972) and Crewe (2009).

Similarly, few quantitative effect sizes appear in studies of solitary confinement. This is true of the studies that tell us much of what we know about these institutions, how they operate, and the lengths to which prisoners must go in order to survive inside them, including those from Rhodes (2004), Shalev (2009), Reiter (2016), and Kupers (2017). It is also true of most of the numerous studies of the negative psychological con-

sequences of prison isolation that are discussed in the most-often-cited literature reviews. The nature of the settings and the routine prison operations that govern them make many kinds of conventional research designs impossible to implement.

Because the best prison research is qualitative, or does not lend itself to generating effect sizes, meta-analyses conducted on many important prison topics will be compromised by serious sample bias, resulting in “the drawing of inferences that do not generalize to the population of interest (typically all research conducted on the topic)” (Strube, Gardner, and Hartmann 1985, p. 66).

The concern is not only that meta-analyses on important prison topics almost invariably ignore or underrepresent the larger literature, but also that they privilege certain kinds of studies far beyond their actual scientific merit, and do so in a way that many readers are unlikely to appreciate. One critique rightly observed that readers “might not be motivated to look beyond the meta-analyses themselves due to confidence in the objective, straightforward nature of the tasks of conducting a meta-analysis, reporting findings, and making recommendations” (Coyne, Thombs, and Hagedorn 2010, p. 108). Reducing entire studies to single or multiple effect sizes almost invariably creates a false equivalency between them. Readers can easily be mesmerized by arrays of numbers that appear simply and accurately to represent highly complex and substantially different underlying realities.

The two meta-analyses contained in the Morgan et al. (2016) article suffer from all of these problems and more. They need to be scrutinized carefully because of the stakes involved and the possibility that they will mislead correctional decision makers and policy makers by their “surprising results,” ones that, as the authors say, “do not fit with people’s intuitive analysis of what happens when you isolate offenders” in solitary confinement. The resulting conclusions are indeed “in marked contrast to the ‘fiery opinions’ . . . commonly presented in the scientific and advocacy literature” in which solitary confinement “has been likened to torture, with debilitating consequences” (p. 455). They warrant conscientious examination.

A. Truncating the Scope of Literature Reviewed

The first problem with Morgan et al. (2016) is the tiny number and unrepresentative nature of studies included in its two separate meta-

analyses. Literature reviews, whether narrative or meta-analytic, are useful only if they faithfully represent the literature being examined. As Card (2012, p. 10) put it, “If the literature reviewed is not representative of the extant research, then the conclusions drawn will be a biased representation of reality.” Morgan et al. (2016) excluded a vast number of published studies, including most of the key works.

The first meta-analysis, “Research Synthesis I,” reported that over 90 percent of the published material that they found on the topic was eliminated: “Of the 150 studies located, only 14 (or 9.3 percent) were suitable for analysis according to our inclusion criteria” (Morgan et al. 2016, p. 442). The second meta-analysis, “Research Synthesis II,” began with an astonishing 40,589 articles, which were reduced by “trained research assistants” using unspecified methods to 61. A “trained research assistant” then used unspecified methods to reduce that number to 19 (0.05 percent of the initial literature; pp. 442–43).

A meta-analysis that includes so little of the available relevant literature is not a synthesis of much of anything. In addition to the drastic reduction in the sheer number of articles included, the selection criteria used by Morgan et al. (2016) excluded key studies but included questionable other ones. Among the articles excluded is Grassian (1983), regarded as one of the seminal studies on the adverse effects of solitary confinement. Morgan et al. also ignored most of the work discussed in widely cited literature reviews by Haney and Lynch (1997), Haney (2003), Grassian (2006), Smith (2006), and Arrigo and Bullock (2008).

Despite the small numbers of studies included, tables reporting effect sizes seem to suggest that a vast number of studies were taken into account. A closer look reveals something different. Many of the studies have little or nothing to do with the key question of whether and when solitary confinement is psychologically harmful. Morgan et al. (2016) included studies that addressed medical outcomes, and behavioral outcomes such as recidivism and institutional misconduct, that have not been widely studied and are not central to understanding solitary confinement’s psychological effects. Thus, despite the drastic reduction in overall number of studies, many of the studies actually included were simply beside the main point.

When the largely irrelevant studies are set aside, only six studies on the psychological effects of solitary confinement remain in the first meta-analysis and 10 in the second. Two in the first were excluded from the sec-

ond and six others were added.¹⁸ No explanation is given for why different sets of articles appeared in the two meta-analyses. In any event, the truncated set of 12 studies was not remotely representative of the larger scientific literature on the psychological effects of solitary confinement.

B. Overreliance on the Colorado Study

Even “the most thorough sampling and complete data recovery cannot make up for basic limitations in the data base” (Strube, Gardner, and Hartmann 1985, p. 68). Indeed, “An experiment that is deficient in either statistical conclusion validity, internal validity, or construct validity is meaningless and, therefore, worthless. Consequently, it should not be used” (Chow 1987, p. 266). Notwithstanding these basic methodological truisms, tables 2 and 4 in Morgan et al. (2016) reveal that both meta-analyses relied primarily on the fatally flawed Colorado study. It provided the bulk of the effect sizes on which their overall conclusions were based.

Thus, in the first meta-analysis, I counted 24 of 50 relevant effect sizes on “psychological outcomes” that came from the Colorado study. In the second meta-analysis, 140 of 210 effect sizes came from the Colorado study.¹⁹ Because of its sample size, the weights given to the multiple effect sizes from the Colorado study dwarf those of most of the other studies included.

As tables 2 and 4 in Morgan et al. (2016) make clear, they repackaged the Colorado results in a way that allowed them to dominate the analyses.²⁰ Thus, when they claimed that their results “are even more compelling when one considers that primary studies with the strongest designs produced much smaller effects,” they were referring primarily to the un-

¹⁸ The first (Morgan et al. 2016, table 2) included six studies that explicitly addressed psychological effects of solitary confinement: Ecclestone, Gendreau, and Knox (1974), Suedfeld et al. (1982), Miller and Young (1997), Zinger, Wichmann, and Andrews (2001), Andersen et al. (2003), and O’Keefe et al. (2010). The second (Morgan et al. 2016, table 4) added six studies: Walters, Callagan, and Newman (1963), Miller (1994), Coid et al. (2003), Cloyes et al. (2006), and Kaba et al. (2014); but it omitted Suedfeld et al. (1982) and Andersen et al. (2003).

¹⁹ “Anti-social indicators” such as “re-admission” and “behavior” like re-arrest and “physical health” outcomes were omitted from this calculation of psychological effects.

²⁰ Zinger, Wichmann, and Andrews (2001) accounted for another four effect sizes in table 2 and 30 in table 4. It too is fundamentally flawed, as I explain in the next section. By my count, it and the Colorado study account for 28 of 50 relevant effect sizes in the first meta-analysis and 170 of 210 in the second.

interpretable O'Keefe et al. (2010) study. However, few if any of the fundamental defects of the Colorado study were even mentioned and none was seriously engaged. Instead, the authors simply described the Colorado study as "the most sophisticated study" ever done on the topic (Morgan et al. 2016, p. 441) and relied on it for the bulk of their conclusions.²¹

C. Including Other Methodologically Flawed Studies

There are serious problems with a number of the other studies included in the Morgan et al. (2016) analyses. For example, Zinger, Wichmann, and Andrews (2001) accounted for the next-largest number of effect sizes in their meta-analyses. However, there are several problems with how the results of this study were treated and serious issues with how the study itself was conducted, raising questions about whether it should have been included at all. Its sample size is erroneously listed in table 2 as 136. Although 136 was the initial number of participants, only 60 remained at the end of 60 days. The *N* shown in table 4 is, correctly, the 60 who remained, but that also is misleading. That number includes a majority of prisoners in the "administrative segregation" group (13 of 23) who were there voluntarily. Only 10 involuntary prisoners remained in administrative segregation at the end of 60 days. Thus this study was weighted far too heavily in the first meta-analysis and given a misleading weight in the second.

The results of Zinger, Wichmann, and Andrews (2001) are in any case impossible to interpret. They are based on data from a sample that combined "voluntarily" and "involuntarily" segregated prisoners. Voluntarily isolated prisoners (such as protective custody prisoners who "choose" to be in isolation) control their own fates; at least in theory, they can leave. In addition, in most cases they know that by staying they are at least safe from threats to their well-being elsewhere in the prison system, ones they presumably fear and necessarily want to avoid more than the pain and harm they may endure in solitary confinement. They are thus

²¹ Morgan et al. (2016) appear to have overweighted the disproportionate number of effect sizes they took from the Colorado study, treating the *N*'s in each group as though their integrity was maintained throughout. However, as I noted, the bulk of the Colorado study participants moved back and forth between groups. Thus the "uncontaminated" cases are far fewer than Morgan et al. cited and used. Because O'Keefe et al. (2010) did not disaggregate their data, Morgan et al. must have relied on the confounded results, treating all participants as if they remained in their original groups for the duration of the study and weighted effect sizes as if this had been the case.

motivated to adapt to their isolation—or to appear to have adapted to it—in ways that involuntarily isolated prisoners are not. They should not be treated as if their experiences represent the effects of solitary confinement on involuntarily segregated prisoners.

A second and more important problem is the significant amount of attrition that occurred. Especially in longitudinal research, participants leave studies for various reasons. This inevitably complicates comparisons over time or between groups because people who remain are likely to be different from those who leave, thereby changing the compositions of the groups in ways that are difficult to specify.²² This is especially a problem in prison research because prison administrators decide where prisoners are housed, under what conditions, and for how long; they do so on the basis of considerations that have nothing to do with the goals of researchers. In Zinger, Wichmann, and Andrews (2001), the reduction in the number of administrative segregation prisoners after 60 days, from 83 to 23, only 10 of whom were involuntary, means that attrition reduced the number of involuntarily segregated prisoners by 80 percent. The reasons for the attrition were not given.

Attrition is seldom random. That it results largely, if not entirely, from decisions made by prison administrators means that Zinger, Wichmann, and Andrews (2001) wound up with a group that was significantly different, in indeterminate ways, from the group with which they began.²³ They do not report whether and in what ways the prisoners who remained differed from those with whom the study began.²⁴

²² Zinger, Wichmann, and Andrews acknowledge this: “Attrition is a major drawback to psychological research in general. The problem with attrition is especially relevant to the evaluation of the psychological effects of segregation” (2001, p. 56). However, they ignored the extent of this problem in presenting and interpreting their results.

²³ If, for example, disproportionate numbers of transferred prisoners were considered too “vulnerable” to remain in administrative segregation, were reacting especially negatively, or were adjusting poorly and were especially effective at convincing the prison administration to return them to the general prison population, those left behind would be, by definition, those least affected by the experience. Alternatively, if those who remained at the end of 60 days were the most recalcitrant and least compliant, perhaps explaining why the prison administrators were less likely to release them, they may have been especially “difficult” prisoners who were less likely to admit vulnerability or weakness in the assessments they underwent. Or if the voluntary administrative segregation prisoners remaining after 60 days were the least willing or able to return to the general prison population, they may have been unlikely to admit that they were suffering lest this jeopardize their continued safekeeping. Any of these possible scenarios could greatly compromise interpretation of the results, and none of them appear to have been considered.

²⁴ The assertion that “none of the attrition was attributable to prisoners being incapable of participating in the study because of episodes of delusion or hallucination or suicide at-

An additional methodological problem was acknowledged in passing but not fully discussed, either in the published article or in Zinger's (1998) dissertation, on which it was based. "Practice effects" are a common problem in longitudinal studies because they require repeated administration over time of the same tests or measures. Participants may recall the questions and intentionally or inadvertently try to reproduce the same or similar answers, or lose interest and reply with stock, rote answers, or, if the tests include performance measures, improve (because of practice) each time they take the test. If any of these things occurs, the existence of real changes (especially negative ones) will be masked or minimized.

Zinger (1998) himself recognized that "artifacts of repeated testing" likely played a role in producing apparent improvements in functioning and the lack of signs of deterioration and that practice effects may have accounted for prisoners "report[ing] less problems over time" (p. 93). He also observed that it is well known that "participants lose interest in answering repeatedly to identical questions and tend to report less problems over time" (p. 92).²⁵ Thus, practice effects may have accounted in large part for the findings of "no change" or "improvement" on the measures used and repeatedly administered.

There are also significant problems with several other studies that were included in the already small group that Morgan et al. (2016) considered. For example, Cloyes et al. (2006) did not compare administrative segregation with nonadministrative segregation at all. Instead, all of the prisoners involved in their study were in solitary confinement. The effect size Morgan et al. reported was the only statistical test of differences between groups that appeared anywhere in Cloyes et al. (2006, p. 772). However, it is a *t*-test of differences in Brief Psychiatric Rating Scale scores between two groups of solitary confinement prisoners—those identified as seriously mentally ill or not, both of which were housed in isolation. Data from this study did not belong in the meta-analysis.

tempts" (Zinger, Wichmann, and Andrews 2001, p. 71) sets far too high a threshold and does not adequately address the matter. "Episodes of delusion or hallucination or suicide attempts" are hardly the only measures of whether someone is being so adversely affected that he would seek to be transferred elsewhere or, in the opinion of a correctional administrator or mental health staff member, need to be moved.

²⁵ Zinger, Wichmann, and Andrews (2001) did acknowledge that reports of "better mental health and psychological functioning over time" are "common in studies which rely on studies with repeated measures designs" (p. 74) but then ignored the implications of this for interpretation of results that showed exactly this.

Walters, Callagan, and Newman (1963) arguably does not belong either. It is over 50 years old and, more importantly, the participants were all volunteers. They were not typical of prisoners involuntarily placed in solitary confinement. In addition, the study lasted only 4 days, not long enough to reach a conclusion that the psychological effects of solitary confinement are minimal. The one effect size Morgan et al. (2016) reported, for “anxiety,” is .57 with a weight of .726 (table 4, p. 452). Yet the only mention of numerical data for anxiety in Walters, Callagan, and Newman’s study was this: “More isolated than non-isolated prisoners reported an increase in anxiety from the pre-test to post-test period ($p = .038$, Fisher’s Exact Probability Test).” It is impossible to calculate an effect size from this statistic.

Another included study, Andersen et al. (2003, table 2), reported only chi-squares and p -values. It is not clear how Morgan et al. (2016) managed to calculate effect sizes from those data.

The decision to include Ecclestone, Gendreau, and Knox (1974) is also questionable. The study is more than 40 years old and, more importantly, included only prisoners who volunteered to spend 10 days in isolation. For previously noted reasons, the experience of volunteers is not comparable to that of involuntary administrative segregation prisoners. In addition, the study used an almost indecipherable measure of psychological functioning—the Repertory Grid Technique—which does not appear to have been used in published prison research before or since.²⁶ Moreover, half of the initial participants “quit the experiment after two days of solitary confinement” (p. 179), which meant that the assignment of participants was no longer “random,” the results suffered from significant attrition bias, and the remaining volunteer participants knew that they could leave whenever they wanted. Notwithstanding these problems, Ecclestone, Gendreau, and Knox concluded that isolated confinement was “not more stressful than normal institutional life” (p. 178). Morgan et al. (2016) included this study in both meta-analyses and singled it out as having one of the stronger research designs (along with Zinger, Wichmann, and Andrews [2001] and O’Keefe et al. [2010]).²⁷

²⁶ Description of the nature and scoring of the Repertory Grid Technique was so complicated that it consumed nearly two full pages of text (Ecclestone, Gendreau, and Knox 1974, pp. 180–81).

²⁷ The studies deemed to have stronger research designs were identified by name only in Morgan et al.’s (2016) Research Synthesis I, although an estimate of the strength of the designs was also apparently used in Research Synthesis II. Morgan et al. concluded that

In sum, Morgan et al.'s (2016) meta-analyses were based on one fundamentally flawed and uninterpretable study (O'Keefe et al. 2010), another with an attrition rate of 80 percent over a 60-day period (Zinger, Wichmann, and Andrews 2001), two that were four decades old and included only volunteers (Walters, Callagan, and Newman 1963; Ecclestone, Gendreau, and Knox 1974), and one (Cloyes et al. 2006) that could not provide an effect size on the impact of AS.

Few readers are intimately familiar with the solitary confinement literature or willing to invest the effort to read and evaluate each of the studies cited in Morgan et al. (2016). Similarly, few are willing to carefully examine the hundreds of effect sizes included in the two meta-analyses or are able to make judgments about the propriety of the particular statistical techniques used in the calculations.²⁸ The presentation of a vast array of numerical data in Morgan et al. gives the impression of an objective representation of equally meaningful effect sizes, but it is not the reality. Their conclusion that solitary confinement has modest or no significant negative psychological effects is not at all what a significant preponderance of the relevant empirical research shows and is at odds with findings

these studies with "stronger designs" were the ones that showed "less impairment" due to isolated confinement (p. 456). My critical discussion of the individual studies in question shows why.

²⁸ Morgan et al. (2016) appear to have used statistical methods that require very stringent assumptions and will give misleading results if these assumptions are violated (e.g., Aguinis, Gottfredson, and Wright 2011). Furthermore, the meta-analytic method they used requires a large number of studies to assess these assumptions, and there were not enough studies to assess them. Specifically, they used a random-effects meta-analysis model. This model assumes that the included studies are a random sample from some definable universe of studies. For example, are the prisons represented in Morgan et al.'s meta-analysis a random sample of all US prisons? If not, they cannot claim that their results generalize to this universe. Random-effects meta-analyses also assume that weights and sample sizes are uncorrelated with the effect sizes. If they are correlated, the results will be biased. The correlation between the sample sizes and effect sizes reported in their table 1 indicate that the correlation is about $-.5$, which could severely bias the results. In a random-effects meta-analysis, both the mean and the variance of the effect sizes in the universe are key parameters that need to be estimated and both require confidence intervals. Morgan et al. reported only the sample estimate of the variance and not the confidence interval. However, the confidence interval for the variance requires a strong assumption of normally distributed effect sizes, and the confidence interval is very sensitive to minor violations of this assumption. A large number of studies are needed to assess the normality assumption—much larger than the number used. Morgan et al. also appear to have used a new and unproven method for combining multiple effect sizes from a single study. This method requires at least a moderate number of studies (10–20, the more the better), more than the separate meta-analyses that were used. Finally, Morgan et al. also used extremely crude and inaccurate methods to approximate effect sizes in studies that did not provide enough information to correctly compute an effect size.

that are consistent across many decades, theoretically coherent, and buttressed by a very large and growing literature on the harmful effects of social isolation in contexts other than prison.

Misleading repackaging of bad data can ripple through the field and produce an echo chamber in which motivated commentators repeat each others' flawed conclusions. Thus O'Keefe (2017, p. 5) recently asserted that "a recent meta-analysis found small to moderate adverse psychological effects resulting from [solitary confinement] that were no greater in magnitude than the overall effects of incarceration. These findings are consistent with our Colorado results." She was referring to the Morgan et al. (2016) meta-analysis, whose conclusions were not only "consistent" with the Colorado results but based largely on them.

IV. Conclusion

These two studies offer several cautionary tales about the fraught nature of prison research, especially on the methodologically challenging and politically charged topic of solitary confinement. The first of these tales is about the potential influence of bad, uninterpretable data on public discourse and correctional policy. Once the results of research that bear the trappings of science enter into public and policy discourse, it is difficult to correct the record, especially when motivated advocates are willing to overlook fatal flaws in the research. Unfortunately, when this transpires, researchers can lose control of the narrative by which their research is described and the manner in which it is applied. For example, O'Keefe has repeatedly and steadfastly defended her Colorado research but has opposed the uses to which others have put it. She was emphatic that she did "not believe in any way and we do not promote the study as something to argue for the case of segregation. . . . My interpretation is that people believe that this study sanctions administrative segregation for mentally ill and non-mentally ill alike. . . . I do not believe that the conclusions lend to that and that is not the intended use of our study" (2013, p. 96).²⁹ Yet, that is exactly the use to which a number of interested parties have put it.

²⁹ Two prominent advisory board members, Jeffrey Metzner and Jamie Fellner (2010), published a "post-Colorado study" article that seemed to contravene the study's findings. They conceded that "isolation can be harmful to any prisoner" and noted that the potentially adverse effects of isolation include "anxiety, depression, anger, cognitive disturbances, perceptual distortions, obsessive thoughts, paranoia, and psychosis" (p. 104)—not at all what

The Colorado study is also a stark reminder that attempts to implement conventional experimental or even quasi-experimental research designs in prison environments face a number of often insurmountable obstacles. The ordinary demands of prison operations nearly always doom even the most carefully planned such studies, and certainly anything resembling a traditional experiment. Savvy prison researchers understand that the desire to treat a prison environment as if it were a research laboratory should be resisted. Real people live (and die) in prison, a setting in which the core dynamics between prisoners and staff are governed by forces beyond the researchers' control.

In separate but related ways, both the Colorado study and the Morgan et al. (2016) meta-analyses underscore the pitfalls of allowing the veneer of scientific rigor to substitute for its reality. They also show the limitations of focusing on quantitative outcomes with little or no concern for precisely how and under what conditions data were acquired. The decontextualized and de-individualized approach to data collection that characterized the Colorado study allowed researchers to treat all participants within each of the study groups as if they were the same, when clearly they—and especially their prison experiences—were not. Ignoring the prison context and individual prisoner trajectories helped render the findings incoherent and uninterpretable.

Similarly, Morgan et al. (2016) illustrate the shortcomings of attempting to apply an otherwise useful approach for summarizing quantitative data to environments as complex and variable as prisons (or especially solitary confinement units). Whatever the benefits of reducing empirical results to effect sizes may be, omitting an entire field's best-known and most in-depth works from consideration because most do not lend themselves to meta-analytic reductions means that nuance and context are inevitably ignored. The compromise in "scientific truth" is far too great.

Some critics of meta-analysis argue that "a literature review should *not* be a formalized or standardized one" (Chow 1987, p. 267; emphasis

the Colorado study claimed. Metzner and Fellner's deep concerns led them to recommend that professional organizations "should actively support practitioners who work for changed segregation policies and they should use their institutional authority to press for a nationwide rethinking of the use of isolation" in the name of their "commitment to ethics and human rights" (p. 107). Zinger has become an eloquent critic of the use of solitary confinement in Canada (e.g., Makin 2013) even though defenders of the practice continue to cite his dissertation research to justify its use.

added). As Chow observed, “It is not the case that narrative reviews lack rigor. To the contrary, rigor is maintained by reviewers of the traditional [narrative] approach when they evaluate the validity of individual studies” (p. 268). Meta-analyses, even when done well, risk compromising the richness of the prison data they seek to summarize.

In any event, the magnitude of what can be and often is lost in the course of the compromises made in the kind of research critically discussed in this essay often goes unrecognized. Amid thousands of data entries and hundreds of effect sizes reported in these two studies, there are few references to the core subjectivity, institutional trajectory, or life outcome of a single individual prisoner confined in an isolation unit. Nor is there acknowledgment that the studies focused on human beings rather than on interchangeable data points.

Martha Nussbaum (1995) noted in a different context that regarding people as “fungible” and denying them their subjectivity are powerful ways to ensure their objectification. Objectivity in prison research is a worthy goal, except when it results in objectification of prisoners and others in the prison environment. Feeley and Simon (1992) observed that the era of mass imprisonment occasioned and was facilitated by the emergence of a “new penology” whose key elements—“statistical prediction, concern with groups, strategies of management”—shifted the focus of the prison enterprise “toward mechanisms of appraising and arranging groups rather than intervening in the lives of individuals” (p. 459). This actuarial approach still defines the modern prison. It should not be made worse and reinforced by scholarship that exacerbates rather than alleviates or exposes these depersonalizing tendencies.

Studying only at a distance, as the research criticized in this essay did, requires precisely that kind of objectifying sacrifice. If John Irwin was right, that the close study of people in general and prisoners in particular uncovers their humanity, and I think he was, then the opposite is also true. Studying prisoners at a distance, without trying fully to understand and adequately to convey the conditions in which they live or to gain an “appreciation of their meaning worlds, motivations, and aspirations” (1987, p. 47), leaves us with little insight into basic truths about them. That includes whether and how much they are adversely affected by near-total deprivation of meaningful sensory and social contact.

The insurmountable methodological flaws of the Colorado study and the fundamental inadequacy of the Morgan et al. (2016) meta-analysis

410 Craig Haney

should preclude policy makers from using either in debates over the proper use of solitary confinement and the nature of its psychological effects.

REFERENCES

- Aguinis, Herman, Ryan Gottfredson, and Thomas Wright. 2011. "Best-Practice Recommendations for Estimating Interaction Effects Using Meta-Analysis." *Journal of Organizational Behavior* 32:1033-43.
- American Psychological Association. 2016. *Statement on the Solitary Confinement of Juvenile Offenders*. Washington, DC: American Psychological Association, Public Information Government Relations Office. <http://www.apa.org/about/gr/issues/cyf/solitary.pdf>.
- Andersen, Henrik, D. Setsoft, T. Lillebaek, G. Gabrielsen, and R. Hemmingsen. 2003. "A Longitudinal Study of Prisoners on Remand: Repeated Measures of Psychopathology in the Initial Phase of Solitary versus Non-solitary Confinement." *International Journal of Law and Psychiatry* 26:165-77.
- Arrigo, Bruce, and Jennifer Bullock. 2008. "The Psychological Effects of Solitary Confinement on Prisoners in Supermax Units: Reviewing What We Know and What Should Change." *International Journal of Offender Therapy and Comparative Criminology* 52:622-40.
- Associated Press. 2007. "Limon Prison under Lockdown after Assault." *Vail (CO) Daily*, September 13. <http://www.vaildaily.com/news/limon-prison-under-lockdown-after-assault/>.
- Association of State Correctional Administrators and the Arthur Liman Public Interest Program. 2016. *Aiming to Reduce Time-in-Cell: Reports from Correctional Systems on the Numbers of Prisoners in Restricted Housing and on the Potential of Policy Changes to Bring About Reforms*. New Haven, CT: Yale Law School.
- Barnes, Harry Elmer. 1921. "The Historical Origin of the Prison System in America." *Journal of Criminal Law and Criminology* 12:35-60.
- Barte, Henri. 1989. "L'isolement carceral." *Perspectives Psychiatriques* 28:252-55.
- Bauer, Michael, Stefan Priebe, Bettina Haring, and Kerstin Adamczak. 1993. "Long-Term Mental Sequelae of Political Imprisonment in East Germany." *Journal of Nervous and Mental Disease* 181:257-62.
- Benjamin, Thomas, and Kenneth Lux. 1975. "Constitutional and Psychological Implications of the Use of Solitary Confinement: Experience at the Maine State Prison." *Clearinghouse Review* 9:83-90.
- Berger, Ronald, Paul Chaplin, and Robert Trestman. 2013. "Commentary: Toward an Improved Understanding of Administrative Segregation." *Journal of the American Academy of Psychiatry and Law* 41:61-64.
- Boulet, Jack, and Marvin Boss. 1991. "Reliability and Validity of the Brief Symptom Index." *Psychological Assessment* 3:433-37.
- Brodsky, Stanley, and Forrest Scogin. 1988. "Inmates in Protective Custody: First Data on Emotional Effects." *Forensic Reports* 1:267-80.

- Burney, Christopher. 1961. *Solitary Confinement*. New York: Macmillan.
- Cacioppo, Stephanie, and John Cacioppo. 2012. "Decoding the Invisible Forces of Social Connections." *Frontiers of Integrative Neuroscience* 6:1-7.
- Card, Noel. 2012. *Applied Meta-Analysis for Social Science Research*. New York: Guilford.
- Chow, Sui. 1987. "Meta-Analysis of Pragmatic and Theoretical Research: A Critique." *Journal of Psychology* 121(23):259-71.
- Cloyes, Kristin, David Lovell, David Allen, and Lorna Rhodes. 2006. "Assessment of Psychosocial Impairment in a Supermaximum Security Unit Sample." *Criminal Justice and Behavior* 33:760-81.
- Cohen, Stanley, and Laurie Taylor. 1972. *Psychological Survival: The Experience of Long-Term Imprisonment*. New York: Pantheon.
- Coid, Jeremy, Ann Petruckevitch, Paul Bebbington, Rachel Jenkins, Traolach Brugha, Glyn Lewis, Michael Farrell, and Nicola Singleton. 2003. "Psychiatric Morbidity in Prisoners and Solitary Cellular Confinement. I: Disciplinary Segregation." *Journal of Forensic Psychiatry and Psychology* 14:298-319.
- Commission on Crime Prevention and Criminal Justice. 2015. *United Nations Standard Minimum Rules for the Treatment of Prisoners*. New York: United Nations Economic and Social Council.
- Cormier, Bruno, and Paul Williams. 1966. "Excessive Deprivation of Liberty." *Canadian Psychiatric Association Journal* 11(6):470-84.
- Correctional News. 2012. "New Colorado Prison Closed Down." November 14. <http://correctionalnews.com/2012/11/14/new-colorado-prison-closed-down/>.
- Coyne, James, Brett Thombs, and Mariet Hagedorn. 2010. "It Ain't Necessarily So: A Review and Critique of Recent Meta-Analyses of Behavioral Medicine Interventions in Health Psychology." *Health Psychology* 29:107-16.
- Cressey, Donald. 1940. *The Prison Community*. Boston: Christopher Publishing.
- Crewe, Ben. 2009. *The Prisoner Society: Power, Adaptation, and Social Life in an English Prison*. Oxford: Oxford University Press.
- Dickens, Charles. 1842. *American Notes*. New York: Harper & Brothers.
- Ecclestone, C., Paul Gendreau, and Clifford Knox. 1974. "Solitary Confinement of Prisoners: An Assessment of Its Effects on Inmates' Personal Constructs and Adrenal Cortical Activity." *Canadian Journal of Behavioral Science* 6:178-91.
- Feeley, Malcolm, and Jonathan Simon. 1992. "The New Penology: Notes on the Emerging Strategy of Corrections and Its Implications." *Criminology* 30:449-74.
- Fiorillo, Damiano, and Fabio Sabatini. 2011. "Quality and Quantity: The Role of Social Interactions in Self-Reported Individual Health." *Social Science and Medicine* 73:1644-52.
- Foster, Don, Dennis Davis, and Diane Sandler. 1987. *Detention and Torture in South Africa*. New York: St. Martin's.
- Franke, Herman. 1992. "The Rise and Decline of Solitary Confinement." *British Journal of Criminology* 32:125-43.
- Gendreau, Paul, and James Bonta. 1984. "Solitary Confinement Is Not Cruel and Unusual: People Sometimes Are!" *Canadian Journal of Criminology* 26:467-78.

412 Craig Haney

- Gendreau, Paul, and Ryan Labrecque. 2016. "The Effects of Administrative Segregation: A Lesson in Knowledge Cumulation." In *The Oxford Handbook of Prisons and Imprisonment*, edited by John Wooldredge and Paula Smith. New York: Oxford University Press.
- Gendreau, Paul, and Yvette Theriault. 2011. "Bibliotherapy for Cynics Revisited: Commentary on *One Year Longitudinal Study of the Psychological Effects of Administrative Segregation*." *Corrections and Mental Health: An Update of the National Institute of Corrections*. <https://community.nicic.gov/blogs/mental-health/archive/2011/06/21/bibliotherapy-for-cynics-revisited-commentary-on-one-year-longitudinal-study-of-the-psychological-effects-of-administrative-segregation.aspx>.
- Gendreau, Paul, et al. 1972. "Changes in EEG Alpha Frequency and Evoked Response Latency during Solitary Confinement." *Journal of Abnormal Psychology* 79:54–59.
- Government Accountability Office. 2013. *Improvements Needed in Bureau of Prisons' Monitoring and Evaluation of Impact of Segregated Housing*. Washington, DC: US Government Accountability Office.
- Grassian, Stuart. 1983. "Psychopathological Effects of Solitary Confinement." *American Journal of Psychiatry* 140(11):1450–54.
- . 2006. "Psychiatric Effects of Solitary Confinement." *Washington University Journal of Law and Policy* 22:325–83.
- Grassian, Stuart, and Terry Kupers. 2011. "The Colorado Study versus the Reality of Supermax Confinement." *Correctional Mental Health Report* 13(1):1, 9–11.
- Hafner, S., R. Emeny, M. Lacruz, J. Baumert, C. Herder, et al. 2011. "Association between Social Isolation and Inflammatory Markers in Depressed and Non-depressed Individuals: Results from the MONICA/KORA Study." *Brain Behavior and Immunity* 25:1701–7.
- Haney, Craig. 1993. "Infamous Punishment: The Psychological Consequences of Isolation." *National Prison Project Journal* 8(2):3–7, 21.
- . 2003. "Mental Health Issues in Long-Term Solitary and 'Supermax' Confinement." *Crime and Delinquency* 49:124–56.
- Haney, Craig, and Mona Lynch. 1997. "Regulating Prisons of the Future: The Psychological Consequences of Solitary and Supermax Confinement." *New York University Review of Law and Social Change* 23:477–570.
- Hibbert, Christopher. 1963. *The Roots of Evil: A Social History of Crime and Punishment*. Boston: Little, Brown.
- Hilliard, Thomas. 1976. "The Black Psychologist in Action: A Psychological Evaluation of the Adjustment Center Environment at San Quentin Prison." *Journal of Black Psychology* 2:75–82.
- Hinkle, Lawrence, and Harold Wolff. 1956. "Communist Interrogation and Indoctrination of 'Enemies of the State.'" *Archives of Neurology and Psychiatry* 76: 115–74.
- Holt-Lunstad, Julianne, Theodore Robles, and David Sbarra. 2017. "Advancing Social Connectedness as a Public Health Priority in the United States." *American Psychologist* 72:517–30.

- Hunt, Stephen, Mack Orsborn, Harvey Checkoway, Mary Biggs, Miles McFall, and Tim Takaro. 2008. "Later Life Disability Status Following Incarceration as a Prisoner of War." *Military Medicine* 173:613–18.
- Immarigeon, Russ. 2011. "One Year Longitudinal Study of the Psychological Effects of Administrative Segregation: Introduction." *Corrections and Mental Health: An Update of the National Institute of Corrections*. <https://community.nic.gov/blogs/mentalhealth/archive/2011/06/21/one-year-longitudinal-study-of-the-psychological-effects-of-administrative-segregation-introduction.aspx>.
- Irwin, John. 1980. *Prisons in Turmoil*. Boston: Little, Brown.
- . 1987. "Reflections on Ethnography." *Journal of Contemporary Ethnography* 16:41–48.
- Jackson, Michael. 1983. *Prisoners of Isolation: Solitary Confinement in Canada*. Toronto: University of Toronto Press.
- Jacobs, James. 1977. *Stateville*. Chicago: University of Chicago Press.
- Jones, David. 1976. *The Health Risks of Imprisonment*. Lexington, MA: Lexington Books.
- Kaba, Fatos, Andrea Lewis, Sarah Glowa-Kollisch, James Hadler, David Lee, Howard Alper, Daniel Selling, Ross MacDonald, Angela Solimo, Amanda Parsons, and Homer Venters. 2014. "Solitary Confinement and Risk of Self-Harm among Jail Inmates." *American Journal of Public Health* 104:442–47.
- Kellett, Stephen, Nigel Beail, David Newman, and Pat Frankish. 2003. "Utility of the Brief Symptom Inventory in the Assessment of Psychological Distress." *Journal of Applied Research in Intellectual Disabilities* 16:127–34.
- Koch, Ida. 1986. "Mental and Social Sequelae of Isolation: The Evidence of Deprivation Experiments and of Pretrial Detention in Denmark." In *The Expansion of European Prison Systems*, edited by Bill Rolston and Mike Tomlinson. Working Papers in European Criminology, no. 7. Stockholm: European Group for the Study of Deviance and Social Control.
- Kupers, Terry. 2017. *Solitary: The Inside Story of Supermax Isolation and How We Can Abolish It*. Oakland: University of California Press.
- Lieberman, Matthew. 2013. *Social: Why Our Brains Are Wired to Connect*. New York: Random House.
- Liebling, Alison. 1999. "Doing Research in Prison: Breaking the Silence?" *Theoretical Criminology* 3:147–73.
- Lovell, David, and Hans Toch. 2011. "Some Observations about the Colorado Segregation Study." *Correctional Mental Health Report* 13(1):3–4, 14.
- Makin, Kirk. 2013. "Canadian Prisons 'Out of Step' on Solitary Confinement." *Globe and Mail*, March 21. <https://www.theglobeandmail.com/news/politics/canadian-prisons-out-of-step-on-solitary-confinement/article10103358/>.
- Masur, Louis. 1989. *Rites of Execution: Capital Punishment and the Transformation of American Culture, 1776–1865*. New York: Oxford University Press.
- Mendez, Juan. 2011. *Interim Report of the Special Rapporteur of the Human Rights Council on Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment*. New York: United Nations.
- Metzner, Jeffrey, and Jamie Fellner. 2010. "Solitary Confinement and Mental Illness in U.S. Prisons: A Challenge for Medical Ethics." *Journal of the Acad-*

414 Craig Haney

- eny of *Psychiatry and Law* 38:104–8. http://www.hrw.org/sites/default/files/related_material/Solitary%20Confinement%20and%20Mental%20Illness%20in%20US%20Prisons.pdf.
- Metzner, Jeffrey, and Maureen O’Keefe. 2011. “The Psychological Effects of Administrative Segregation: The Colorado Study.” *Correctional Mental Health Report* 13(1):1–2, 12–14.
- Miller, Holly. 1994. “Reexamining Psychological Distress in the Current Conditions of Segregation.” *Journal of Correctional Health Care* 1:39–53.
- Miller, Holly, and Glenn Young. 1997. “Prison Segregation: Administrative Remedy or Mental Health Problem?” *Criminal Behaviour and Mental Health* 7:85–94.
- Mitchell, Kirk. 2010. “Limon Prison Incentives Keep Inmates in Check.” *Denver Post*, November 24. <http://www.deseretnews.com/article/700085422/Limon-prison-incentives-keep-inmates-in-check.html>.
- Morgan, Robert, Paul Gendreau, Paula Smith, Andrew Gray, Ryan Labrecque, Nina MacLean, Stephanie Van Horn, Angelea Bolanos, Ashley Batastini, and Jeremy Mills. 2016. “Quantitative Synthesis of the Effects of Administrative Segregation on Inmates’ Well-Being.” *Psychology, Public Policy, and Law* 22:439–61.
- National Commission on Correctional Health Care. 2016. “Position Statement: Solitary Confinement (Isolation).” *Journal of Correctional Health Care* 22:257–63. <http://www.ncchc.org/solitary-confinement>.
- National Research Council. 2014. *The Growth of Incarceration in the United States: Exploring Causes and Consequences*. Washington, DC: National Academies Press.
- Nowak, Manfred. 2006. “What Practices Constitute Torture?” *Human Rights Quarterly* 28:809–41.
- Nussbaum, Martha. 1995. “Objectification.” *Philosophy and Public Affairs* 24:249–91.
- O’Donnell, Ian. 2014. *Prisoners, Solitude, and Time*. New York: Oxford University Press.
- O’Keefe, Maureen. 2008. “Administrative Segregation from Within: A Corrections Perspective.” *Prison Journal* 88:123–43.
- . 2017. “Reflections on Colorado’s Administrative Segregation Study.” *Nij Journal* 278:1–8. <https://nij.gov/journals/278/Pages/reflections-on-colorado-administrative-segregation-study.aspx>.
- O’Keefe, Maureen, Kelli Klebe, Jeffrey Metzner, Joel Dvoskin, Jamie Fellner, and Alysha Stucker. 2013. “A Longitudinal Study of Administrative Segregation.” *Journal of the American Academy of Psychiatry and Law* 41:49–60.
- O’Keefe, Maureen, Kelli Klebe, Alysha Stucker, Kristin Sturm, and William Leggett. 2010. *One Year Longitudinal Study of the Psychological Effects of Administrative Segregation*. Final Report to the National Institute of Justice. Washington, DC: US Department of Justice, National Institute of Justice.
- Raemisch, Rick. 2017. “Putting an End to Long-Term Solitary.” *New York Times*, October 13. <https://www.nytimes.com/2017/10/12/opinion/solitary-confinement-colorado-prison.html>.
- Reiter, Keramet. 2016. *23/7: Pelican Bay Prison and the Rise of Long-Term Solitary Confinement*. New Haven, CT: Yale University Press.

- Reyes, Hernan. 2007. "The Worst Scars Are in the Mind: Psychological Torture." *International Review of the Red Cross* 89:591–616. <https://www.icrc.org/en/international-review/article/worst-scars-are-mind-psychological-torture>.
- Rhodes, Lorna. 2004. *Total Confinement: Madness and Reason in the Maximum Security Prison*. Berkeley: University of California Press.
- Rhodes, Lorna, and David Lovell. 2011. "Is Adaptation the Right Question? Addressing the Larger Context of Administrative Segregation: Commentary on One Year Longitudinal Study of the Psychological Effects of Administrative Segregation." *Corrections and Mental Health: An Update of the National Institute of Corrections* (June 21). http://community.nicic.gov/cfs-file.ashx/__key/CommunityServer.Components.PostAttachments/00.00.05.95.19/Supermax-_2D00_-T-_2D00_-Rhodes-and-Lovell.pdf.
- Rundle, Frank. 1973. "The Roots of Violence at Soledad." In *The Politics of Punishment: A Critical Analysis of Prisons in America*, edited by Erik Wright. New York: Harper.
- Scott, George, and Paul Gendreau. 1969. "Psychiatric Implications of Sensory Deprivation in a Maximum Security Prison." *Canadian Psychiatric Association Journal* 14:337–41.
- Shalev, Sharon. 2009. *Supermax: Controlling Risk through Solitary Confinement*. Portland, OR: Willan.
- Shalev, Sharon, and Monica Lloyd. 2011. "If This Be Method, Yet There Is Madness in It: Commentary on One Year Longitudinal Study of the Psychological Effects of Administrative Segregation." *Corrections and Mental Health: An Update of the National Institute of Corrections* (June 21). http://community.nicic.gov/cfs-file.ashx/__key/CommunityServer.Components.PostAttachments/00.00.05.95.21/Supermax-_2D00_-T-_2D00_-Shalev-and-Lloyd.pdf.
- Slater, Robert. 1986. "Psychiatric Intervention in an Atmosphere of Terror." *American Journal of Forensic Psychology* 7(1):5–12.
- Smith, Peter Scharff. 2006. "The Effects of Solitary Confinement on Prison Inmates: A Brief History and Review of the Literature." In *Crime and Justice: A Review of Research*, vol. 34, edited by Michael Tonry. Chicago: University of Chicago Press.
- . 2011. "The Effects of Solitary Confinement: Commentary on One Year Longitudinal Study of the Psychological Effects of Administrative Segregation." *Corrections and Mental Health: An Update of the National Institute of Corrections* (June 21). http://community.nicic.gov/cfs-file.ashx/__key/CommunityServer.Components.PostAttachments/00.00.05.95.22/Supermax-_2D00_-T-_2D00_-Smith.pdf.
- Strube, Michael, William Gardner, and Donald Hartmann. 1985. "Limitations, Liabilities, and Obstacles in Reviews of the Literature: The Current Status of Meta-Analysis." *Clinical Psychology Review* 5:63–78.
- Suedfeld, Peter, Carmenza Ramirez, John Deaton, and Gloria Baker-Brown. 1982. "Reactions and Attributes of Prisoners in Solitary Confinement." *Criminal Justice and Behavior* 9:303–40.
- Sykes, Gresham. 1958. *Society of Captives*. Princeton, NJ: Princeton University Press.

416 Craig Haney

- Toch, Hans. 1975. *Men in Crisis: Human Breakdowns in Prison*. New York: Aldine.
- . 1977. *Living in Prison: The Ecology of Survival*. New York: Free Press.
- . 2003. "The Contemporary Relevance of Early Experiments with Supermax Reform." *Prison Journal* 83:221–28.
- Turner, Scott, Laura Cardinal, and Richard Burton. 2017. "Research Design for Mixed Methods: A Triangulation-Based Framework and Roadmap." *Organizational Research Methods* 20:243–67.
- Volkart, Reto, Adolf Dittrich, Thomas Rothenfluh, and Paul Werner. 1983. "Eine Kontrollierte Untersuchung uber Psychopathologische Effekte der Einzelhaft" [A controlled investigation on psychopathological effects of solitary confinement]. *Psychologie—Schweizerische Zeitschrift fur Psychologie und ihre Anwendungen* 42:25–46.
- Volkart, Reto, Thomas Rothenfluh, W. Kobelt, Adolf Dittrich, and K. Ernst. 1983. "Einzelhaft als Risikofaktor fur Psychiatrische Hospitalisierung" [Solitary confinement as a risk for psychiatric hospitalization]. *Psychiatria Clinica* 16:365–77.
- Waligora, Bogusław. 1974. "Funkcjonowanie Czlowieka W Warunkach Izolacji Wieziennej" [How men function in conditions of penitentiary isolation]. *Seria Psychologia I Pedagogika* 34:1–123.
- Walters, Richard, John Callagan, and Albert Newman. 1963. "Effect of Solitary Confinement on Prisoners." *American Journal of Psychiatry* 119:771–73.
- Zinger, Ivan. 1998. "The Psychological Effects of 60 Days in Administrative Segregation." PhD diss., Carlton University, Department of Psychology.
- Zinger, Ivan, Cherami Wichmann, and D. Andrews. 2001. "The Psychological Effects of 60 Days in Administrative Segregation." *Canadian Journal of Criminology* 43:47–83.

EXHIBIT 12

Specialty Guidelines for Forensic Psychology (excerpts)

Specialty Guidelines for Forensic Psychology

American Psychological Association

In the past 50 years forensic psychological practice has expanded dramatically. The American Psychological Association (APA) has a division devoted to matters of law and psychology (APA Division 41, the American Psychology–Law Society), a number of scientific journals devoted to interactions between psychology and the law exist (e.g., *Law and Human Behavior*; *Psychology, Public Policy, and Law*; *Behavioral Sciences & the Law*), and a number of key texts have been published and undergone multiple revisions (e.g., Grisso, 1986, 2003; Melton, Petrila, Poythress, & Slobogin, 1987, 1997, 2007; Rogers, 1988, 1997, 2008). In addition, training in forensic psychology is available in predoctoral, internship, and postdoctoral settings, and APA recognized forensic psychology as a specialty in 2001, with subsequent recertification in 2008.

Because the practice of forensic psychology differs in important ways from more traditional practice areas (Mohan, 1980) the “Specialty Guidelines for Forensic Psychologists” were developed and published in 1991 (Committee on Ethical Guidelines for Forensic Psychologists, 1991). Because of continued developments in the field in the ensuing 20 years, forensic practitioners’ ongoing need for guidance, and policy requirements of APA, the 1991 “Specialty Guidelines for Forensic Psychologists” were revised, with the intent of benefiting forensic practitioners and recipients of their services alike.

The goals of these Specialty Guidelines for Forensic Psychology (“the Guidelines”) are to improve the quality of forensic psychological services; enhance the practice and facilitate the systematic development of forensic psychology; encourage a high level of quality in professional practice; and encourage forensic practitioners to acknowledge and respect the rights of those they serve. These Guidelines are intended for use by psychologists when engaged in the practice of forensic psychology as described below and may also provide guidance on professional conduct to the legal system and other organizations and professions.

For the purposes of these Guidelines, *forensic psychology* refers to professional practice by any psychologist working within any subdiscipline of psychology (e.g., clinical, developmental, social, cognitive) when applying the scientific, technical, or specialized knowledge of psychology to the law to assist in addressing legal, contractual, and administrative matters. Application of the Guidelines does not depend on the practitioner’s typical areas of practice or expertise, but rather, on the service provided in the case at hand. These Guidelines apply in all matters in which psychologists provide expertise to judicial, administrative, and

educational systems including, but not limited to, examining or treating persons in anticipation of or subsequent to legal, contractual, or administrative proceedings; offering expert opinion about psychological issues in the form of amicus briefs or testimony to judicial, legislative, or administrative bodies; acting in an adjudicative capacity; serving as a trial consultant or otherwise offering expertise to attorneys, the courts, or others; conducting research in connection with, or in the anticipation of, litigation; or involvement in educational activities of a forensic nature.

Psychological practice is not considered forensic solely because the conduct takes place in, or the product is presented in, a tribunal or other judicial, legislative, or administrative forum. For example, when a party (such as a civilly or criminally detained individual) or another individual (such as a child whose parents are involved in divorce proceedings) is ordered into treatment with a practitioner, that treatment is not necessarily the practice of forensic psychology. In addition, psychological testimony that is solely based on the provision of psychotherapy and does not include psycholegal opinions is not ordinarily considered forensic practice.

For the purposes of these Guidelines, *forensic practitioner* refers to a psychologist when engaged in the practice of forensic psychology as described above. Such professional conduct is considered forensic from the time the practitioner reasonably expects to, agrees to, or is legally mandated to provide expertise on an explicitly psycholegal issue.

The provision of forensic services may include a wide variety of psycholegal roles and functions. For example, as

This article was published Online First October 1, 2012.

These Specialty Guidelines for Forensic Psychology were developed by the American Psychology–Law Society (Division 41 of the American Psychological Association [APA]) and the American Academy of Forensic Psychology. They were adopted by the APA Council of Representatives on August 3, 2011.

The previous version of the Guidelines (“Specialty Guidelines for Forensic Psychologists”; Committee on Ethical Guidelines for Forensic Psychologists, 1991) was approved by the American Psychology–Law Society (Division 41 of APA) and the American Academy of Forensic Psychology in 1991. The current revision, now called the “Specialty Guidelines for Forensic Psychology” (referred to as “the Guidelines” throughout this document), replaces the 1991 “Specialty Guidelines for Forensic Psychologists.”

These guidelines are scheduled to expire August 3, 2021. After this date, users are encouraged to contact the American Psychological Association Practice Directorate to confirm that this document remains in effect.

Correspondence concerning these guidelines should be addressed to the Practice Directorate, American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242.

maintain integrity by examining the issue or problem at hand from all reasonable perspectives and seek information that will differentially test plausible rival hypotheses.

Guideline 9.02: Use of Multiple Sources of Information

Forensic practitioners ordinarily avoid relying solely on one source of data, and corroborate important data whenever feasible (AERA, APA, & NCME, in press). When relying upon data that have not been corroborated, forensic practitioners seek to make known the uncorroborated status of the data, any associated strengths and limitations, and the reasons for relying upon the data.

Guideline 9.03: Opinions Regarding Persons Not Examined

Forensic practitioners recognize their obligations to only provide written or oral evidence about the psychological characteristics of particular individuals when they have sufficient information or data to form an adequate foundation for those opinions or to substantiate their findings (EPPCC Standard 9.01). Forensic practitioners seek to make reasonable efforts to obtain such information or data, and they document their efforts to obtain it. When it is not possible or feasible to examine individuals about whom they are offering an opinion, forensic practitioners strive to make clear the impact of such limitations on the reliability and validity of their professional products, opinions, or testimony.

When conducting a record review or providing consultation or supervision that does not warrant an individual examination, forensic practitioners seek to identify the sources of information on which they are basing their opinions and recommendations, including any substantial limitations to their opinions and recommendations.

10. Assessment

Guideline 10.01: Focus on Legally Relevant Factors

Forensic examiners seek to assist the trier of fact to understand evidence or determine a fact in issue, and they provide information that is most relevant to the psycholegal issue. In reports and testimony, forensic practitioners typically provide information about examinees' functional abilities, capacities, knowledge, and beliefs, and address their opinions and recommendations to the identified psycholegal issues (American Bar Association & American Psychological Association, 2008; Grisso, 1986, 2003; Heilbrun, Marczyk, DeMatteo, & Mack-Allen, 2007).

Forensic practitioners are encouraged to consider the problems that may arise by using a clinical diagnosis in some forensic contexts, and consider and qualify their opinions and testimony appropriately.

Guideline 10.02: Selection and Use of Assessment Procedures

Forensic practitioners use assessment procedures in the manner and for the purposes that are appropriate in light of

the research on or evidence of their usefulness and proper application (EPPCC Standard 9.02; AERA, APA, & NCME, in press). This includes assessment techniques, interviews, tests, instruments, and other procedures and their administration, adaptation, scoring, and interpretation, including computerized scoring and interpretation systems.

Forensic practitioners use assessment instruments whose validity and reliability have been established for use with members of the population assessed. When such validity and reliability have not been established, forensic practitioners consider and describe the strengths and limitations of their findings. Forensic practitioners use assessment methods that are appropriate to an examinee's language preference and competence, unless the use of an alternative language is relevant to the assessment issues (EPPCC Standard 9.02).

Assessment in forensic contexts differs from assessment in therapeutic contexts in important ways that forensic practitioners strive to take into account when conducting forensic examinations. Forensic practitioners seek to consider the strengths and limitations of employing traditional assessment procedures in forensic examinations (AERA, APA, & NCME, in press). Given the stakes involved in forensic contexts, forensic practitioners strive to ensure the integrity and security of test materials and results (AERA, APA, & NCME, in press).

When the validity of an assessment technique has not been established in the forensic context or setting in which it is being used, the forensic practitioner seeks to describe the strengths and limitations of any test results and explain the extrapolation of these data to the forensic context. Because of the many differences between forensic and therapeutic contexts, forensic practitioners consider and seek to make known that some examination results may warrant substantially different interpretation when administered in forensic contexts (AERA, APA, & NCME, in press).

Forensic practitioners consider and seek to make known that forensic examination results can be affected by factors unique to, or differentially present in, forensic contexts including response style, voluntariness of participation, and situational stress associated with involvement in forensic or legal matters (AERA, APA, & NCME, in press).

Guideline 10.03: Appreciation of Individual Differences

When interpreting assessment results, forensic practitioners consider the purpose of the assessment as well as the various test factors, test-taking abilities, and other characteristics of the person being assessed, such as situational, personal, linguistic, and cultural differences that might affect their judgments or reduce the accuracy of their interpretations (EPPCC Standard 9.06). Forensic practitioners strive to identify any significant strengths and limitations of their procedures and interpretations.

Forensic practitioners are encouraged to consider how the assessment process may be impacted by any disability an examinee is experiencing, make accommodations as

EXHIBIT 13

Dr. Kupers Report (excerpts)

CONFIDENTIAL

Expert Report of Terry A. Kupers, M.D., M.S.P.

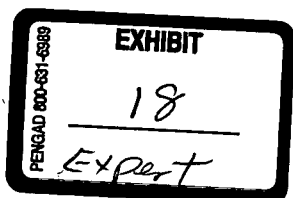
Re: *Vermillion v. Levenhagen*, 1:15-CV-0605-RLY-TAB

TABLE OF CONTENTS

- I. Background and Qualifications
- II. Preparation for My Testimony
- III. Summary of Litigation
- IV. Childhood History
- V. Adult Life Prior to Prison
- VI. Mental Health History
- VII. Criminal Justice Record
- VIII. Jail Experience
- IX. Experience in Prison General Population
- X. Research and Literature about Solitary Confinement
- XI. Mr. Vermillion's Experience in Solitary Confinement
- XII. Mr. Vermillion's Psychological Response to Solitary Confinement
- XIII. Mr. Vermillion's Life in Prison After Release from Solitary
- XIV. Mental Status Examination
- XV. Physical Health
- XVI. Summary of Opinions

I. Background and Qualifications

I am a board-certified psychiatrist, Institute Professor at the Wright Institute, Distinguished Life Fellow of the American Psychiatric Association, and an expert on correctional mental health issues. I graduated from UCLA School of Medicine with an M.D. degree in 1968, completed an internship in medicine and pediatrics at Kings County Hospital/Downstate Medical Center in 1969, completed residency in psychiatry at UCLA NPI in 1972, and completed a Fellowship in Social and Community Psychiatry in 1974 with an MSP degree from UCLA (Masters in Social and Community Psychiatry). I have testified more than thirty times in state and federal courts about the psychiatric and physiological effects of jail and prison conditions, the quality of correctional management and mental health treatment, and prison sexual assaults. I have served as a consultant to the U.S. Department of Justice, Human Rights Watch and Disability Rights. I am the author of Prison Madness: The Mental Health Crisis Behind Bars and What We Must Do About It (Jossey-Bass/Wiley, 1998) and Solitary: The Inside Story of Supermax Isolation and How We Can Abolish It (University of California Press, 2017),



CONFIDENTIAL

42

4. In addition to the pain and suffering, there are three psychiatric diagnoses that fit the clinical picture Mr. Vermillion presents: **#1. Depression.** His recurring depressive episodes and ongoing level of depression warrant a diagnosis of “dysthymic disorder” (DSM IV) or “persistent depressive disorder” (DSM V), and Major Depressive Disorder must also be considered; **#2. Sub-Syndromal Posttraumatic Stress Disorder (PTSD)**, in other words, there have clearly been multiple traumas in Mr. Vermillion’s life history, including massive abuse during childhood and a very traumatizing four plus year stint in solitary confinement, and Mr. Vermillion exhibits many of the symptoms of PTSD, but the condition is “sub-syndromal” to the extent that he may not exhibit sufficient symptoms to qualify for a clear diagnosis of PTSD; and **#3. the SHU Post-Release Syndrome.** Mr. Vermillion’s clinical presentation fits a combination of the three of these disorders. The second and third disorder are not listed in the Diagnostic and Statistical Manual (DSM IV or DSM V) but are terms that appear in the clinical and correctional literature. All three conditions have been caused by or greatly exacerbated by Mr. Vermillion’s four plus years in solitary confinement between 2009 and 2013.
5. There are three other diagnoses that warrant consideration: **Malingering, Anti-Social Personality Disorder and Substance Abuse Disorder.** It is always important to rule out **malingering** in the course of a forensic examination.³⁵ It is very clear to me, based on clinical techniques for identifying malingering, that Mr. Vermillion is not malingering, if anything he minimizes his psychiatric symptoms. His narrative is internally consistent, and is consistent with the documents related to this matter. And the kinds of symptoms and disabilities he reports to me are also entirely consistent with what I and other researchers have heard reported by many other denizens of solitary confinement units.

³⁵ Kupers, T. “Malingering in Correctional Settings,” Correctional Mental Health Report, 5, 6, 81-, March/April, 2004.

CONFIDENTIAL

43

6. Mr. Vermillion has been diagnosed in the past with **Antisocial Personality Disorder**. While that would be a relevant diagnosis for him as a much younger man, when he was abusing substances and getting in a lot of trouble, his current ability to plan ahead, work very hard and diligently on his legal pursuits, remain clean and sober and help other prisoners with their legal cases pretty much rules out the diagnosis of antisocial personality disorder. In other words, he has clearly outgrown the antisocial tendencies he evidenced as a much younger man.
7. Mr. Vermillion's mental health chart also contains the diagnosis **Poly-Substance Abuse**. Typically, in correctional settings, substance abuse is further described as "in remission in a controlled setting." In other words, if he was not in prison, this prisoner would be abusing substances in the community. This would be a mischaracterization in Mr. Vermillion's case because he has been clean and sober for very many years in jail and prison by an act of will, his resolution to avoid all alcohol and drugs even though they are readily available in the prisons where he has been incarcerated.
8. A significant part of the damage caused by Mr. Vermillion's stint in solitary confinement at Westville Correctional Facility is physical. There has been worsening of his chronic medical illnesses, and there have been changes in the structure and function of his brain that are known to accompany the psychological damage of solitary confinement.
9. The fact that Mr. Vermillion always believed and maintained that his sentence to solitary confinement was entirely unfair, actually made worse his pain and suffering as well as the psychological and physiological damage. He always felt it was very unfair that he was not provided notice and opportunity to plead his case, and then he was denied access to the program that would have made it possible for him to earn his way out of solitary confinement at Westville. In fact, there was a program in the WCU that would permit prisoners to earn their way out of solitary, and he tried to apply for it, but he was never permitted to take part

EXHIBIT 14

Heilbrun et al., "An MMPI-Based Empirical Model of Malingering and Deception"

An MMPI-Based Empirical Model of Malingering and Deception

Kirk Heilbrun, Ph.D., William S. Bennett, M.A.,
Adam J. White, M.A. and Jennifer Kelly, Ph.D.

The empirically-based model of malingering and deception described by Rogers (1984b) was operationalized using MMPI and clinical interview data. Subjects ($N = 159$) were patients committed to an inpatient forensic hospital as 'Incompetent to Stand Trial' or 'Not Guilty by Reason of Insanity'. Inter-rater reliability was acceptable for the 11 criteria used to define response styles. Each subject was categorized into one of five response style groups: 'Reliable', 'Malingering', 'Defensive', 'Irrelevant', or 'Unclassifiable'. Factor analysis of the rating criteria yielded four factors, three of which are comparable to the assigned groups, providing some validation for the constructs underlying these groups.

The assessment of patients' response style is important whenever there may be motivation to respond in a deceptive fashion. Rogers (1984b; 1988) has described four patterns of responding to clinical evaluation. These include: (1) *malingering*, a conscious fabrication or gross exaggeration of physical and/or psychological symptoms, distinguished from factitious disorder in that the motivation for malingering goes beyond the desire to assume the patient role and is understandable in light of the individual's circumstances; (2) *defensiveness*, a conscious denial or gross minimization of physical and/or psychological symptoms, distinguished from ego defenses, which involve intrapsychic processes that distort perception; (3) *irrelevance*, when one does not become engaged in the evaluation process; responses are not necessarily relevant to question content and may be random; and (4) *reliability/honesty*, when a sincere attempt is made to be accurate in responding, with factual inaccuracies attributable to poor understanding or misperception. A fifth pattern, hybrid responding, has also been noted (Rogers, 1984a; 1988) as referring to a combination of these response styles.

One of the most frequently employed instruments for assessing response style is the Minnesota Multiphasic Personality Inventory (Ziskin, 1981). The various MMPI scales and indexes that have been applied to assessing malingering and defen-

Kirk Heilbrun, Ph.D., is Chief Psychologist at the Forensic Service, Florida State Hospital. William S. Bennett, M.A., is a staff psychologist at the Forensic Service, Florida State Hospital, and a doctoral candidate in the Department of Psychology at Florida State University. Adam J. White, M.S., is a predoctoral intern at the Miami VA Medical Center and a doctoral candidate in the Department of Psychology at Florida State University. Jennifer Kelly, Ph.D., is Assistant Professor at the University of Texas Health Science Center at Houston. Please address correspondence and reprint requests to Dr Heilbrun at the Forensic Service, Florida State Hospital, Chattahoochee, Florida 32324, USA.

siveness have been described by Greene (1988). For malingering, these indices include the Wiener and Harmon Obvious and Subtle Scales (Wiener, 1948), the Dissimulation Scale (Gough, 1954), the Lachar and Wrobel critical items (Lachar & Wrobel, 1979), the F scale (Dahlstrom *et al.*, 1972), and the Gough F-K Dissimulation Index (Gough, 1950). Approaches to assessing defensiveness include the Obvious and Subtle scales, the Positive Malingering (Mp) Scale (Cofer *et al.*, 1949), the critical items, the L and K scales, and the F-K index.

The MMPI contains three specific approaches to detecting irrelevant responding. The F scale is one of the most reliable indicators of a random profile (Dahlstrom *et al.*, 1972; Rogers *et al.*, 1983). When a profile is random, a clinician should expect approximately half of the 64 F scale items to be endorsed. The other two indicators are the TR index (Buechley and Ball, 1952; Dahlstrom *et al.*, 1972; Greene, 1979) and the Carelessness (CLS) scale (Greene, 1978). The TR index is a measure of the number of the 16 repeated items that have been endorsed inconsistently. It has been shown to be a sensitive measure of irrelevant responding, since (in contrast to the F scale) it is relatively unaffected by psychopathology (Greene, 1988). The CLS consists of 12 pairs of items judged to be 'psychologically opposite' in content, and is a measure of the number of these pairs endorsed in the same direction. Like TR, CLS appears sensitive to irrelevant responding independent of psychopathology. Specific cut-off scores (such as greater than four for marginal, greater than five for clearly irrelevant) have been employed for both TR and CLS, and a combined cut-off (TR + CLS greater than eight for marginal, greater than nine for clearly irrelevant) has also been used (Greene, 1988).

The MMPI offers a variety of measures of malingering, defensiveness, and irrelevance in responding. It is also one of the most heavily researched instruments with respect to malingering and defensiveness (Ziskin, 1981). It therefore seems appropriate to include the MMPI in any empirical approach to measuring response style.

However, there are certain methodological problems with the previously-described MMPI studies which make it difficult to apply this research in a straightforward fashion. Much of the MMPI response style research has involved comparing the performance of individuals instructed to take the MMPI under 'fake good' or 'fake bad' conditions with that of individuals taking it under standard conditions. The problem with this approach is that there are likely to be significant differences between normal populations responding under deviant set instructions, and real defendants (for example) facing criminal charges. Some studies have compared experimental samples to clinical ones (for example Gough 1954; Grow *et al.*, 1980; Heaton *et al.*, 1978), but only a few have specifically examined MMPI validity scale patterns in forensic populations (for example Audubon & Kirwin, 1982; Lanyon & Lutz, 1984). Most previous MMPI research has also failed to control for such demographic differences as age, race, and sex (Butcher & Tellegen, 1978; Carlson, 1971).

Some MMPI research has used inmates or patients responding under circumstances that would suggest motivation to distort their presentation in a particular direction. In one study, three groups of maximum security federal inmates thought to be motivated to exaggerate psychopathology, deny it, or report it accurately, were administered the MMPI (Walters, 1988). The 'Exaggerating' group ($N = 35$, requesting single cell placement on the basis of psychopathology) scored significantly higher on F, F-K, D-O, Hy-O, Pd-O, Pa-O, Ma-O, the Dissimulation Scale, the total number of Obvious Items, and the Obvious to Subtle ratio, and significantly

lower on scales K, Hy-S, Ma-S, and the total number of Subtle items, when compared with inmates in 'Neutral' condition ($N = 30$, entering group therapy) and 'Denial' condition ($N = 32$, undergoing parole evaluation). Differences between Neutral and Denial groups were relatively small, with only D-O, Hy-O, and Ds producing statistically significant results. These results support the use of the MMPI in this population in assessing response style, particularly exaggeration.

In another study (Wasyliw *et al.*, 1988), the MMPI was administered to two groups of forensic patients: those being evaluated for competency to stand trial and/or sanity at the time of the offense ($N = 35$), and those who had already been acquitted by reason of insanity ($N = 39$). The investigators reasoned that the first group would stand to gain from the exaggeration of psychopathology, and thus predicted a higher incidence of malingering in this group. Their findings were consistent with this prediction. On four MMPI measures sensitive to malingering (F scale, Ds scale, Obvious and Subtle subscales, and F-K index), the 'evaluated' group was significantly higher on malingering than the 'acquitted' group.

The methodological difficulty with this approach, as Wasyliw *et al.* (1988) point out, is that such subject groups are by no means homogeneous with respect to response style. Some patients being evaluated for competency to stand trial may malingering, others may be defensive, others irrelevant, and yet others may respond reliably. A finding of a higher incidence of malingering in a sample with externally-induced motivation to exaggerate or fake psychopathology can provide one form of validation for the malingering measure, yet the sensitivity of the measure is adversely affected by the mixture of response styles in the group. An alternative methodology might involve obtaining more homogeneous groups of subjects using each response style and examining the differences between groups on a number of other variables relevant to response style.

What would be most useful in the classification of response style is a *model* composed of a number of empirical correlates rather than one (or several) scale(s) or index(es). Such a model has been proposed. Rogers (1984b), in a comprehensive review of the literature on response styles, described two models: (a) an 'heuristic' model, composed of clinical indicators of deception, and (b) an 'empirical' model, composed of factors obtained from a review of the research findings from psychological testing, case studies, and social psychology. The two share a number of elements, including severity of symptoms, consistency of self-report, rare symptoms, sequence of symptoms, obvious versus subtle symptoms, appearance of symptoms, and random pattern of responses. One other element is very similar: self-report inconsistent with clinical observation (heuristic model) versus symptoms inconsistent with clinical observation (empirical model). Each model also has some unique elements. The heuristic model includes the following elements that are not in the empirical model: endorsement of contradictory symptoms; memory of past psychological problems; potentially 'self-damaging' statements; and endorsement of highly specified symptoms. Elements in the empirical model not contained in the heuristic model include admission of common foibles; endorsement of idealistic self-attributes; improbable failure rate; latency of response; responses that are nearly correct; vagueness of responses; fidgetiness; willingness to discuss symptoms; and endorsement of 'stereotypes of neurosis'.

We have attempted to operationalize this empirical model of response style, using the MMPI and clinical interview data. To the extent that this can be accomplished,

the results will have clear implications for practice: both the clinical interview and the MMPI are routinely performed in psychological assessments.

In developing an MMPI-based classification approach, it is useful to consider the question posed by Megargee and Bohn (1979) regarding any MMPI-based model. Relevant questions include:

- (1) Do the MMPI profiles fall into distinct groups or clusters?
- (2) Is it possible for a clinician to sort individual MMPI profiles into such groups reliably?
- (3) Is it possible to define such groups operationally so that other clinicians, or even a computer, can sort individual MMPI profiles validly?

These questions will be addressed in the course of this study. In addition, the factor structure of the rating criteria will be examined to determine whether it suggests internally consistent constructs underlying the respective response styles.

METHOD

Subjects

The records of 159 patients were examined as part of this research. These patients had been randomly selected upon admission as a part of a larger project designed to collect clinical data for newly-admitted forensic patients at the Forensic Service, Florida State Hospital. The Forensic Service is a 460-bed forensic hospital located in Chattahoochee, which is a small town in northwest Florida. Patients are committed by the criminal courts as either 'Incompetent to Stand Trial' or 'Not Guilty by Reason of Insanity'. Also located on the hospital grounds are units for civilly-committed patients, a specialized treatment facility for mentally ill correctional inmates, and a minimum-security prison.

The sample consisted of male patients with a mean age of 30.8 years ($SD = 9.3$), with a range from 17 to 71. Racial composition was 57.2% White and 42.8% Black. Diagnosis was assessed through structured interviews: (a) the Schedule for Affective Disorders and Schizophrenia (Endicott and Spitzer, 1978) from September 1984 to October 1985; and (b) the Structured Clinical Interview for DSM-III (Spitzer and Williams, 1985) from October 1985 to June 1988. A total of 53% of patients were diagnosed as schizophrenic, with another 18.5% diagnosed as having affective disorder with psychotic features. Non-psychotic primary diagnoses were assigned to 23.9% of the sample. Secondary diagnoses of substance abuse (83.3%) were frequent.

Most subjects were committed as Incompetent to Stand Trial (78.6%) or Not Guilty by Reason of Insanity (18.2%). Arrest charges ranged from homicide to misdemeanor property offenses. Patients had a mean of 4.5 ($SD = 4.9$) prior hospitalizations in state psychiatric facilities.

Procedure

Two sources of information were used in rating response style. The MMPI, administered within three to five weeks of admission, was examined. Psychiatric evaluations conducted within 48 hours of admission were also reviewed.

The MMPI-Based Empirical Model of Response Style (MERS) was developed by using 11 of the 17 criteria described in Rogers's (1984b) empirical model of dissimulation. The criteria employed were severity of symptoms, consistency of self-report, rare symptoms, sequence of symptoms, obvious versus subtle symptoms, appearance of symptoms, admission of common foibles, endorsement of idealistic self-attributes, random responding, reported symptoms inconsistent with clinical observation, and willingness to discuss symptoms. These criteria were selected because seven of them were also included in Rogers's (1984b) heuristic model, and because these 11 factors appeared most likely to discriminate between reliable, malingering, defensive, and irrelevant responders.

Each criterion was operationalized by using information from the MMPI or the psychiatric evaluation. The MMPI was used to define consistency of self-report (TR, CLS Scales); obvious versus subtle symptoms (Obvious and Subtle Scales); admission of common foibles (L scale); endorsement of idealistic self-attributes (L scale, K scale, and the O-H Scale developed by Megargee *et al.* [1967]), and random responding (TR and CLS Scales and F scale). The psychiatric evaluation was used to rate symptom severity, rare symptoms, sequence of symptoms consistent with diagnosis, sudden onset of symptoms, reported symptoms inconsistent with clinical observation, and willing to discuss symptoms. Decision rules were established for each criterion to help determine whether the criterion should be rated as definitely present, possible, or definitely absent (see Appendix A).

Two of us (A.W and S.B.) served as raters. Using the 11 criteria, the raters scored each profile by adding the number of criteria which were met for each response style. Each subject thus received four scores, one each for Reliable, Malingering, Defensive, and Irrelevant. Raters then classified each subject into one of these four categories of response style, or as Unclassifiable.¹

RESULTS

The first question concerned the inter-rater reliability in classifying subjects. Agreement between raters appeared good ($r = 0.76$) in their scoring of each of the 11 criteria. Raters were also in substantial agreement ($r = 0.82$) with respect to the Reliable, Malingering, Defensive, and Irrelevant scores computed for each subject.

The next step was to examine the factor structure of the 11 criteria to see whether constructs similar to our response style categories would emerge. A principle-components analysis to oblique rotation (see Table 1) yielded four factors, which might be best described as Reliable, Malingering, General Deception, and Severity of Psychopathology. A total of 64.8% of variance was accounted for by these factors. The Reliable factor included two criteria seen in reliable responding: 'consistency of self report', and (negative to) 'random responding'. The second factor, Malingering, included a negative loading for the 'endorsement of idealistic self-attributes' and also included the tendency to respond to far more obvious than subtle items and to admit to common foibles. The third factor, General Deception, contained two items usually associated with malingering: 'rare symptoms' and 'sudden onset

¹ A full description of the MERS classification process, including cut-off scores, is available from the first author.

Table 1. Principle-Components Analysis of Items used in MMPI-Based Empirical Response Style (MERS) Procedure.

	Factor 1 (Reliable)	Factor 2 (Malingering)	Factor 3 (Deception)	Factor 4 (Severity of disturbance)
Consistency of self-report	0.97	0.02	-0.03	-0.05
Random responding	-0.96	0.05	0.03	0.04
Idealistic self-attributes	0.03	-0.91	-0.05	0.04
Admits common foibles	0.19	0.86	0.00	-0.06
Obvious > subtle	-0.27	0.64	-0.11	0.02
Reported symptoms inconsistent with clinical observations	-0.06	0.09	0.71	-0.27
Rare symptoms	0.17	0.06	0.64	0.44
Willingness to discuss symptoms	0.04	0.13	-0.62	-0.01
Sudden onset of symptoms	-0.23	-0.13	0.37	-0.10
Consistently severe symptoms reported	-0.06	0.11	0.12	0.77
Sequence of symptoms consistent with diagnosis	-0.02	-0.12	-0.28	0.75
Factor	Eigen-value	Percentage of variance	Cumulative percentage	
1	2.44	22.1	22.1	
2	1.95	17.7	39.9	
3	1.41	12.8	52.7	
4	1.33	12.1	64.8	

of symptoms'. It also included one criterion that could be applied to either Malingering or Defensive responding — 'inconsistency of reported symptoms with clinical observation', and another specific to Defensiveness — (negative on) 'willingness to discuss symptoms'. This factor was thus labelled 'general deception', as it contained aspects of deceptive responding in both directions. The final factor appeared to reflect severity of disturbance, relatively independent of response style.

DISCUSSION

The MERS procedure is a different approach to the assessment of response style in a forensic population. We have attempted, with some success, to operationalize the empirical model for the assessment of malingering and deception described by Rogers (1984b). Inter-rater reliability in rating the operational criteria is good, and the underlying factor structure is consistent with the existence of constructs underlying two of the response styles, and partially supportive of a third. Employing only MMPI and clinical interview data, the MERS may yield a more efficient use of such routinely collected clinical information.

It is noteworthy that the 'hybrid' groups described by Rogers (1984b, 1988), as observed clinically but not investigated empirically, are clearly part of the response styles seen in this study. The Reliable, Malingering, Defensive, and Irrelevant groups defined in this article are by no means 'pure'. There appear to be several reasons for this. A number of the criteria used in rating response style could be scored in the same direction for different groups, reflecting the clinical and empirical realities that these groups are similar in some ways. Also, response style is probably a dynamic variable, fluctuating according to situation (for example a male patient who exagger-

ates psychopathology when he is Incompetent to Stand Trial but under-reports it following his acquittal by reason of insanity) and even between topics within the same evaluation (reliably reporting thoughts and feelings in most areas, for example, but defensively minimizing sexual impulses toward children). The Malingering and Defensive groups in this study both had prominent elements of Reliable responding. This raises an interesting question: on what dimensions will these four groups differ? One eventual goal of research categorizing response style in this fashion is to provide 'templates' for each response style, allowing the evaluator to make a 'best fit' judgment between the empirically derived templates and the immediate pattern of clinical data.

The overall distribution of the groups is enlightening. The MERS procedure could not classify 52 (33%) of 159 subjects, and another 35 (22%) were classified as Reliable. That means that 46% of the sample were assigned to categories that should carry a warning label: 'Caution. What you see is *not* what you get!'. The consequent need for third-party information to help balance distortions created by inaccurate responding is particularly compelling. While third-party information has always been a recommended part of forensic assessment (Heilbrun, 1988; Melton *et al.*, 1987), the present results should serve to strongly underscore this need; as Greene (1988) points out, inaccurate responding means that little can be inferred about the nature of the psychopathology present. It is also worth noting that Defensive (21%) and Irrelevant (16%) response styles occurred more frequently than Malingering (9%), which has received far more attention in both the professional literature and popular press.

Research on the MERS process is just beginning. The next major step is external validation: do the response style groups differ significantly on non-MMPI variables, such as life-style, social history, behavior, and patterns of thinking? The classification procedure itself will need to be further investigated and refined. Nearly one-third of our present sample could not be classified. Further MERS research should reduce this percentage to a more acceptable level (such as 1% — the figure reported by Megargee and Bohn, 1979), and possibly lead to differential weighting of criteria. The procedures must be cross-validated as well. The question of generalizability to other populations should be addressed in samples of different age and sex, in out-patient settings as well as inpatient, and in jail and prison settings where the base rate of psychotic disorders is much lower. Finally, the interview might be more relevant to response style if it were designed specifically for that purpose. The use of such an interview, either structured (Rogers, 1986) or unstructured (Resnick, 1987), as a supplement to the standard clinical interview used in this study, might yield even more promising results.

ACKNOWLEDGEMENTS

The authors would like to express their appreciation to Richard Rogers for a thorough critique of an earlier version of this manuscript, and to Stewart Parsons and the staff of the Forensic Service, Florida State Hospital, for the support they have provided throughout the course of the larger project of which these data are a part.

REFERENCES

- Audubon, J. J. & Kirwin, B. R. (1982). Defensiveness in the criminally insane. *Journal of Personality Assessment*, *46*, 304–311.
- Buechley, R. & Ball, H. (1952). A new test of 'test validity' for the group MMPI. *Journal of Consulting Psychology*, *16*, 299–301.
- Butcher, J. N. & Tellegen, A. (1978). Common methodological problems in MMPI research. *Journal of Consulting and Clinical Psychology*, *46*, 620–628.
- Carlson, R. (1971). Where is the person in personality research? *Psychological Bulletin*, *75*, 203–219.
- Cofer, C. N., Chance, J. & Judson, A. J. (1949). A study of malingering on the MMPI. *Journal of Psychology*, *27*, 491–499.
- Dahlstrom, W. G., Welsh, G. S. & Dahlstrom, L. E. (1972). *An MMPI Handbook*. Vol. I.: *Clinical Interpretation* (rev. ed.). Minneapolis: University of Minnesota Press.
- Endicott, J. & Spitzer, R. (1978). A diagnostic interview: The Schedule for Affective Disorders and Schizophrenia. *Archives of General Psychiatry*, *35*, 837–844.
- Gough, H. G. (1950). The F minus K dissimulation index for the MMPI. *Journal of Consulting Psychology*, *14*, 408–413.
- Gough, H. G. (1954). Some common misconceptions about neuroticism. *Journal of Consulting Psychology*, *18*, 287–292.
- Greene, R. L. (1978). An empirically derived MMPI carelessness scale. *Journal of Clinical Psychology*, *34*, 407–410.
- Greene, R. L. (1979). Response consistency on the MMPI: The TR index. *Journal of Personality Assessment*, *43*, 69–71.
- Greene, R. L. (1988). Assessment of malingering and defensiveness by objective personality inventories. In R. Rogers (Ed.), *Clinical Assessment of Malingering and Deception* (pp. 123–158). New York: Guilford.
- Grow, R., McVaugh, W. & Eno, T. D. (1980). Faking and the MMPI. *Journal of Clinical Psychology*, *36*, 910–917.
- Heaton, R. K., Smith, H. H., Lehman, R. A. W. & Voyt, A. C. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology*, *46*, 892–900.
- Heilbrun, K. (1988, March). Third party information in forensic assessment: Much needed, sometimes collected, poorly guided. Paper presented at the Mid-Winter Meeting of the American Psychology-Law Society/Division 41, Miami, Florida.
- Lachar, D. & Wrobel, T. A. (1979). Validating clinicians' hunches: Construction of a new MMPI critical item set. *Journal of Consulting and Clinical Psychology*, *47*, 277–284.
- Lanyon, R. L. & Lutz, R. W. (1984). MMPI discrimination of defensive and nondefensive felony sex offenders. *Journal of Consulting and Clinical Psychology*, *52*, 841–843.
- Megargee, E. I. & Bohn, M. J. (1979). *Classifying Criminal Offenders: A New System Based on the MMPI*. Beverly Hills: Sage.
- Megargee, E. I., Cook, P. E. & Mendelsohn, G. A. (1967). Development and validation of an MMPI scale of assaultiveness in overcontrolled individuals. *Journal of Abnormal Psychology*, *72*, 519–528.
- Melton, G. B., Petrila, J., Poythress, N. G. & Slobogin, C. (1987). *Psychological evaluations for the courts: A handbook for mental health professionals and lawyers*. New York: Guilford.
- Resnick, P. (1987, October). The detection of malingered mental illness. Workshop presented at the American Academy of Psychiatry and Law, Ottawa, Canada.
- Rogers, R. (1984a). *Rogers Criminal Responsibility Assessment Scales (RCRAS) and Test Manual*. Odessa, Florida: Psychological Assessment Resources, Inc.
- Rogers, R. (1984b). Towards an empirical model of malingering and deception. *Behavioral Sciences & the Law*, *2*, 93–111.
- Rogers, R. (1986). *Structured Interview of Reported Symptoms (SIRS)*. Clarke Institute of Psychiatry: Toronto, Unpublished scale.
- Rogers, R. (1988). Introduction. In R. Rogers (Ed.), *Clinical Assessment of Malingering and Deception* (pp. 1–9). New York: Guilford Press.
- Rogers, R., Dolmetsch, R. & Cavanaugh, J. L. (1983). Identification of random responders on MMPI protocols. *Journal of Personality Assessment*, *47*, 364–368.
- Spitzer, R. & Williams, J. (1985). *Structured clinical interview for DSM-III*. New York: New York State Psychiatric Institute, Biometrics Research.
- Walters, G. D. (1988). Assessing dissimulation and denial on the MMPI in a sample of maximum security, male inmates. *Journal of Personality Assessment*, *52*, 465–474.
- Wasyliw, O. E., Grossman, L. S., Haywood T. W. & Cavanaugh, J. L. (1988). The detection of malingering in criminal forensic groups: MMPI validity scales. *Journal of Personality Assessment*, *52*, 321–333.
- Wiener, D. N. (1948). Subtle and obvious keys for the MMPI. *Journal of Consulting Psychology*, *12*, 164–170.

Ziskin, J. (1981). *Coping with Psychiatric and Psychological Testimony* (3rd edition). Venice, California: Law & Psychology Press.

APPENDIX A

Criteria for MERS Rating

1 = definitely no.	2 = uncertain.	3 = definitely yes.
Indicators	Source	Score
1. Symptoms consistently severe. (Is there a range of severity or is every symptom reported as severe?)	Psych	_____
2. Consistency of self-report across time. (TR or CLS $\geq 5 = 1$, TR or CLS $\leq 4 = 3$.)	MMPI	_____
3. Rare symptoms.	Psych	_____
4. Sequence of symptoms consistent with diagnosis.	Psych	_____
5. Far more obvious than subtle symptoms. (4 scales with T score differences 1-10 = 1, 11-19 = 2, 20+ = 3.)	MMPI	_____
6. Sudden onset of symptoms (Not due to substance abuse.)	Psych	_____
7. Admits to common foibles. (T score for Scale L 70+ = 1, 60-69 = 2, 40-59 = 3.)	MMPI	_____
8. Endorsement of idealistic self-attributes. (T scores for Scales L, K are 70+ = 1, 60-69 = 2, 40-59 = 1.)	MMPI	_____
9. Random pattern of responses on MMPI. (TR or CLS $\geq 5 = 3$, TR and CLS $\leq 4 = 1$.)	MMPI	_____
10. Reported symptoms inconsistent with clinical observations.	Psych	_____
11. Willingness to discuss symptoms.	Psych	_____

Note: 'Psych' refers to psychiatric evaluations, TR to the MMPI test-retest index, and CLS to the MMPI Carelessness Scale.

EXHIBIT 15

Melton, et al.,
Psychological
Evaluations for the
Courts

MON
KF
8922
.M45
2018

Psychological Evaluations for the Courts

**A Handbook for Mental Health
Professionals and Lawyers**

FOURTH EDITION

Gary B. Melton
John Petrila
Norman G. Poythress
Christopher Slobogin
Randy K. Otto
Douglas Mossman
Lois O. Condie



THE GUILFORD PRESS
New York London

VERMILLION 005137

threshold for suspecting dissimulation with a cautious stance about reaching conclusions on that issue.

(a) General Strategies for Detecting Feigning of Symptoms

A number of strategies are available for systematically investigating response style. The most common and venerable method is the clinical interview, usually consisting of a mental status examination or other relatively unstructured interview procedure. The cues and indicators of exaggeration or fabrication of mental symptoms that experienced interviewers look for have developed over decades of clinical experience⁷⁸ and involve comparing scientific knowledge about the typical presentation, symptoms, and course of various mental disorders with examinees' reports about their problems and their behavior during the evaluation.⁷⁹ A problem with this approach is that it is difficult to quantify how accurately it works, and accuracy varies substantially among clinicians. For this reason, forensic examiners increasingly use structured interviews designed to detect particular symptoms and other tools specifically designed to detect exaggeration or fabrication. A third technique is the judicious interpretation of traditional psychological tests, including particular scales on those tests, to detect dissimulation.

Finally, having independent, reliable third-party observations is one of the most useful means of assessing response style. Indeed, information that contradicts an examinee's descriptions of problems may well be the best ways of demonstrating exaggeration, fabrication, or denial, and may be the only viable one with examinees who are highly invested in sabotaging interview and testing efforts. In particular, forensic clinicians may use evidence from interviewing, testing, or other sources (e.g., documents or surveillance data that is incongruent with the examinees' self-report or presentation) to confront examinees with questions about their response style and approach to the assessment process. Little is lost if an examinee does not confess to dissimulation, but much is gained if the examinee admits to misrepresenting symptoms or problems. Whatever the limitations of the interview or testing data that gave rise to the concerns, the examinee's admission is a compelling piece of clinical information that

brings closure to the issue. If such a confrontation is handled tactfully, the clinician may now have a relieved, cooperative, and candid respondent with whom to discuss remaining issues in the forensic referral.

While the use of third-party information to detect dissimulation probably needs no further elaboration, below we discuss in somewhat more detail the use of the other three approaches (structured interviews, instruments designed to detect fabrication, and traditional psychological tests), in connection with discerning feigned psychosis, intellectual deficits, depression, anxiety, and guardedness or minimization.

(b) Assessment of Feigned Psychosis

Rogers identified several indicators of exaggeration or fabrication of serious mental disorder that might be detected during an interview.⁸⁰ These include the following:

- *An overplayed and dramatic presentation*, including such features as theatrical style, eagerness to discuss symptoms or impairments, reports of extreme symptom severity, and indiscriminate endorsement of symptoms.
- *Deliberateness and carefulness*, including such features as slower rate of speech, extensive use of qualifiers, more hesitations, and repeating of interviewer questions.
- *Inconsistency with psychiatric diagnosis or clinical functioning*, including report of rapid onset and resolution of symptoms, and report of rare symptoms or unusual symptom combinations.
- *Inconsistency of self-report*, including report of contradictory symptoms and disparity between reported and observed symptomatology and functioning.
- *Claims of obvious, blatant symptoms*, including reports of hallucinations and delusional ideas in the absence of altered thought patterns and "negative" symptoms (such as paucity of speech and impoverished affect).

Experienced forensic clinicians develop their own interview questions to look for these kinds of cues. For example, persons who appear all too ready to claim or acknowledge a wide range of positive or dramatic symptoms may be enticed by inquiries

about exotic symptoms or ludicrous symptom combinations, such as “When you hear voices, do you also experience slight dizziness and the sudden smell of hamburger?” As another example, Schacter suggested that defendants who claim amnesia for a particular event be asked whether they feel they could remember the incident with sufficient prodding; his research suggests that simulators are more likely to deny a “feeling of knowing” than those who truly cannot remember.⁸¹ When examinees fall into one or more of these interviewing traps laid by the clinician, a diagnosis of malingering may follow.

However, this approach to detecting feigned impairment has limited utility. Because the specific questions examiners employ vary, clinicians have few to no systematically gathered data (e.g., explicit norms) to facilitate an objective interpretation of examinees’ responses. A clinician’s basis for concluding that an examinee is malingering may merely be “clinical experience,” which is subject to a number of biases in judgment and memory. Research indicates that the accuracy of clinical judgments of malingering is not impressive.⁸² Increasingly, mental health professionals have concluded that because interview-based approaches to detecting malingering are of such limited utility, “employment of instruments specifically designed for this purpose should be considered the standard of practice whenever there is a basis for suspecting overreporting of symptoms.”⁸³

In this regard, one psychological assessment instrument that has received considerable praise is the Structured Interview of Reported Symptoms (SIRS), published originally in 1992 and now in its second edition (SIRS-2).⁸⁴ Both versions conduct the test administrator through a 172-item structured interview that yields scores on several scales, each representing different strategies to detect feigned symptoms of serious mental disorders.⁸⁵ The primary scales contain items that permit more systematic exploration of the various malingering strategies found in Rogers’s survey of the literature (summarized above). For example, one scale (RS) records an examinee’s endorsement of rare symptoms that do not often occur even in truly psychotic individuals; another scale (IA) captures the reporting of improbable or absurd symptoms; yet another scale (SU) is sensitive to the overreporting of everyday problems that are not indicative of a major psychiatric disorder.

Rogers and his colleagues have investigated the capacity of the SIRS to detect malingering in a series of studies that have included so-called “simulation” and “known-group” designs.⁸⁶ The original studies conducted to validate the SIRS showed that its scores could detect known⁸⁷ or simulated⁸⁸ malingerers without mislabeling “honest” responders as feigners. Rogers and colleagues also published studies showing that the SIRS did well even when the feigning subjects received information about the detection strategies employed in the test⁸⁹ or were psychologically knowledgeable about specific disorders to be feigned.⁹⁰

The original research on the SIRS has also consistently reported respectable indices of sensitivity⁹¹ and specificity,⁹² although investigators who completed a meta-analysis of research on the SIRS expressed concerns about lower specificity rates in later research.⁹³ In response to concerns about its specificity, Rogers and his colleagues revised the SIRS scoring and classification rubrics via publication of the SIRS-2 (while not changing the instrument’s item content). However, at least two publications raise concerns about the accuracy of the revised scoring system and information presented in the SIRS-2 manual.⁹⁴

Despite these and other concerns about its research base,⁹⁵ the SIRS/SIRS-2 deserves serious consideration by forensic clinicians who want to evaluate the genuineness of reported symptoms of severe mental illness. The structured nature and objective scoring of the SIRS make it superior to clinical assessment of response styles. While other screening measures show some promise,⁹⁶ the SIRS remains the most praised and best-validated measure with respect to this type of assessment.

Some commentators have suggested that projective tests such as the Rorschach Inkblot Technique and the TAT [see Table 3.3], because of their more ambiguous nature, show some potential in identifying persons feigning psychotic-spectrum symptoms (as well as symptoms of other disorders). Research, however, indicates that these claims are not well founded,⁹⁷ a conclusion that, in conjunction with the ongoing debate about the validity and value of projective approaches,⁹⁸ raises serious questions about the use of such measures in forensic evaluation contexts.

Among traditional measures of psychopathology and behavior, the instrument most extensively

I. GENERAL CONSIDERATIONS

§ 3.06

60

researched and widely used in the assessment of dissimulation is the MMPI/MMPI-2/MMPI-2-RF [see Table 3.3].⁹⁹ Potential indices of dissimulation on the MMPI/MMPI-2/MMPI-2-RF include several response style scales and indices (L, F, K; F-K index; Fb, Fp, F-r, VRIN, TRIN, Ds, Ds-R, Md, Mp, FBS, RBS). Greene has cautioned against the overinterpretation of these indicators, warning that “[i]t is not necessary to use several of these scales/indexes simultaneously since they are correlated highly and consequently are very redundant.”¹⁰⁰ With this caveat, the MMPI/MMPI-2/MMPI-2-RF appears to have the most empirical support among conventional psychological tests for use in the assessment of malingered psychopathology.

Finally, note must be made of two other psychological tests are sometimes suggested as means of detecting dissimulation of serious mental illness. The PAI [again, see Table 3.3], like the MMPI-2, is a structured self-report measure of psychopathology and behavioral functioning.¹⁰¹ Like the MMPI-2, the PAI has standard response style scales as well as specially designed response style indices. Although the classification utility of the PAI scales has not been investigated as closely as that of their MMPI-2 counterparts, research to date is promising.¹⁰² In contrast, the response style scales of the MCMI-IV and its predecessors (the MCMI-II and -III) have not been examined in much research, and the research that has been conducted has been discouraging.¹⁰³

(c) Assessment of Feigned Intellectual Impairment

One might assume that detection of feigned cognitive impairment would rely primarily on intelligence tests. But a review by Schretlen revealed that much of the research on the usefulness of such tests for this purpose is dated and focused on instruments not currently in wide use.¹⁰⁴ Furthermore, Schretlen found that even the more popular contemporary intelligence tests [see Table 3.3] are not well adapted to the kind of “scatter analysis” that has been the best indicator of malingering.¹⁰⁵ Thus Schretlen's review offers little support for the use of intelligence tests in assessing malingered cognitive impairment. Similar in conclusion is Rogers and colleagues' review of several studies in which skilled clinicians have attempted to detect malingering of cognitive

deficits by using intelligence tests, or a combination of those tests with neuropsychological tests and other clinical protocols. In most of these studies as well, clinicians performed poorly (even below chance in some cases).¹⁰⁶

Sweet has identified six circumstances that should raise questions about response styles of examinees undergoing neuropsychological evaluations in forensic contexts;¹⁰⁷ many of these are analogous to the strategies identified by Rogers in connection with assessing the validity of claims of thought disorder. Examiners should be alert for (1) insufficient effort on CAIs of intelligence, memory, and other cognitive abilities (i.e., an examinee obtains a score on a measure that is below what would be expected even of impaired persons); (2) insufficient effort on FRIs that were designed to assess response style in forensic contexts (i.e., the examinee obtains a score on a measure that is below chance, or is below what would be expected even of impaired persons); (3) “nonsensical,” rare, or unique responses not typically associated with known neuropsychological or cognitive impairment (e.g., the examinee reports limitations or experiences that could not result from the reported or claimed injury); (4) inconsistent responding across tests that assess related aspects of cognitive functioning (e.g., the examinee obtains divergent scores on highly correlated measure of memory, intelligence, or other cognitive functioning); (5) observed inconsistency between test performance and real-life adaptive behaviors (e.g., test scores on measures of intellect, problem solving, or memory indicate very low functioning, yet the examinee exercises higher capabilities in day-to-day life tasks); and (6) inconsistencies in how deficits affect different areas of life (e.g., the examinee reports intellectual, memory, or cognitive impairments that limit his or her ability to work but have no effect on enjoyable activities, such as traveling and engaging in recreational activities). Although several of these strategies have shown promise in isolated studies, few are systematically used in commonly employed measures for detecting feigned cognitive deficits.

In the past 30 years, however, several measures have been expressly designed to assess examinees' response styles and effort during assessments of neuropsychological and cognitive functioning. These measures include the Test of Memory Malingered (TOMM), the Victoria Symptom

VERMILLION 005141

Validity Test, the Validity Indicator Profile (VIP), and the Word Memory Test (WMT)—can be administered in conjunction with and compared to the results of CAIs (e.g., measures of intelligence or memory). These specialized measures typically rely on two approaches with respect to assessing an examinee's response style. First, the instruments employ a two-alternative forced-choice design that allows identification of "worse than chance" performance, which is strongly indicative of efforts to appear more impaired than is actually the case. In addition, these tests contain scales that allow evaluators to compare the performance of examinees with the performance of persons who have conditions that are known to produce cognitive impairment, but who have no motivation to exaggerate their impairment. Test scores significantly below those obtained by impaired persons are indicative of attempts to appear more impaired than is actually the case.

(d) Assessment of Feigned Anxiety and Depression

Assessment of the response styles of persons reporting symptoms of anxiety and depression is challenging. Few tests that assess symptoms of anxiety and depression have embedded within them measures of response style, and research indicates that those that do (such as the MMPI-2/MMPI-2-RF and PAI) are not particularly good at identifying exaggerated or fabricated symptoms of anxiety and depression.¹⁰⁸ These findings may be due in part to the fact that such test takers have probably experienced some actual anxiety or depression in their lives, which is often untrue of examinees who want to feign symptoms of psychosis or cognitive impairment.¹⁰⁹ Not surprisingly, research suggests that feigners are more likely to be successful when they rely on past experience in simulating symptoms.¹¹⁰

(e) Assessment of Denial, Disavowal, Guardedness, and Minimization

Assessment of examinee guardedness and defensiveness has received considerably less attention than assessment of feigned psychopathology and cognitive impairment. Although mental health publications largely agree in their depictions and descriptions of feigned mental problems, the professional

literature offers no such consensus on the characteristics, behaviors, or styles of examinees who are guarded, minimizing, or defensive in their presentations. Similarly, in comparison to the large number of free-standing measures of feigned psychiatric symptoms and cognitive impairment, few measures of guardedness and defensiveness have been developed, and existing measures have not been subjected to significant validation research.¹¹¹

Omnibus measures of psychopathology and personality (e.g., the MMPI-2/MMPI-2-RF, PAI, and MCMI-II/MCMI-III/MCMI-IV) have a number of response scales embedded within them that are designed to identify examinees who deny or minimize psychopathology or present themselves as having positive attributes. But these are of limited utility. The L and S scales of the MMPI-2 and MMPI-2-RF are designed to identify persons motivated to portray themselves as having positive attributes, and a variety of other scales and indices (e.g., K, F-K, O-S, Mp, Sd) are thought to be of some value in identifying persons motivated to deny and minimize problem behaviors or experiences. Yet available research recommends that evaluators use different cutoff scores for the scales with different populations. In his review of MMPI-2 measures of response style, Greene also noted research finding low intercorrelations between the scales that measure defensiveness and guardedness, and suggested that they thus may measure different phenomena.¹¹²

Even less research has examined the performance of the PAI's scales and indices of guardedness/defensiveness, and the results are only modestly encouraging.¹¹³ Finally, the small amount of research examining the classification of the response style scales of the MCMI-II/MCMI-III/MCMI-IV offers little ground for optimism.¹¹⁴

(f) Summary

In conclusion, clinicians should be cautious about their use of psychological tests to assess response style, and very cautious about drawing conclusions about response style based on their clinical judgment. With respect to omnibus measures of psychopathology and personality, use of the MMPI-2/MMPI-2-RF and PAI appears warranted in some circumstances. In contrast, the limited and discouraging research examining the utility of the response style scales of the MCMI-II/MCMI-III/

MCMI-IV indicates that this measure is of little value for detecting feigning. Measures specifically aimed at assessing response style (e.g., the SIRS/SIRS-2, TOMM, WMT, and VIP) are more likely to be appropriate, but only if their use is tailored to the referral question. For example, the SIRS/SIRS-2 may have little value in a forensic examination where the reported complaints revolve around intellectual and memory impairment rather than symptoms of severe mental illness, whereas the TOMM, VIP, or WMT may be very helpful. Conversely, questions regarding the forensic examinee's report of psychotic-spectrum symptoms may properly trigger use of the SIRS/SIRS-2 or M-FAST (see Table 3.3), whereas the TOMM, VIP, and WMT may well be useless in this circumstance.

Many experienced clinicians advocate the use of a battery of tests of different types as an approach to overcoming the limitations associated with any one test. This approach is theoretically sound, and some research supports its potential value.¹¹⁵ However, these studies used a combination of experimental and conventional testing indices, and relied on strictly statistical (as opposed to clinical) interpretations to attain their results. To our knowledge, no study shows that clinicians, using clinical interpretations of combinations of tests, have demonstrated extraordinary ability to identify people feigning impairment. Moreover, examiners should be cautious about using highly correlated scales or measures of response style, because doing so might increase the examiners' confidence while actually decreasing classification accuracy.¹¹⁶ Yet researchers continue to look for potential combinations of indices or scores from multiple measures of response style that would improve identification of feigned impairment.¹¹⁷

Other caveats are also important. First, because they do not provide information about *why* an examinee might be feigning symptoms, even the best measures of symptom exaggeration/fabrication are not measures of "malingering"—a *descriptor* (not a diagnosis or condition) that, according to DSM-5, reflects efforts to feign illness or impairment to achieve some external goal (e.g., avoid criminal prosecution, obtain disability benefits). Second, given the limitations of existing assessment approaches, examiners should be cautious in their descriptions of examinees' response styles and recognize the potential for error. Third, examiners

should not make the fundamental error of assuming that evidence of symptom exaggeration indicates the absence of genuine impairment, because some persons who exaggerate and fabricate difficulties also have real problems. Relatedly, one cannot conclude that all accounts offered by a person who is less than candid in his or her response style are inaccurate in some way. Conversely, an examiner should not assume that simply because an examinee has a well-documented record of symptoms or impairment, all of the problems reported by the examinee during a forensic examination are genuine.

3.07. Challenges to the Basis of Expert Testimony

As this chapter has explained, mental health professionals should base their expert testimony on multiple sources of information. Suppose an expert bases an opinion about insanity on a clinical interview of the defendant, a diary that police seized from the defendant's home, psychological test data, interviews of the defendant's family members, arrest reports, medical records, and the results of a polygraph test or narcoanalysis. It may be that, given common practice in the jurisdiction, the other side will not object to these various bases for the testimony. However, if in doubt on this score, the wise attorney should seek, prior to trial, either a stipulation from opposing counsel or a judicial determination that the opinion and its foundations are admissible. In practice, perhaps only opinions based on polygraph testing will give rise to serious challenge, but the law provides grounds for attacking the admissibility of any of these sources of information and the opinions based on the sources.

The legal bases for such challenges are diverse. The results of a clinical interview might be tainted if the defendant's Fifth Amendment rights were not honored [see § 4.02]. The diary might be "suppressed" if it was seized illegally [see §§ 2.04(a)(1), 7.02]. Psychological testing the expert employed might be considered irrelevant [see § 3.03(a)]. The family interviews and the police and medical records might be challenged as inadmissible "hearsay" [see below]. And the polygraph and narcoanalysis results could be challenged either under the *Frye* test (which bars testimony unless the reliability of the polygraph or narcoanalysis is sufficiently established to have gained

EXHIBIT 16

BCCLA v. Canada (2018)
(excerpts)

IN THE SUPREME COURT OF BRITISH COLUMBIA

Citation: *British Columbia Civil Liberties Association
v. Canada (Attorney General)*,
2018 BCSC 62

Date: 20180117
Docket: S150415
Registry: Vancouver

Between:

**British Columbia Civil Liberties Association and
the John Howard Society of Canada**

Plaintiffs

And

Attorney General of Canada

Defendant

And

**West Coast Women's Legal Education and Action Fund and
Criminal Defence Advocacy Society**

Intervenors

Before: The Honourable Mr. Justice Leask

Reasons for Judgment

Counsel for the Plaintiffs:

J. Arvay, Q.C.
A. Latimer
C. DiPuma

Counsel for the Defendant:

M.R. Taylor, Q.C.
S. Currie
E. Burnet
M.J. Huculak
G.J. Stark
F. Paradis
V. Wynn-Williams
M.L. French
M. Iyengar, Articled Student
A. Azevedo, Articled Student
A. Kirby, Articled Student

Counsel for the Intervenor,
West Coast Women's Legal Education and
Action Fund:

J. Winteringham
R. Mangat
S. Arrandale

Counsel for the Intervenor,
Criminal Defence Advocacy Society:

T. Arbogast
A. Ballantyne

Place and Dates of Hearing:

Vancouver, B.C.
July 4-7, 10-14, 17-19, 24-28, 31,
August 1-4, 9-11, 14, 16-18, 21, 24,
28-31, and September 1, 2017

Place and Date of Judgment:

Vancouver, B.C.
January 17, 2018

Table of Contents

I. INTRODUCTION	6
A. The Parties	7
B. Nature Of The Evidence	7
C. The Office of the Correctional Investigator’s Reports	8
II. HISTORY OF SOLITARY CONFINEMENT	8
A. History	8
B. International Law	19
III. ADMINISTRATIVE SEGREGATION	21
A. Correctional Service of Canada: Organization and Statistics	21
B. The Statute, Regulations, and Commissioner’s Directives	25
C. Placement in Administrative Segregation	28
D. Alternatives to Administrative Segregation	32
E. Conditions of Confinement	34
1. Physical Conditions	34
2. Cell Effects	36
3. Human Contact	37
4. Programming	42
(a) <i>General</i>	42
(b) <i>Aboriginal Programming</i>	43
5. Duration of Placements	45
F. Effects of Segregation on Mental Health	47
1. Plaintiffs’ Experts	48
(a) <i>Dr. Stuart Grassian</i>	48
(i) <i>Qualifications</i>	48
(ii) <i>Opinion</i>	48
(b) <i>Dr. Craig Haney</i>	53
(i) <i>Qualifications</i>	53
(ii) <i>Opinion</i>	53
2. Government Experts	58
(a) <i>Dr. Jeremy Mills</i>	58
(i) <i>Qualifications</i>	58
(ii) <i>Opinion</i>	59
(b) <i>Dr. Paul Gendreau</i>	64

(i) Qualifications	64
(ii) Opinion	65
3. Discussion	67
4. Conclusion	74
IV. SECTION 7	76
A. Interests	77
1. Life	77
2. Security of the Person	80
(a) Segregation Causes Serious Psychological Suffering	80
(b) Mental Health Monitoring	83
(c) Segregation Causes Physical Harm	88
B. Principles of Fundamental Justice	88
1. Legislative Objective	89
2. Arbitrariness and Overbreadth	90
3. Gross Disproportionality	95
C. Procedural Fairness	96
1. The Law	96
2. Positions of the Parties	97
3. Discussion	97
D. External Oversight	100
1. Background	100
2. Discussion	108
E. Right to Counsel	115
V. SECTION 10	117
VI. SECTION 15	121
A. Positions of the Parties	121
B. Law	123
C. Analysis	125
1. Women	125
2. Aboriginal Inmates	127
3. Inmates with Mental Illness	134
4. Intersecting Characteristics	141
VII. SECTION 12	141
VIII. SECTION 9	143
IX. SECTION 1	145

A. Governing Legal Principles	146
B. Proportionality – Prolonged, Indeterminate Segregation	147
1. Rational Connection	147
2. Minimal Impairment	148
(a) <i>Time Limits</i>	148
(b) <i>Voluntary Dissociation</i>	151
(c) <i>Other Subpopulations</i>	153
(d) <i>Treatment Approach</i>	157
3. Proportionality of Effects	158
C. Proportionality: Lack of Independent Review and Right to Counsel	159
X. REMEDIES	159
XI. COSTS	162
XII. END NOTE	162

I. INTRODUCTION

[1] Segregation has been described as “the most onerous and depriving experience that the state can legally administer in Canada”: *Annual Report of the Office of the Correctional Investigator 2014-2015* at p. 31 [2014-2015 *Annual Report*]. The British Columbia Civil Liberties Association and the John Howard Society of Canada ask the Court to end administrative segregation as it is presently practised in federal penitentiaries in Canada.

[2] The plaintiffs contend that ss. 31-33 and 37 of the *Corrections and Conditional Release Act*, S.C. 1992, c. 20 (the “CCRA”), authorizing administrative segregation are contrary to ss. 7, 9, 10, 12 and 15 of the *Canadian Charter of Rights and Freedoms*. They say that the impugned provisions permit indeterminate and prolonged solitary confinement, as that term is understood in international law and accepted worldwide by virtually every organization or professional group conversant with the issue. Segregation, especially when endured for extended periods, has significant adverse effects on the physical, psychological, and social health of inmates; there is no independent oversight of placements in what has been described by the Supreme Court of Canada as a “prison within a prison”: *Martineau v. Matsqui Disciplinary Bd.*, [1980] 1 S.C.R. 602 at 622. The plaintiffs further allege that the impugned provisions have a disproportionate impact on Aboriginal inmates and those with mental illness.

[3] The Attorney General of Canada (the “Government”) responds that administrative segregation as it is practised in federal correctional facilities is not solitary confinement since inmates have daily opportunity for meaningful human contact. Moreover, the psychological effects of segregation on inmates remain the subject of ongoing and vigorous scientific debate. The Government submits that maintaining institutional security and inmate and staff safety is a complicated task, and that administrative segregation is a necessary tool when no other reasonable alternatives exist. The length of placements is not indeterminate as alleged but, rather, determined by the time required to eliminate the safety or security issue that triggered its use. Accordingly, the Government contends that the plaintiffs have

failed to establish that the impugned provisions are unconstitutional on their face or in their application, and that their claims must therefore be dismissed.

A. The Parties

[4] The British Columbia Civil Liberties Association is a non-profit advocacy group whose objects include the promotion and defence of civil liberties in British Columbia and Canada. The Association has had long-standing interest in inmates' rights and policy.

[5] The John Howard Society of Canada is a non-profit organization with a strong history of commitment to and involvement in matters of criminal justice, especially as they pertain to penal policy and corrections.

[6] The Government does not dispute that the plaintiffs are entitled to public interest standing in this case. However, it contends that the lack of an individual plaintiff has implications for the available remedies.

[7] Two intervenors also participated in these proceedings. West Coast LEAF's submissions focussed on what it says is the disproportionate impact of administrative segregation on individuals with intersecting characteristics of disadvantage, namely, Aboriginal women with mental illness. The Criminal Defence Advocacy Society ("CDAS") primarily challenged the lack of access to counsel during the segregation review process.

B. Nature Of The Evidence

[8] Although the parties were required to operate under very tight timelines, they nevertheless assembled a substantial evidentiary record.

[9] Twenty-eight witnesses were cross-examined on their affidavits before the Court. The plaintiffs' witnesses comprised 10 experts on a range of subject matters relating to the practice and effects of administrative segregation, and eight lay witnesses. These latter individuals were primarily former Correctional Service of Canada ("CSC") employees and inmates who had experienced placements in administrative segregation.

[10] The Government tendered the evidence of two experts with respect to the psychological effects of administrative segregation, and 10 lay witnesses who were all current or former CSC employees.

[11] Because of the volume of evidence, I will not refer to every affidavit or witness in these Reasons. However, I have reviewed the entire record, and my conclusions are based on all of the evidence before me.

C. The Office of the Correctional Investigator's Reports

[12] The Office of the Correctional Investigator (“OCI”) serves as ombudsman for federally sentenced inmates. Among its statutory responsibilities are the investigation of individual or systemic concerns relating to corrections, and the preparation of annual and special reports. To enable performance of these duties, the OCI has full access to all of CSC’s facilities, records and staff.

[13] The plaintiffs seek the admissibility of numerous OCI reports pursuant to either the public records exception to the rule against hearsay or the principled approach. The Government objects to their admissibility on both grounds.

[14] I refer to OCI reports throughout these Reasons. For the most part, the particular facts or statistics I cite were put to the Government’s witnesses in cross-examination and accepted by them as accurate. Where this was not the case, I am satisfied the reports are nonetheless admissible pursuant to the principled approach to hearsay. They are necessary because the Correctional Investigator is not a competent or compellable witness pursuant to s. 189 of the *CCRA*. They are also reliable because they are compiled by the OCI in the discharge of a public duty on the basis of data maintained by CSC.

II. HISTORY OF SOLITARY CONFINEMENT

A. History

[15] The following history derives primarily from the expert reports of Dr. Stuart Grassian, whose qualifications are discussed later, and Michael Jackson, Q.C. Professor Jackson is Emeritus Professor of Law at the University of British

[252] I accept that the early history of solitary confinement in the United States and more particularly in Germany, demonstrates that these harmful effects have been recognized since the late 19th and early 20th centuries.

[253] Dr. Mills and Dr. Gendreau base their opinions to a significant extent on the Morgan et al. Study which, in turn, relies for its results on the Colorado Study and the Zinger Study of Canadian inmates. I agree with Dr. Grassian and Dr. Haney's criticisms of the Colorado Study and Dr. Haney's criticisms of the Zinger Study. In particular, the distinction that Dr. Haney draws between voluntary and involuntary confinement in administrative segregation will be discussed later in these Reasons.

[254] None of my above findings should be read as a criticism of meta-analysis as a method of scientific research. I agree with Dr. Gendreau that it can be a valuable tool for understanding complex problems. The Morgan et al. Study is unhelpful in understanding solitary confinement because of flaws in the Colorado and Zinger studies.

[255] I now turn to the legal significance of the facts that I have found.

IV. SECTION 7

[256] Section 7 of the *Charter* guarantees everyone "the right to life, liberty and security of the person and the right not to be deprived thereof except in accordance with the principles of fundamental justice".

[257] To establish a breach of s. 7, the plaintiffs must show that (a) the impugned laws interfere with, or deprive them of, their life, liberty or security of the person; and (b) that the deprivation in question is not in accordance with the principles of fundamental justice.

[258] The inquiry under s. 7 is not a quantitative one – for instance, how many people are negatively impacted – but qualitative. Accordingly, an arbitrary, overbroad or grossly disproportionate impact on one person suffices to establish a breach of s. 7: *Canada (Attorney General) v. Bedford*, 2013 SCC 72 at para. 127.

2. The impugned laws are invalid pursuant to s. 15 of the *Charter*:
 - a) to the extent that the impugned laws authorize and effect any period of administrative segregation for the mentally ill and/or disabled; and
 - b) also to the extent that the impugned laws authorize and effect a procedure that results in discrimination against Aboriginal inmates.

[610] I am prepared to grant a 12 month suspension of my declaration of invalidity on the basis of the first two reasons enumerated in *Schachter v. Canada*, [1992] 2 S.C.R. 679, that is that an immediate declaration would pose a potential danger to the public or threaten the rule of law.

XI. COSTS

[611] I am satisfied that this is a proper case to award the plaintiffs' special costs and am pleased that counsel have worked out a mechanism to accomplish this.

XII. END NOTE

[612] I want to express my appreciation for the hard work and legal abilities of all counsel who have been involved in this matter. The issues were complex, the time pressures formidable, yet the professionalism of counsel substantially lightened the burden on the Court.

"Leask J."

EXHIBIT 17

The Colorado Study

2010

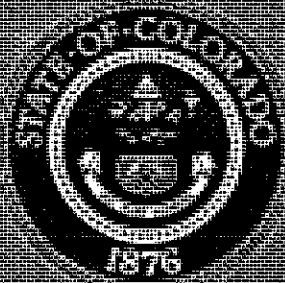
One Year Longitudinal Study of the Psychological Effects of Administrative Segregation

By:

Maureen L. O'Keefe, M.A.
Colorado Department of Corrections
2862 South Circle Drive
Colorado Springs, CO 80906
(719) 726-4364

Kelli L. Kiebe, Ph.D.
Alysha Stucker, B.A.
Kristin Sturm, B.A.
William Leggett, B.A.
University of Colorado – Colorado Springs
Department of Psychology
1420 Austin Bluffs Parkway
Colorado Springs, CO 80918
(719) 255-4175

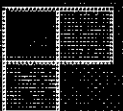
Submission Date to FBI: October 31, 2010



United States
Department of
Justice

Department of
Justice
Federal Bureau of
Investigation

This project was supported by Grant No. 2006-IJ-CJ-0025 awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. Points of view in this document are those of the author and do not necessarily represent the official position or policies of the U.S. Department of Justice.



This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

ACKNOWLEDGEMENTS

The research team would like to gratefully acknowledge the following individuals who served on the project's advisory board for their significant time commitment and invaluable insights into this project:

Jamie Fellner, Esq.
Human Rights Watch

Jeffrey Metzner, M.D.
University of Colorado – Denver

Joel Dvoskin, Ph.D.
University of Arizona

Larry Reid, M.A.
Susan Jones, M.C.J.
Joanie Shoemaker, M.B.A.

Jim Michaud, Ph.D.
Elizabeth Hogan, M.D.
Colorado Department of Corrections

John Stoner, Ph.D.
Colorado Department of Corrections, Retired

The researchers would also like to acknowledge the support of Mr. Aristedes Zavaras and the staff at the Department of Corrections. This project would never have been possible without the commitment and contributions of management, mental health, case management, offender services, and security staff.

ABSTRACT

One of the most widely debated topics in the field of corrections – the use of long-term administrative segregation (AS) – has suffered from a lack of empirical research. Critics have argued that the conditions of AS confinement exacerbate symptoms of mental illness and create mental illness where none previously existed. Empirical research has had little to offer this debate; the scant empirical research conducted to date suffers from research bias and serious methodological flaws. This study seeks to advance the literature in this regard.

This study tested three hypotheses: (1) offenders in AS would develop an array of psychological symptoms consistent with the security housing unit (SHU) syndrome, (2) offenders with and without mental illness would deteriorate over time in AS, but at a rate more rapid and extreme for the mentally ill, and (3) inmates in AS would experience greater psychological deterioration over time than the comparison groups.

Study participants included male inmates who were placed in AS and comparison inmates in the general population (GP). Placement into AS or GP conditions occurred as a function of routine prison operations. GP comparison participants included those at risk of AS placement due to their institutional behavior. Inmates in both of these study conditions (AS, GP) were divided into two groups – inmates with mental illness (MI) and with no mental illness (NMI). A third comparison group of inmates with severe mental health problems placed in San Carlos Correctional Facility, a psychiatric care prison facility, was also included. A total of 302 inmates were approached to participate in the study, and 55 refused to participate or later withdrew their consent. Participants were tested at 3-month intervals over a yearlong period.

Standardized test data were collected through self-report, correctional staff and clinical staff measures. Tests with demonstrated reliability and validity were selected to assess the eight primary constructs of interest: (1) anxiety, (2) cognitive impairment, (3) depression-hopelessness, (4) hostility-anger control, (5) hypersensitivity, (6) psychosis, (7) somatization, and (8) withdrawal-alienation. Extensive analyses of psychometric properties revealed that inmates self-reported psychological and cognitive symptoms in remarkably reliable and valid ways.

The results of this study were largely inconsistent with our hypotheses and the bulk of literature that indicates AS is extremely detrimental to inmates with and without mental illness. Similar to other research, our study found that segregated offenders were elevated on multiple psychological and cognitive measures when compared to normative adult samples. However, elevations were present among the comparison groups too, suggesting that high degrees of psychological disturbances are not unique to the AS environment. In examining change over time patterns, there was initial improvement in psychological well-being across all study groups, with the bulk of the improvements occurring between the first and second testing periods, followed by relative stability for the remainder of the study. Patterns indicated that the MI groups tended to be similar to one another but were significantly elevated compared to the NMI groups, regardless of their setting. Contrary to our hypothesis, offenders with mental illness did not deteriorate over time in AS at a rate more rapid and more extreme than for those without mental illness. Finally, although AS inmates in this study were found to possess traits believed to be associated with long-term segregation, these features cannot be attributed to AS confinement because they were present at the time of placement and also occurred in the comparison study groups. Implications for policy and future research are discussed.

CONTENTS

<p>Executive Summary..... iv</p> <p> Introduction iv</p> <p> Purpose of Present Study..... iv</p> <p> Method v</p> <p> Findings viii</p> <p> Policy Implications ix</p> <p>Introduction 1</p> <p> Characteristics of Long-Term Segregation 1</p> <p> Criticisms of the A5 Model 2</p> <p> Case Law Review 3</p> <p> Research Review 5</p> <p> The Colorado System 7</p> <p> Purpose of Present Study..... 15</p> <p>Method 17</p> <p> Group Assignment 17</p> <p> Participants 19</p> <p> Materials 19</p> <p> Procedures 28</p> <p>Results 31</p> <p> Data Analysis Plan 31</p> <p> Sampling..... 32</p> <p> Validity of Responses 36</p>	<p> Official Record Data 40</p> <p> Group Comparisons 45</p> <p> Normative Comparisons 46</p> <p> Change over Time 52</p> <p> Predictor Analyses..... 75</p> <p>Discussion..... 78</p> <p> Limitations..... 79</p> <p> Future Research 80</p> <p> Policy Implications 81</p> <p>References 84</p> <p>Appendix A 91</p> <p>Appendix B 95</p> <p> Summary of Study Measures 95</p> <p> Description of Individual Measures 96</p> <p> Normative Comparisons 125</p> <p> Composite Scores..... 126</p> <p> Summary 138</p> <p> Appendix B References 139</p> <p>Appendix C 145</p> <p> Prison Symptom Inventory Analyses 145</p> <p> Summary 150</p>
--	--

EXECUTIVE SUMMARY

INTRODUCTION

One of the most widely debated topics in the field of corrections – the use of long-term administrative segregation (AS) – has suffered from a lack of empirical research. The placement of offenders in AS environments, particularly those with serious mental illness, has been a point of contention. Critics have argued that the conditions of AS confinement exacerbate symptoms of mental illness and create mental illness where none previously existed. The use of AS across the country has persisted as a corrections management tool despite litigation, although in many states, the placement of mentally ill into AS is no longer permitted. Empirical research has had little to offer this debate; the scant empirical research conducted to date suffers from research bias and serious methodological flaws.

Now decades after the deinstitutionalization of states' mental health hospitals, corrections agencies have seen a surge of offenders with serious mental illness in their prisons. The rate of serious mental illness in the community is 6% (National Institute of Mental Health, 2010). Among the incarcerated, the rate of serious mental illness is tripled at about 18% (Ditton, 1999; O'Keefe & Schnell, 2008). A similar phenomenon is occurring within prisons, whereby a disproportionate rate of mentally ill are found within AS, estimated to be 50% higher than the rate within the general prison population (O'Keefe, 2008a). It is not known the extent to which this difference is caused by the AS environment. Researchers have been unable to settle the question of whether these high rates of mental illness are caused by AS relative to the general prison population or whether there is a selection bias such that offenders with mental illness, unable to adapt to general prison settings, are placed in AS at higher rates. This study seeks to advance the literature in this regard.

PURPOSE OF PRESENT STUDY

The broad purpose of the project was to evaluate the psychological effects of long-term segregation on offenders, particularly those with mental illness. This study examined conditions as they existed in the Colorado prison system with respect to AS, using the Colorado State Penitentiary (CSP) as the AS study facility. Only males were included because females represent 2% of Colorado's AS population. We did not assign inmates to segregation, but studied those conditions as they naturally occurred. The following are the primary goals and hypotheses.

Goal 1: To determine which, if any, psychological domains are affected, and in which direction, by the different prison environments. A multitude of psychological dimensions were examined, drawing from those most often cited in the literature. The broad constructs of interest were depression/hopelessness, anxiety, psychosis, withdrawal and alienation, hostility and anger control, somatization, hypersensitivity, and cognitive impairment. We hypothesized that offenders in segregation would develop an array of psychological symptoms consistent with the security housing unit (SHU) syndrome, with elevations across the eight constructs.

Goal 2: To assess whether offenders with mental illness decompensate differentially from those without mental illness. We were particularly interested in whether long-term segregation had a differential impact based on the presence of mental illness in offenders. We sought answers to the following questions: Does AS exacerbate symptoms in offenders with mental illness? Does AS create symptoms of mental illness in those who did not exhibit any at placement? It was hypothesized that offenders with and without mental illness would deteriorate over time, but the rate at which it occurred would be more rapid and more extreme for the mentally ill.

Goal 3: To compare the impact of long-term segregation against the general prison setting and a psychiatric care prison. In this study, the psychological and behavioral symptoms of offenders in AS were compared to similar offenders who were sent to San Carlos Correctional Facility (SCCF) or returned to the general prison population pursuant an AS hearing. This study used a repeated measures design over the course of a year to explore whether psychological distress was attributable to the various prison environments. It was hypothesized that inmates in segregation would experience greater psychological deterioration over time than the comparison groups.

This study also included an examination of individual characteristics such as mental health status, personality, and trauma history to determine if certain factors could predict patterns of change. Prediction analyses were exploratory in nature and we did not formulate a hypothesis about the variables that might predict differential rates of psychological decompensation.

METHOD

Group Assignment

Study participants included male inmates who were placed in AS and comparison inmates in the general population (GP). Placement into AS or GP conditions occurred as a function of routine prison operations, pending the outcome of their AS hearing, without involvement of the researchers. All study participants classified to AS were waitlisted for and placed in CSP. Inmates who returned to GP following an AS hearing were assumed to be as similar as possible to AS inmates and, therefore, comprised the comparison groups. Comparison participants also included inmates targeted for a diversionary program that identified inmates at high risk of AS placement due to their disruptive behavior. This program discontinued shortly after the study commenced, hence few participants were identified through this method. Inmates in both of these study conditions (AS, GP) were divided into two groups – inmates with mental illness (MI) and with no mental illness (NMI). There are fewer inmates with mental illness than without, but because both subgroups were of equal interest to this study, separate groups enabled over-selection of inmates with mental illness.

A third comparison group was included. This group included inmates with severe mental health problems placed in SCCF. Of the inmates placed in SCCF, only those with patterns of prison misbehavior, as measured by disciplinary violations, were included in the study. The purpose of the SCCF comparison group was to study inmates with serious mental illness and behavioral problems who were managed in a psychiatric prison setting.

Participants

A total of 302 male inmates were approached to participate in the study. Thirty refused to participate. Two more offenders were considered a passive refusal and were removed for inappropriate sexual behavior towards the researcher during the first testing session. An additional 23 offenders later withdrew their consent, although the data collected to the point of their withdrawal was used. In addition to refusals and withdrawals, 10 inmates released prior to the end of the study due to discretionary releases by the Parole Board and one participant death.

Five testing sessions were initially established at 3-month intervals, beginning with the date of consent and initial administration. Therefore, tests were scheduled at 3 months, 6 months, 9 months and 12 months after the baseline assessment. However, this schedule was problematic for the AS groups. When the study

began, there was a 3-month average wait for inmates to be transferred to CSP due to a shortage of AS beds. While on the waitlist, AS inmates were held in a punitive segregation bed at their originating facility. It was determined that the primary goal was to study inmates in a single long-term segregation facility (CSP) to limit confounding variables and that therefore the baseline measure should be collected upon placement into CSP. However, it was also recognized that significant changes could occur while inmates were held in segregation at their originating facility. Therefore, a “pre-baseline” measure was collected as close to the AS hearing as possible, which meant that the CSP groups completed six test intervals rather than five. The time between the pre-baseline and baseline measure varied according to how long the inmate was on the waitlist. The median time between pre and baseline tests was 99 days, although eight offenders were moved into CSP so quickly that they did not have a pre-baseline measure. In the analyses, tests were aligned across groups according to the test number, such that the CSP groups had an additional test at the end rather than at the beginning.

Participants’ ages ranged from 17 to 59 at the time of consent, with a mean age of 31.8 ($SD = 9.1$). The racial/ethnic breakdown of participants was 40% white, 36% Hispanic, 19% African American, 4% Native American, and 1% Asian. Of the inmates with mental illness who were included in this study, 56% were identified with a serious and pervasive disorder.

Materials

Assessment tools were selected to comprehensively cover the variety of psychological constructs associated with AS (e.g., Arrigo & Bullock, 2008; Grassian, 1983; Haney, 2003). The primary constructs assessed in this study were as follows: (1) anxiety, (2) cognitive impairment, (3) depression/hopelessness, (4) hostility/anger control, (5) hypersensitivity, (6) psychosis, (7) somatization, and (8) withdrawal/alienation. Additionally, malingering, self-harm, trauma, and personality disorders were assessed.

Research materials were selected to meet the following criteria: (1) use of assessments with demonstrated reliability and validity, (2) use of multiple sources for providing information (e.g., self-report, clinician ratings, files), (3) use of multiple assessments of each construct of interest, (4) ability to use within the prison setting, and (5) ease of administration, including no specialized equipment, no physical contact, length of time, and appropriate reading level.

The 12 self-report instruments used in this study were: (1) Beck Hopelessness Scale, (2) Brief Symptom Inventory, (3) Coolidge Correctional Inventory, (4) Deliberate Self-Harm Inventory, (5) Personality Assessment Screener, (6) Prison Symptom Inventory, (7) Profile of Mood States, (8) Saint Louis University Mental Status, (9) State-Trait Anxiety Inventory, (10) Structured Inventory of Malingered Symptomatology, (11) Trail Making Test, and (12) Trauma Symptom Inventory.

In addition to self-report assessments, ratings of psychological functioning were obtained from clinical staff and ratings of behavior in the housing unit were obtained from correctional staff. The Brief Psychiatric Rating Scale (BPRS) was completed by clinical staff and the Prison Behavior Rating Scale (PBRs) was completed by correctional staff.

Most assessments were collected at each testing period, although personality disorders, self-harm, and trauma history were not. It was determined that personality and trauma history were relatively stable constructs that needed to be assessed only once to limit the testing burden on study participants. Also, due to

the burden on already limited mental health resources, the BPRS was only administered at the first, third, and fifth testing intervals.

Data from official records were collected primarily from the Department of Corrections Information System, which is an administrative database of offender data. Offender characteristics to include demographic history, criminal history and offense data, institutional behavior, and needs levels were electronically downloaded.

Certain data elements were collected only for study participants during the course of their participation in the study. The following were collected and coded for the period of time between each testing interval for each participant: the amount of time spent in various settings (e.g., segregation, GP, hospital), phone records, and mental health crisis data. Additionally, activity logs from paper files for the CSP participants were collected and coded.

Procedure

Study enrollment began July 2007 and ended March 2009, with final testing of all participants completed in March 2010. The project operated under the approval of the institutional review board at the University of Colorado at Colorado Springs.

The research team was notified of AS hearings by the case management supervisor at each facility and of SCCF placements by the clinician who scheduled the facility transfers. Notification typically occurred before the hearings or SCCF placement to give the field researcher maximal lead time. Researchers reviewed electronic records to screen inmates for study eligibility.

The field researcher was a female university employee who completed the full training academy and had a badge that permitted her unescorted access to the facilities. In advance of each visit, the researcher contacted prison security to arrange visits with specific inmates. All inmates were escorted by security staff to the visiting room, which contained a noncontact booth for inmates in AS or punitive segregation conditions. The researcher met individually with each inmate to review the consent form, which included the general purpose of the study, voluntary nature of participation, risks and benefits, and remuneration. Inmates were advised that the purpose of the study was to learn about adjustment to prison and offenders in prisons across the state were participating in this study.

At the time of consent, the initial test battery was administered. The field researcher instructed participants to read the directions for each test. Instructions were highlighted by researchers when there was an indication on the test to respond with respect to a certain timeframe (e.g., in the past week). The researcher administered the timed tests, and she assisted if they had questions, most frequently with the definition of a word. The researcher collected the test packet immediately following its completion, so it was not ever handled by security staff.

The field researcher distributed the PBRs to housing staff at each testing interval and collected the completed forms upon return visits to the facility. Mental health clinicians were generally notified that a BPRS was needed a couple weeks prior to the researcher testing to give them time to complete the assessment.

Participants' data were kept in two separate databases. The eligibility database tracked the eligible pool of offenders, such as identifying information, current location, date of AS hearing or SCCF placement, expected

release date, mental health status and clinician approval, selection into study or reason for exclusion, and date of consent or refusal. A testing schedule for study participants was incorporated into the database, which also had reporting capabilities in order to manage the project. A separate database tracked participants' responses to the standardized tests; no identifying information was included in this database other than a secure researcher-assigned identification number. Both databases were stored on a secured server with access restricted to project researchers.

FINDINGS

The results of this study were largely inconsistent with our hypotheses and the bulk of literature that indicates AS is extremely detrimental to inmates with and without mental illness. We hypothesized that inmates in segregation would experience greater psychological deterioration over time than comparison inmates, who were comprised of similar offenders confined in non-segregation prisons. Consistent with other research, our study found that segregated offenders were elevated on multiple psychological and cognitive measures when compared to normative adult samples (Haney, 2003; Suedfeld, Ramirez, Deaton, & Baker-Brown, 1982). However, there were elevations among the comparison groups too, suggesting that high degrees of psychological disturbances are not unique to the AS environment. The GP NMI group was the only one that was similar to the normative group on a number of scales.

In examining change over time patterns, there was initial improvement in psychological well-being across all study groups, with the bulk of the improvements occurring between the first and second testing periods, followed by relative stability for the remainder of the study. On only one measure – withdrawal – did offenders worsen over time, but this finding was only true for the two NMI groups, so it is not attributable to AS. Even given the improvements that occurred within the study timeframe, the elevations in psychological and cognitive functioning that were evident at the start of the study remained present at the end of the study.

Another hypothesis was that offenders with mental illness would deteriorate over time in AS at a rate more rapid and more extreme than for those without mental illness. Patterns indicated that the MI groups (CSP MI, GP MI, SCCF) tended to look similar to one another but were significantly elevated compared to the NMI groups (CSP NMI, GP NMI), regardless of their setting. For the AS offenders, the MI group scored worse than the NMI group on all self-report measures except the Trails test and all staff measures except the PBRs Anti-Authority scale. In addition to the changes over time described above, PBRs scores decreased significantly for segregated inmates regardless of their mental health status, which would be an indicator that staff may be perceiving improvements, but the significant differences were from the first to the second assessment periods when the majority of participants changed facilities, which suggests this is perhaps a measurement error rather than a true improvement. As hypothesized there was a differential time effect for the MI and NMI groups on several composite measures (i.e., anxiety, hostility-anger control, hypersensitivity, somatization), but the interactions were in the opposite direction of our hypothesis; on average, the CSP NMI group did not change while the CSP MI group improved.

We stated that offenders in segregation would develop an array of psychological symptoms consistent with the SHU syndrome. As already discussed, all of the study groups, with the exception of the GP NMI group, showed symptoms that were associated with the SHU syndrome. These elevations were present from the start and were more serious for the mentally ill than non-mentally ill. In classifying people as improving, de-

clining, or staying the same over time, the majority remained the same. There was a small percentage (7%) who worsened and a larger proportion (20%) who improved. Therefore, this study cannot attribute the presence of SHU symptoms to confinement in AS. The features of the SHU syndrome appear to describe the most disturbed offenders in prison, regardless of where they are housed. In fact, the group of offenders who were placed in a psychiatric care facility (SCCF) had the greatest degree of psychological disturbances and the greatest amount of negative change.

Finally, in this study, we conducted some exploratory predictive analyses to determine if there were individual characteristics that could identify who may be at greater risk of psychological harm from segregation. There were no individual predictors that showed strong effects for predicting change. This could indicate that we did not have the correct predictors or that patterns of decompensation are individualized (i.e., not predictable), but it is more likely that the relative stability over time makes it difficult to predict change.

A review of the findings warrants a discussion of plausible alternative explanations that might account for our results. The use of a repeated measures design enabled us to determine that change was occurring and in which direction. Even given the debate about whether or not harmful effects resulted from AS, it was never suggested that inmates might improve as this study found. The presence of comparison groups avoids an attribution error; the changes, improvements in this case (i.e., 20%), are not due to segregation. These conclusions replicate those drawn by Zinger and colleagues (2001) where there was a similar lack of evidence of harm. These studies suffered criticism for high refusal rates, high attrition rates, small sample sizes, and short durations – limitations that were corrected in the present study (note, however, that no generalizations should be made beyond the 1 year follow-up period in this study). Furthermore, the use of reliable and valid standardized measures enabled the present research study to assess psychological functioning in an objective manner. Although the majority of these tests were not normed for prisoner populations, the current reliability and validity findings increased our confidence in these measures.

POLICY IMPLICATIONS

Does this study legitimize the use of segregation with offenders, including those with serious and persistent mental illness? Because this study may not generalize to other prison systems, especially those that have conditions of confinement dissimilar to CSP, it is not possible to conclude that AS is not detrimental for all offenders. Systems that are more restrictive and have fewer treatment and programming resources should not generalize these findings to their prisons. Replication is needed to understand how increased services, privileges, and out of cell time ameliorate the unintended consequences of AS, and research needs to inform prison systems about the standards and practices necessary to protect inmates in segregation from harmful psychological effects.

It is also important to note that there may be other negative consequences of AS that we did not study. For example, Lovell, Johnson, and Cain (2007) found that inmates released directly from segregation to the streets had dramatically higher rates and severity of detected recidivism than AS inmates who first released to GP (but see Mears & Bales, 2009). We also did not study the degree to which AS met its purported goal of changing inmate behavior for the better over time. The only questions addressed by this study were related to psychological changes over time in segregation. Thus, we make no empirical or value judgments about whether and to what degree the use of AS balances the benefits (e.g., a safer prison system) with costs (e.g., significant reductions in freedom).

It is impossible to ignore the extremely disproportionate rate at which inmates with serious mental illness are assigned to AS (Lovell, 2008; Metzner & Fellner, 2010; O'Keefe, 2008a), which has to some degree "shocked the conscience" of the courts (see *Jones 'El v. Berge*, 2001; *Madrid v. Gomez*, 1995; *Ruiz v. Johnson*, 1999). In an era when prisons are expected to implement evidence-based practices and to rehabilitate offenders who will be releasing back to the community, is it enough to avoid harm? Must we ask ourselves another question: what are the conditions required to *improve* inmates' mental well-being while in segregation? Prison systems are held to a standard of treatment that is at least equivalent to community standards. It is likely that this most difficult segment of society has failed at all levels of community treatment and earlier criminal justice interventions, but the quest to treat and improve services for the most needy is an important reality facing corrections agencies.

Regarding their psychological functioning and levels of distress, these data suggest, although the differences were small, that inmates with serious mental illness are less likely to improve in segregation *and* are less likely to get worse compared to mentally ill inmates in GP. We do not assume that the reasons for these apparently contradictory findings are the same. For example, it is possible that fewer inmates with mental illness get worse because segregation is a safer and more structured environment. On the other hand, hypotheses regarding their unlikeliness to improve include the significant limitations that segregation places on various types of therapeutic activities and services such as group therapy. Further, the data do not tell us which aspects of AS prevent psychological improvement and deterioration, respectively, among inmates with mental illness. However, since prisons have a constitutional duty to respond to serious medical (including psychiatric) needs, the possibility that segregation may prevent improvement is cause for concern and further study.

There remain significant implications for mental health staff who work in prison systems that permit the placement of mentally ill in long-term segregation. It is critical for mental health staff to screen and assess offenders prior to AS placement to determine their vulnerability to harm that might occur as a result of their segregation. While in segregation, it is important that the mental status of all offenders be assessed on a frequent, regular basis through rounds and individual sessions. Prison systems need to have a range of confinement options, such that offenders who are at risk of or are showing signs of decompensation can be removed from segregation and placed in an alternative high security environment that permits greater out of cell time and interaction with others.

Other systems have rejected confinement models that isolated offenders and held them in extremely restrictive spaces. Even if the segregation models of the early 1900's and the state psychiatric hospitals of the mid-19th century are viewed as "primitive" compared to modern-day AS facilities, it is important to examine and understand why these models failed and were ultimately dismantled. Although there are a number of researchers who predict that there is no end in sight to the supermax model (King, 1999; Mears, 2008; Pizarro & Narag, 2008; Pizarro & Stenius, 2004), they have also raised empirical questions regarding their efficacy. Questions about the efficacy of AS will be asked until more is known about whether the use of AS in prison systems improves conditions for the rest of the system, whether and how they improve inmate behavior within and beyond the prison walls, whether they are cost-effective, whether they increase risks to public safety, and whether there are settings or individuals that are prone to psychological deterioration.

INTRODUCTION

One of the most widely debated topics in the field of corrections – the use of long-term administrative segregation (AS) – has suffered from a lack of empirical research. The placement of offenders in AS environments, particularly those with serious mental illness, has been a point of contention. Critics have argued that the conditions of AS confinement exacerbate symptoms of mental illness and create mental illness where none previously existed. The use of AS across the country has persisted as a corrections management tool despite litigation, although in many states, the placement of mentally ill into AS is no longer permitted. Empirical research has had little to offer this debate; the scant empirical research conducted to date suffers from research bias and serious methodological flaws.

Now decades after the deinstitutionalization of states' mental health hospitals, corrections agencies have seen a surge of offenders with serious mental illness in their prisons. The rate of serious mental illness in the community is 6% (National Institute of Mental Health, 2010). Among the incarcerated, the rate of serious mental illness is tripled at about 18% (Ditton, 1999; O'Keefe & Schnell, 2008). A similar phenomenon is occurring within prisons, whereby a disproportionate rate of mentally ill are found within AS, estimated to be 50% higher than the rate within the general prison population (O'Keefe, 2008a). It is not known the extent to which this difference is caused by the AS environment. Researchers have been unable to settle the question of whether these high rates of mental illness are caused by AS relative to the general prison population or whether there is a selection bias such that offenders with mental illness, unable to adapt to general prison settings, are placed in AS at higher rates. This study seeks to advance the literature in this regard.

CHARACTERISTICS OF LONG-TERM SEGREGATION

"Supermax" is the popular term used to describe the technologically advanced, supermaximum security prisons designed for single-cell occupancy that were rapidly being constructed across the nation during the 1990's. Even when new construction was not possible, existing prisons were retrofitted to conform to this new model. Therefore, a supermax *facility* may refer to an entire facility or a distinct unit within a facility (National Institute of Corrections, 1997). Although there was a virtual explosion of supermax facilities over the past two decades, similar units have operated on a smaller scale for decades (Zinger, Wichman, & Andrews, 2001).

The modern-day supermax model is traced back to the U.S. Penitentiary in Marion, Illinois, that went into permanent lockdown status in 1983. Prior to Marion, the Federal Bureau of Prisons operated solitary confinement at the Alcatraz Island Prison until it closed in 1963. History points to even earlier uses of solitary confinement including Pennsylvania's Eastern State Penitentiary, which opened in 1829 and was later modeled in European prisons (Smith, 2008). However, these early models featured such extreme social isolation and sensory deprivation (Cohen, 2008) and were so primitive that there is little comparison between them and today's modern supermaxes (National Institute of Corrections, 1999).

Across prison systems, different terms are used to describe the same concept: administrative segregation or AS, control units, security housing units or SHUs, and security controls unit (Haney, 2003; NIC, 1999). In Colorado, it is known as AS. Just as the names vary, so do the conditions. However, the defining feature that is frequently associated with this model is single-cell confinement for 23 hr per day, with 1 hr allowed out of cell for showers and exercise. AS is differentiated from *punitive or disciplinary segregation*, which is a time-

limited punishment enforced for a prison violation pursuant a full due process hearing; placement in AS is an administrative decision that often extends for an indefinite time period.

AS prisons are costly to build and operate due to costs associated with high security that include single-occupancy cells, high staff to inmate ratios, and technology (Mears & Bales, 2009). Because inmate movement requires multiple restraints and staff, many services are provided at the cell door, including meals, library, mental health services, and programs. Newer AS facilities are equipped with advanced technological equipment, which enables delivery of even more services to inmates in their cells (e.g., visitation through videoconferencing) or within the facility (e.g., medical and dental procedures). Although technological equipment is designed and used to reduce security breaches, it also increases the degree of isolation experienced by inmates.

It is difficult to establish the number of inmates held in AS nationally. In 1999, King estimated that 1.8% of all state prisoners were housed in AS. Although prevalence estimates are higher now than in 1999, prison systems under-report the actual use of AS, likely due to the negative connotation associated with the supermax label used in national reporting (Naday, Freilich, & Mellow, 2008). For example, the Federal Bureau of Prisons reported no inmates in AS, protective custody, or supermax beds in 2008 (American Correctional Association, 2009), which is inaccurate. Additionally, states reported drastically different numbers of offenders in AS from year to year (see Naday et al., 2008). Given these limitations, it is estimated that at least 3.2% of all state prisoners in 2008 were housed in AS or protective custody (American Correctional Association, 2009), although this appears to be an under-estimation of the true prevalence rates.

CRITICISMS OF THE AS MODEL

The use of AS has sparked a controversy resulting in considerable criticism of the prison system and its administrators. The limited number of research studies and the inadequacies of existing research on AS have only fueled the controversy. Numerous researchers and forensic professionals have called for more research to examine whether evidence based practices are in place and to examine whether harm is being done by confining inmates to segregation (Kurki & Morris, 2001; Lovell et al., 2007; Mears, 2008; Metzner & Dvoskin, 2006; Pizarro & Narag, 2008), but the topics and setting are difficult ones in which to conduct research (Mears & Watson, 2006; Naday et al., 2008).

One criticism has been the lack of evidence that segregation has achieved its intended goal of reducing violence in the prison system (Kurki & Morris, 2001; Mears, 2008). There is some literature to suggest that wardens and prison systems find this model to be effective in reducing violence and increasing order within the larger prison system (Atherton, 2001; Mears & Watson, 2006; Ward & Werlich, 2003). However, these studies lack the appropriate statistical controls to assert that the improvements are measurable and attributable to AS rather than merely perceptions of wardens or the result of other management controls also put into place at the same time. In an empirical study of institutional violence in three states, Briggs, Sundt, and Castellano (2003) did not find that AS reduced inmate-on-inmate violence. However, in a follow-up study, Sundt, Castellano and Briggs (2008) found that permanent reductions in inmate-on-staff violence were attributable to the opening of an AS prison in Illinois.

Corrections departments have been moving towards evidence-based models and practices to improve the rehabilitation opportunities for offenders. These practices include standardized assessments, matching offender needs to services, cognitive-behavioral programs, re-entry services, structured decision making

guidelines, and intensive treatment programs such as therapeutic communities. By endorsing an administrative action to determine placement of inmates into AS, corrections agencies have moved away from the evidence-based risk and classification instruments in favor of more subjective decisions. This criticism applies to both the decision to place inmates in AS and their continuation in AS (Human Rights Watch, 2000; O’Keefe, 2008b; Pizarro & Narag, 2008).

An emerging concern is the return of offenders from AS to society, which may occur with little or no step-down process such that offenders are released directly to the streets from 23/7 confinement. Although the adjustment required for offenders to adapt to rapid and extreme socialization changes is of concern, the issue of public safety is perhaps of even greater concern. Research has indicated that AS inmates have higher recidivism rates than non-AS offenders (Mears & Bales, 2009; Motiuk & Blanchette, 2001; O’Keefe, 2005), but this is likely due to the selection effects of who is confined to AS. When matching procedures were engaged, no differences in overall recidivism rates were found between AS and matched non-AS inmates (Lovell et al., 2007; Mears & Bales, 2009). Mears and Bales (2009) found a small, but significant difference when violent recidivism was the outcome measure rather than general recidivism; 24.2% of AS inmates had a violent re-offense compared to 20.5% of matched non-AS inmates. Lovell et al. (2007) found that inmates who released directly from AS had a higher recidivism rate than matched offenders who transitioned from AS to a lower security facility prior to release. In contrast, Mears and Bales (2009) found neither a recency effect (i.e., amount of time that elapsed between AS confinement and release) nor an exposure effect (i.e., total amount of time spent in AS confinement) on recidivism rates.

Human rights concerns are tantamount to a discussion of the criticisms of the AS model. The use of AS has been called a human rights violation, and some have even labeled it torture (Gawande, 2009; Metzner & Fellner, 2010). Many find the conditions of solitary confinement to be excessively harsh and inhumane (Cohen, 2008; Haney, 2003, 2008; Human Rights Watch, 1997, 1999, 2000; King, 1999; Kupers, 2008; Kurki & Morris, 2001; Toch, 2001). Specifically, the lack of treatment, programs, and activities to engage the mind; the restricted personal contact; lack of control over light and sound; lack of windows; and little or no access to the outdoors are considered to be more extreme than is required for the safe operation of prisons. Additionally, when people are held in highly restrictive environments where they have little control over their life, there is a greater opportunity for staff to inflict abuses upon those confined within (Haney, 2008; Human Rights Watch, 2000; Kurki & Morris, 2001).

The most significant issue is the question of whether prisoners are able to psychologically adapt to the conditions of AS. There is concern that mentally healthy individuals will decompensate in segregation, but recent discussions have centered on the placement of offenders with mental illness in such environments. Because the harmful effects of AS is the central focus of this study, we will examine the evidence as it is available both in case law and in the research literature.

CASE LAW REVIEW

As is the case with many important issues that affect the correctional system, conditions of AS confinement have been challenged in U.S. courts. In a pivotal First Amendment case heard in the Supreme Court, *Turner v. Safley* (1987) set a standard for lower courts to evaluate the claims of prisoners such that deference is given to prison administrators to set policies to ensure the safe operation of their prisons. Although the Court’s decision does not prevent inmates from making claims against AS confinement, it limits the scope of claims that they might successfully litigate to conditions that are needlessly harsh or unreasonable (Pizarro

& Narag, 2008). Additionally, the Prison Reform Litigation Act of 1996 was enacted to restrict the filing of prisoners' cases in federal court. Consequently, most of the case law surrounding AS has been on the grounds of a Fourteenth or Eighth Amendment violation (Collins, 2004).

Fourteenth Amendment

Under the Fourteenth Amendment of the U.S. Constitution, the state must adhere to certain procedures in deciding to deprive inmates of their liberty interest (Collins, 2004). In *Sondin v. Conner* (1995), it was determined that a liberty interest is created only when there is an "atypical and significant deprivation in relation to the normal incidents of prison life." In such cases, certain due process procedures are required.

Segregation that does not pose an atypical and significant hardship is not subject to due process, including such confinement that may occur during a period of investigation into inmates' misconduct (*Jones v. Baker*, 1998). However, in *Wilkinson v. Austin* (2005), it was decided that the plaintiffs' due process and liberty interest had been violated because the combination of conditions were significantly more restrictive than other Ohio state correctional facilities (e.g., isolation, lack of control over heating and lighting, no outside recreation) and because of the length of confinement. The court upheld the *Hewitt v. Helms* (1983) decision that these inmates were entitled to minimal procedural requirements, specifically timely notice of an AS evidentiary hearing, reason for confinement, and sufficient opportunity for response.

Extended confinement in segregation without a review hearing was also determined to be a violation of the Fourteenth Amendment. A New York court found that periodic review of inmates' continued need for such confinement is required (*McClary v. Kelly*, 1998).

Eighth Amendment

The Eighth Amendment ensures prisoners protection from cruel and unusual punishment. Because this concept is subjective, the Supreme Court has established the following standards:

- (a) shocks the conscience of the Court, (b) violates the evolving standards of decency of a civilized society, (c) punishment that is disproportionate to the offense, and (d) involves the wanton and unnecessary infliction of pain (Collins, 2004, p. 106).

In examining the conditions of confinement, the totality of circumstances must be weighed; although each individual condition might not be a violation, the combination of conditions might constitute one. Furthermore, prison officials must demonstrate "deliberate indifference" to a prisoner's basic human need in order for there to be an Eighth Amendment violation.

The use of prolonged segregation was tested in three significant cases in California (*Madrid v. Gomez*, 1995), Texas (*Ruiz v. Johnson*, 1999), and Wisconsin (*Jones 'El v. Berge*, 2001). Long-term segregation was not deemed a violation, except in the case of inmates with serious mental illness where extended stays were ruled unconstitutional. In *Madrid v. Gomez* (1995), not only was it ruled cruel and unusual punishment to place mentally ill inmates in the SHU, those at reasonably high risk of suffering mental illness as a result of SHU conditions were also restricted. Explicit in these cases is the requirement of correctional mental health staff to screen, assess, and monitor offenders for mental illness or emerging symptoms resulting from their placement in segregation.

It is also significant to note that in a number of states, settlement cases have also prevented or mitigated the placement of inmates with serious mental illness into long-term segregation. These states include Ohio,

Connecticut, Indiana, New Mexico, New York, and Mississippi. Other cases, in states such as New Jersey and Florida, have led to agreements to modify the terms under which prisoners with mental illness can be put or kept in segregation (Jamie Fellner, personal communication, June 10, 2010).

RESEARCH REVIEW

There exists a large body of peer-reviewed literature surrounding long-term segregation and solitary confinement. Many of these publications are literature reviews, theoretical articles, and case studies; few meet the American Psychological Association (2009) standard of empirical study article defined as reporting on original research or presenting new data analyses not addressed in previous reports, whether qualitative or quantitative. For example, in the 2008 special edition of *The Disturbed Offender in Confinement* published by Criminal Justice and Behavior, many of the nine articles focused on AS or other types of high security settings but only one (Lovell, 2008) presented an empirical study. Also in 2008, The Prison Journal released a special issue entitled *Supermax Prisons*. Only two of the eight articles (Sundt et al., 2008; O'Keefe, 2008b) meet the American Psychological Association standard for empirical research (2009). The large number of articles and corresponding lack of empirical research reinforce this as an important area of forensic psychiatry in which it is very difficult to conduct viable research.

The entire body of literature has been critical to advancing our understanding of AS confinement and its related issues. We relied on this literature to shape our hypotheses and research design in the present study. The case study research in particular has been useful to illustrate problems that might be attributed to AS (i.e., serious psychological harm) and highlight the need for research (see Benjamin & Lux, 1975; Human Rights Watch, 1997, 1999; King, 1999; Kurki & Morris, 2001; Rhodes, 2004). However, there are serious limitations with case studies. Small sample sizes, as are the norm in case studies, mean findings may not generalize to all, or even most, segregated offenders. Particularly concerning is that sampling procedures are often not discussed, suggesting that special care was not taken to select a representative sample. Additionally, these approaches do not provide a relative comparison of the participants' behavior in other settings; inmates who report serious psychological difficulties in segregation may experience those same problems in other prison settings or in society. Because we are interested in conducting an empirical study, our review of the research focuses on other empirical studies of the psychological effects of AS along with several key articles that informed our selection of psychological measures.

The SHU Syndrome

In 1983, Dr. Grassian described the psychopathological features resulting from AS that he believed to form a clinical syndrome, which later became known as the SHU syndrome in the wake of *Madrid v. Gomez* (1995) case. He interviewed 14 plaintiffs in a conditions-of-confinement lawsuit and described his clinical observations resulting from those interviews. Grassian noted perceptual changes, affective disturbances, cognitive difficulties, disturbing thought content, and impulse control problems that immediately subsided following release from such confinement. In more recent research, Haney (2003) found elevated symptoms of psychological trauma (e.g., anxiety, headaches, impending nervous breakdown, lethargy) and psychopathological features (e.g., ruminations, social withdrawal, irrational anger) among 100 SHU prisoners as compared to national probability samples. This constellation of symptoms composes the primary features of what has been coined the SHU syndrome.

Quantitative Research

Well-designed quantitative studies, although often not as rich in detail as case studies or qualitative research, can provide information about the impact of segregation on psychological well-being through the use of randomly sampled participants, representative samples, comparison groups, objective data collection strategies, standardized procedures, and analytical strategies that account for random error. Research on the effects of AS have been criticized for lacking these quality components that allow one to rule out plausible alternative explanations (Arrigo & Bullock, 2008; Metzner & Dvoskin, 2006; Pizarro & Narag, 2008; Zinger et al., 2001).

A key component that distinguishes research from demonstrations is the use of control or comparison groups. Because of the lack of a comparison group, some frequently cited studies are actually demonstrations of the potential impacts of AS (e.g., Brodsky & Scogin, 1988; Haney, 1993; Grassian, 1983). In the simplest research design, a study will compare a “treated” group to a control or comparison group to determine if the groups are different on the variable of interest. In a pure experimental design where participants are randomly assigned to conditions (e.g., segregation, general prison population), differences between groups would indicate the impact of segregation on the outcome variable; however in applied studies where random assignment to conditions is not feasible, the differences between the segregation group and a comparison group may be due to segregation or to other uncontrolled factors. The quality of the comparison depends on the similarity between the control/comparison group and the experimental/treated group.

Several quantitative studies have used comparison groups to explore the impact of segregation on psychological outcomes. Several of these studies have been experimental in nature in that inmates who volunteer to be randomly assigned to either segregation or comparison conditions for a short period of time (e.g., Ecclestone, Gendreau, & Knox, 1974; Gendreau & Bonta, 1984; Gendreau, Freedman, Wilde, & Scott, 1968, 1972; Gendreau, McLean, Parsons, Drake, & Ecclestone, 1970). These studies tend to show little impact of segregation on mental well-being but can be criticized for lacking ecological validity by using participants who volunteered to be placed in segregation, using small samples sizes, and for being short-term, all of which do not match the current reality of how AS exists in U.S. prisons today. To demonstrate ecological validity, conditions under investigation should reflect real life conditions. Similarly, comparisons to prisoners of war or use of college students and inmate volunteers, lacks the ecological validity necessary to generalize the findings to inmates in segregation.

Cross-Sectional Designs

Non-experimental research, which may demonstrate more ecological validity, have used a variety of comparison groups including general, non-inmate populations and norms (e.g., Haney, 2003; Hodgins & Côté, 1991; Suedfeld et al., 1982), general population prisoners (e.g., Hodgins & Côté, 1991), and inmates in different security levels who report being in segregation or not ever experiencing segregation (Suedfeld et al., 1982). Most, although not all, of these studies concluded that inmates in AS demonstrate higher levels of psychological distress. Because the quality of the conclusions depends on the similarity between the comparison group and the AS group, these cross-sectional studies lack the ability to attribute these differences to the conditions of confinement. In these studies, it is not possible to rule out alternative explanations due to selection bias and potential pre-existing differences, including psychological differences that may have existed prior to entering AS (i.e., there has been an inability to establish the time precedence between AS

and psychological well-being). An improved design strategy is to select a comparison group that has been matched to the segregation group on important variables (e.g., Lovell et al., 2007; Mears & Bales, 2009).

Longitudinal Designs

In order to truly understand how AS impacts the well-being of inmates, an improvement over cross-sectional design strategies with a comparison group is to study how inmates change over time using a longitudinal (or repeated measures) design. Studying intra-individual change allows for better understanding on whether change occurs as well as explication of how change occurs. In longitudinal designs, individuals serve as their own control group, and comparisons from baseline allow one to see how change is occurring. Adding a comparison group in a longitudinal design will allow one to rule out additional alternative explanations when change is (or is not) occurring.

There have been few longitudinal studies about the effects of segregation. Early studies by Gendreau and colleagues (Ecclestone et al., 1974; Gendreau & Bonta, 1984; Gendreau et al., 1968, 1970, 1972) used repeated measures experimental designs over periods of up to 10 days to explore the effects of segregation on psychological and physiological measures. Few negative impacts of segregation were found over these brief time periods. Although use of a repeated measures experimental paradigm improves over cross-sectional studies which may have selection bias issues, the short confinement periods are unrealistic for providing information on the effects of segregation as it is currently being used in U.S. prisons.

Only two recent studies were found that followed inmates for longer time periods after placement in segregation (Andersen et al., 2000; Andersen, Sestoft, Lillebaek, Gabrielsen, & Hemmingsen, 2003; Zinger et al., 2001). Andersen et al. (2000) studied participants over a 4 month period, but the majority of participants had data for less than a month. Zinger et al. (2001) followed inmates over a 60 day period. Both of these studies had high attrition rates (usually due to release from segregation), leading to a small percentage of participants who had complete data. Attrition is a major problem in longitudinal designs both for generalizability issues (i.e., are the participants who remain different from those who drop out) as well as analysis problems for those methodologies which require complete data from all participants (e.g., analysis of variance techniques). Newer methodologies developed for studying intra-individual change are less impacted by attrition rates. Although conclusions from these studies are limited by methodological weaknesses, both Andersen et al. (2000) and Zinger et al. (2001) demonstrated that segregated populations have more psychological disorders at the start than comparison subjects. However, these two studies provide conflicting evidence on whether conditions get worse over time. Thus, further longitudinal studies are needed to sort out these discrepancies and understand the long-term impacts of segregation.

THE COLORADO SYSTEM

In Colorado at the time of this study, there were four designated AS facilities. Colorado State Penitentiary (CSP) opened in 1993 as a 756-bed male AS facility in its entirety. At the Sterling Correctional Facility, 192 of its 2,545 beds were constructed to house male AS inmates in three units that are separate from the rest of the facility. The San Carlos Correctional Facility (SCCF) is a male acute care psychiatric prison, with nine units of varying security levels. One 26-bed unit at SCCF is designated for AS classified inmates. Generally, AS at SCCF is reserved for inmates already housed at SCCF needing high security or for inmates in AS at CSP or Sterling Correctional Facility whose psychiatric needs exceed those available at their current facility. The fourth AS facility is a 24-bed unit located at the multi-custody Denver Women's Correctional Facility. Be-

cause it houses the largest number of AS inmates and no other custody levels, CSP was the only AS site included in this study.

The Colorado Department of Corrections (CDOC) had 25 state and 7 private prisons that managed 19,279 inmates as of June 30, 2007, which marked the start of data collection. There are five security levels – minimum, minimum-restrictive, medium, close, and AS – to which offenders are assigned. CDOC uses a standardized, objective classification instrument that was developed specifically for the management of Colorado’s inmate population (Austin, Alexander, Anuskiewicz, & Chin, 1995). The classification instrument is used to assign inmates to minimum through close security levels. However, AS is a long-term segregation placement for inmates who display violent, dangerous, and disruptive behaviors and placement is determined through an administrative action that is separate and distinct from both the usual classification system and the disciplinary system. Although disciplinary infractions may affect classification at all levels, the disciplinary process is a punitive response to a finding of guilt for an institutional rule violation and may result in punitive segregation, which can extend up to 60 days. Therefore, punitive segregation is of short duration used for punishment and AS is of long duration used for management purposes.

The administrative action to classify an offender to AS begins with a hearing, frequently following either a serious violation or a series of less serious infractions. Colorado does not house protective custody inmates; therefore, no AS placements occur at the request of inmates. Also, during the study, newly arrived inmates were not placed directly into AS upon intake into DOC except in rare cases for violent behavior in county jail or for an interstate compact case transferred from AS in another prison system. Although the disciplinary system only allows for punitive segregation *following* a finding of guilt, pre-hearing segregation (removal from population) may occur immediately following a serious incident for the safety and security of the facility. Therefore, in the time leading up to and during their AS hearing, inmates have typically been in segregation.

AS Offenders in Punitive Segregation

All facilities across the state of Colorado have punitive segregation beds with the exception of CSP and minimum security facilities. Minimum custody offenders are transported to a higher security facility to complete their punitive segregation time. When offenders are placed in punitive segregation, they are removed from the general population (GP) and taken to an isolated part of the facility to be placed in a single cell. Punitive segregation offenders remain in their cell for 23 to 24 hours a day, only coming out for recreation and showers, both of which are located within the living unit. Therefore, most do not leave the unit during their segregation time. Services including meals, library, laundry, and even medical and mental health appointments occur at the cell door. If a situation warrants an offender to be out of cell, the offender is placed in full-restraints and escorted to a room within the unit where he or she can meet with staff privately. Many offenders do not like being taken out of their cell unless absolutely necessary because of the use of full restraints. Additionally, they may not like leaving their cell because officers may take the opportunity to search the cell for contraband.

Due to the disciplinary nature of punitive segregation, offenders are stripped of most privileges during their stay. Punitive segregation inmates are neither allowed to work nor are they permitted to participate in programs or education. Furthermore, their televisions are removed, and they cannot order canteen beyond essential hygiene items.

Punitive segregation is a highly restrictive environment, only intended to be used for a short period of time. Once reclassified to AS, offenders may remain in a punitive segregation bed while waiting for an AS bed to become available. This can be problematic as GP facilities are not designed to house offenders in long-term segregation and the small number of punitive segregation beds at each GP facility can fill up quickly. Furthermore, while punitive segregation offenders are not afforded privileges, AS offenders are granted limited privileges such as visiting, which happens outside of the unit. Visitation is labor intensive because it requires escort by two correctional staff. In addition, while being held at an AS facility, offenders who behave well and complete their required programming and education are able to progress through a step program whereby they earn more phone sessions, visiting time, and privileges (e.g., TV, canteen). Only two punitive segregation facilities offer a step program for privileges, and there are none that provide the opportunity for programming or education. This means that while AS offenders are held in a punitive segregation bed, they are unable to begin working their way toward leaving segregation.

CSP Conditions of Confinement

Once an AS offender is moved from a GP facility and assigned to CSP, he is transported to CSP where he completes his AS time. Offenders are taken into CSP through intake, which is located on the lowest level of the facility. While in intake, offenders are placed in a holding cell that is similar to their permanent cell. During this time, the offender watches an orientation video that outlines what he can expect and what is expected of him during his time at CSP. He also has a brief visit from mental health, conducted at the cell door. While the offender is going through orientation, property staff assesses his belongings to ensure that no unallowable items enter with the offender, as they are permitted fewer property items than in GP facilities. This also prevents dangerous contraband such as drugs or weapons from entering the facility. Once the offender has completed orientation, usually within the first few hours, he is escorted to his permanent cell in a different area of the facility.

Physical Environment. CSP has six identical pods, or living units. When the offender enters the pod, he is escorted down a long hallway that opens into a circular area. In the center of the area is a tower with an office for housing unit staff on the lower level and the pod's control center on the upper level. Officers manning the control center operate all doors or sliders into the pod, including those to offenders' cells. Correctional staff standing in either the lower or upper levels of the tower can see into all eight of the day halls. Each day hall contains 15 to 16 offender cells separated onto two tiers with each tier having 7 or 8 cells, a shower, and a recreation room.

The cells in CSP are 80 square feet with 35 square feet of unencumbered floor space and contain a bunk, toilet, sink, desk, and stool. Each of these items is made of metal and is mounted to the wall or floor for security. Every cell has a 5" x 45" window on the exterior wall above the offender's bunk through which the offender can see outside. There is also a window on the cell door that faces the day hall. Depending on the pod, the window is either 3.5" x 20.5" or 5" x 15". Neither of these windows opens, which precludes the offender from receiving outside air while in his cell.

Per CSP policy, offenders wanting to participate in recreation are generally permitted at least one hour five times per week (as well as to shower for 15 minutes three times per week which generally coincides with an offer to exercise), assuming that there are no facility occurrences disrupting this schedule. When an offender is offered recreation and chooses to participate, he is placed in full-restraints and escorted from his cell to the recreation room at the end of the tier. The recreation room is a 90-square foot cell that contains a pull-

up bar mounted to the wall. No other exercise equipment is allowed. The only opportunity offenders have to receive fresh outside air is through two 5" x 60" grated windows on the exterior wall of the recreation room. On the interior, a glass wall faces the V-shaped day hall, so the offender in recreation is fully visible. Though prohibited by the facility, an offender in the recreation room may call out exercises to other offenders who in turn workout in their cells.

There are light and sound standards for CSP. Standards for CSP require that ambient sound does not exceed 70dBA during the day or 45dBA at night. A sound measurement of offender housing units at CSP, on a single day, returned an average of 55dBA at 7:50 AM and 42dBA at 10:40 PM. Although staff attempt to regulate the ambient sound of the facility, it can be difficult to regulate the noise level of 756 offenders; these measurements do not reflect periods of sound elevations produced by inmates' yelling and banging. Additionally, each offender is entitled to at least 20 foot-candles of light in the desk area of his cell. A light measurement of offender cells returned an average of 55 foot-candles of light in offenders' cells. Offenders have two 32-watt lights over the desk in each cell that they are able to control. In addition, each cell contains a 7-watt security light underneath the desk that stays on 24 hours per day.

Interpersonal Communication. Each cell has an intercom system through which correctional officers can contact each offender from the unit's control center. Officers use the intercom system to ask prisoners questions such as whether or not they want to attend recreation or take a shower. They also use the intercom to inform inmates when they will be leaving their cell for such things as a mental health visit, a family or friend visit, or if the offender will be escorted to another part of the facility or off grounds. Conversely, inmates can use the intercom system by pushing a button in their cell to contact staff, which they may do to request items (e.g., razor, toilet paper) or simply to chat. Staff also has the ability to monitor conversations using the intercom system.

While the intercom system provides a means for correctional staff and offenders to communicate with each other relatively easily, it does not afford offenders the opportunity to communicate with one another. Many offenders at CSP have become skilled in sign language. Since each day hall is V-shaped and cell doors have windows, offenders are able to communicate with each other using sign language. This aids in keeping the noise level down in the day hall and gives inmates the opportunity to speak to each other without the risk of staff overhearing. At times, however, many inmates simply yell through their cell door so that other offenders can hear. When this happens, the day hall can become very noisy.

Due to the safety concerns of the facility and the fact that moving an AS offender from his cell is staff intensive, offenders in AS receive many services at their cell door. At CSP, officers make rounds every 30 minutes to do a visual check into the cell of every offender. Mental health clinicians are required to do monthly rounds as well. During these rounds, clinicians go to the cell door of every offender in their assigned pod and check in with the inmate to see how he is doing. If the offender is well, the clinician moves on; however, if the clinician feels the offender needs follow-up, he or she will schedule an appointment with the offender for a later time. This appointment will be conducted in the visiting room, not cell side. In addition to rounds, offenders receive their library service and educational services at their cell door. Once a week, a librarian picks up library kites, or requests, and distributes books and magazines to offenders who put in a kite the previous week. When an offender is participating in programming or education, the teacher or counselor distributes homework to each inmate through the cell door and also collects completed assignments in the same manner.

Mental Health Services. In addition to mental health services received cell-side, offenders who are diagnosed with a mental illness receive more in-depth mental health services. Offenders with mental illness who are stable are offered a one-on-one session at least once every 90 days. Those with acute mental health needs are required to be seen at least once every 30 days. Although there are requirements on mental health staff to schedule appointments, offenders may refuse these appointments. Conversely, if an offender feels his mental health status has changed since his last monthly round, he may put in a request to see a clinician sooner than scheduled. If necessary, clinicians will schedule an offender for a mental health session for 1 to 2 hours per week as they are available; this is infrequent but most likely to occur following a crisis event. Additionally, if a mental health clinician feels a prisoner requires psychotropic medication, an appointment is made for him to meet with a psychiatrist. This visit may happen in a noncontact visiting booth or via teleconference.

Mental health appointments occur in a noncontact booth in the visiting room, unless the offender has declared a mental health emergency. If an offender has threatened self-harm, he is often taken to intake and placed in a special observation cell where he is stripped of his belongings and can easily be observed by staff for his safety and the safety of staff. An offender is kept in the observation cell until the clinician can make a reasonable assumption that the offender no longer plans to self-harm or for 72 hours, whichever comes first. If the clinician determines the offender needs to be observed beyond 72 hours, approval is needed from administrators and a mental health supervisor outside of the facility. Offenders who remain in a mental health crisis situation beyond the three to five day window are then sent to the infirmary at a different facility. There are generally four to six mental health clinicians who are responsible for managing the mental health needs of offenders at CSP. When the facility is fully staffed with six clinicians, each is assigned to a pod of 126 offenders, but when there are vacant positions, clinicians are required to cover their pod's mental health needs and split an additional pod with another clinician.

Quality of Life Program. When an offender arrives at CSP, his length of stay is indeterminate because it is based upon his behavior and ability to comply with programming requirements. The average length of stay at CSP is two years (O'Keefe, 2005). CSP provides incentive-based behavior modification and cognitive programs. Every offender must successfully complete three cognitive classes with each lasting three months. Successful completion of the required programming along with modeling appropriate behavior is the primary way for an offender to work his way out of CSP. The goal of these programs is to provide offenders with tools so they may be successfully reintegrated into lower security prisons.

CSP's incentive-based programming consists of three quality of life (QOL) levels. Each level brings with it more privileges; however, these privileges must be earned by the offender through appropriate behavior and compliance with CSP rules. Each level has a prescribed minimum number of days: 7 for level one, 90 for level two, and 90 for level three. Because offenders are required to complete three 90-day cognitive courses and there are often program waitlists that may result in an offender staying on levels two or three for longer than 90 days, the total program length is expected to last a minimum of one year. Additionally, offenders who misbehave may be regressed through the levels, extending their time in the program.

QOL level one is much like punitive segregation in that offenders are not permitted to have a television or to participate in programs or work. Furthermore, offenders at this level are only allowed one 20-minute phone session and one 2-hour noncontact visit per month, should they remain at level one for that length of time. They are able to order items from the canteen at a maximum of \$10 per week. Though limited in compari-

son to what GP offenders are able to buy, offenders at CSP have a variety of food, hygiene, faith, and personal items available for purchase. Additionally, AS inmates are permitted three library books at any given time. All other property must fit inside a 2 cubic feet duffle bag; however, as long as the property can fit in the bag, they are allowed two personal books, two magazines, and one newspaper. Other items that offenders at this level may have are photographs and an address book.

Level one offenders are automatically reviewed on their seventh day by the unit sergeant. Unless they received a negative write-up or report, offenders automatically progress to level two after seven days. Those who do not progress are extended another seven days until their next review.

Inmates become eligible for cognitive classes when they have been elevated to QOL level two. Offenders at this level are allowed a television, and if they are indigent and unable to afford one, the facility will loan one to them. This is beneficial as CSP provides some of their programs and recreational activities through the television. Offenders also receive 20 television channels that they are able to control from their cell and view at their leisure 24 hours per day. Also available through the television is a music channel that plays at designated times and rotates through music genres. Additionally, offenders have the opportunity to play bingo on a monthly basis. The bingo numbers are selected and aired over closed circuit TV at the facility. Offenders who wish to participate in Bingo receive six board games and are awarded a candy bar for each verified bingo.

In addition to programming received through their television, offenders at this level are permitted art supplies (colored pencils, art paper, drawing patterns, and coloring pictures), games (solitaire and kings table), puzzles (crossword, word fill-ins, word search, and Sudoku), and pamphlets for in-cell exercises offenders (push-ups, stretching, and isometrics). Offenders may request a new supply of colored pencils every six months and are able to receive four new sheets of art paper and new puzzles on a weekly basis.

At level two, offenders are permitted to increase their weekly canteen order to \$20 and have an increase in both their phone privileges, to two 20-minute phone sessions per month, and their visiting privileges, to two 2-hour noncontact visits per month. However, offenders at this level remain unable to work. Once an offender has completed a minimum of 90 days on level two, has been compliant with programming, has not had any negative write-ups for at least 90 days, has had appropriate interaction with staff, and has sustained suitable cell conditions, he may be progressed to QOL level three. Offenders' case manager initiates the paperwork for a level progression, which requires approval by the housing captain.

Arguably one of the most important benefits of QOL level three is an offender's ability to have more contact with friends and family. While offenders' visits remain noncontact, they are increased to four 3-hour visits per month and four 20-minute phone sessions. Offenders are also permitted to order as much as \$25 worth of canteen per week. One additional benefit is that offenders may now be eligible to work as a porter or barber. There are 54 positions available to offenders at CSP. Benefits to being offered a job position include the ability to earn money, increased time out of their cell, and two additional phone sessions per month. However, simply being at QOL level three does not automatically qualify an offender for a job. If intelligence officers feel the offender is a threat to the facility, he will not be permitted to work as a porter or barber. A QOL level three inmate may be deemed a continued threat due to an institutional history of assaultive behavior or from intelligence that suggests he may use the opportunity to intimidate or pass gang information or contraband to other offenders. There are no time limits restricting how long an offender can be in a job position and there is a waitlist of offenders who have put in a request to work; however, because of the

progression of offenders out of CSP and offenders who their work privilege, the same 54 offenders are not usually working for more than a couple of months.

The restrictions inherent in AS diminish staff's ability to impose traditional sanctions for institutional rule violations. Offenders in GP who are found guilty of a rule violation can receive a maximum of 60 days in punitive segregation or up to 180 days in loss of privileges (e.g., TV, canteen, visiting). Offenders in AS are required to follow the same institutional rules as GP offenders; however, because AS offenders are already in segregation for an indeterminate amount of time, they cannot receive additional segregation time as punishment. They can still lose privileges and be regressed through the QOL system. Sanctions are tailored to the seriousness of the infraction.

Offenders who engage in minor rule violations may initially receive a warning that is documented in a chronological record report. If the behavior continues, the offender may lose a privilege for a short amount of time (e.g., three days) without losing a QOL level. For example, this may happen if an offender covers the security light in his cell to make it darker for sleeping. This may also happen if an offender is caught "rat lining" or "fishing," which are forms of communication or exchange of items between offenders locked in their cells.

When an offender in AS violates a serious institutional rule, the officer initiates documentation on which he may recommend that the offender be regressed to a lower QOL level. This recommendation is approved or denied by a housing lieutenant. In general, if the lieutenant approves the offender's level regression, he is dropped one level. This process is kept separate from the disciplinary process, which may or may not result in a guilty finding, in order to have an immediate response to an offender's negative behavior. The disciplinary process can be lengthy because of due process requirements, but he may also receive a loss of privileges sanction through the disciplinary process.

Regardless of the offender's level, if he engages in behavior that dangerously disrupts the operation of the facility, he will be placed on special controls in the intake unit where he can be carefully monitored. This often happens during what is referred to as a use of force incident, which is any time an officer uses any level of force against an offender. A use of force incident generally occurs when an offender assaults a staff member or refuses to comply with a lawful order (e.g., refuses to be restrained for escort). An officer's response can include the use of simple pressure point tactics, the use of agents such as oleoresin capicum (OC), or a forced cell extraction of the offender. During both fiscal years 2008 and 2009, CSP had an average of seven use of force incidents per month. Upon an offender's return to his cell, he will automatically begin at QOL level one again. Though the offender will not be required to retake any of the cognitive classes that he has already completed, he will be terminated from any classes in which he is currently enrolled and will be required to begin his process through the QOL levels again. Additional sanctions may be imposed through the disciplinary process.

Offenders who have difficulty progressing through the QOL level system may require special consideration. Offenders in segregation can accumulate a high number of sanctions through behaviors such as breaking the sprinkler head in their cells or overflowing the toilet in their cells, causing flooding on the tier. It is difficult to manage and change the behavior of offenders who have so many sanctions that there is no tangible incentive to improve their behavior. When this is the case, case managers and housing staff enact a behavior management plan. In a behavior management plan, case managers and correctional staff will use one privi-

lege (e.g., TV) that is highly valuable to the offender as an incentive. If the offender can behave well for a short period of time (generally 7 to 10 days), he may receive a television despite his loss of privilege status.

When an offender has been at level three for at least 90 days with good behavior and has successfully completed the requirements of the program, he is interviewed for progression out of CSP. A classification committee must approve the decision to reclassify him to close custody, and then he is moved to the Centennial Correctional Facility when a bed becomes available, where he continues to work toward completing his reintegration programming. It is less common that an offender transitions out of CSP any other way; however, offenders do sometimes parole from CSP or release when they reach the end of their sentence while in AS. Additionally, an offender may be released from AS based on a warden's review. An offender may receive a warden's review if he has been in CSP for more than two years but has been unable to progress out of CSP. If it is felt that the offender no longer needs to be in CSP, he may be released back to GP without transitioning through Centennial Correctional Facility.

Progressive Reintegration Opportunity (PRO) Unit. At the Centennial Correctional Facility, there is a continued focus on behavior modification and cognitive programs to transition disruptive offenders to less secure environments. Most offenders complete QOL levels four through six in the PRO unit. Upon transfer from CSP, offenders are reclassified from AS to close custody, the next highest custody level. Upon arrival, little is different for the newly classified close custody offenders; however, as offenders work their way through the PRO unit levels, they work toward contact visits with friends and family and are eventually allowed recreation time in the gym with other inmates. Ultimately, offenders who are successful in completing all six QOL levels are released back to GP.

Offenders with Mental Illness (OMI) Management Program. During the course of the research project, the OMI management program was opened at Centennial Correctional Facility. In addition to the PRO unit, the OMI program was designed to be a transitional program from CSP specifically for prisoners with a mental illness. Offenders are selected for the OMI program by a multi-disciplinary committee and must be approved for reclassification as a close custody inmate. In order to be considered for transfer to the program, offenders must have been in AS for a minimum of six months, enrolled in a cognitive program, have a mental health disorder, and be actively working with a mental health clinician.

Upon transfer to the OMI program, inmates are automatically placed in the intermediate program level. The OMI program has three levels: high, intermediate, and low. High is the most restrictive level with low the least. Depending on the individual's behavior, he can be moved to high or low levels. The program focuses on treatment and socialization. Offenders in this program work their way toward earning more privileges than are available in AS, contact visits with friends and family, and recreation in the gym with other inmates, much like PRO unit offenders do; however, the OMI program has the added benefit of group therapy. Initially, offenders are afforded the opportunity to participate in group therapy by being tethered to a special table. As offenders progress through the program, they are eventually allowed in groups of eight untethered inmates. The goal is to transition offenders to GP or the community although placements in the program may be long term.

San Carlos Correctional Facility (SCCF)

SCCF is a 255-bed special needs prison designed to stabilize and treat offenders with the most acute psychiatric symptoms or with developmental disabilities who are at risk for self injury as a result of their illness

and who have shown a substantial impairment in their ability to function at another correctional facility. SCCF houses inmates at all five custody levels. SCCF is unique in that offenders of all custody levels live and interact with one another on their living unit, with the exception of AS offenders who are housed in a separate unit.

All offenders who arrive at SCCF are processed through the intake/assessment unit. New arrivals are interviewed by both the mental health clinician assigned to the unit and a psychiatrist. New offenders are not permitted interaction with other inmates for the first 72 hours. If after 72 hours, clinicians and correctional staff feel the offender can reasonably interact with other offenders, he will be allowed in the day hall with as many as five other inmates. Offenders on the intake unit are permitted out of their cell in the day hall for at least one hour a day, five days per week. During this time, they have open access to the phones and showers.

Offenders typically progress through the programming levels as their mental health status improves. Offenders are continually monitored by a psychiatrist with an appointment every 30 days for the most severe offenders or every 60 days for those who are progressing well. One-on-one sessions occur with a mental health clinician as needed and are not scheduled on a regular basis; however, there is a clinician assigned to each unit, with each unit housing fewer than 30 inmates.

As an offender continues to progress through the facility, he will work his way towards open access to the day hall, phones, and shower. As he progresses through the facility he will then be allowed out with seven inmates and then fifteen, eventually earning all day open access. Those who have progressed to the lowest levels are also permitted one hour of recreation five days per week in the yard or gym plus three hours per week at the library. Additionally, they are able to participate in group therapy sessions, which happen once or twice a week depending on the topic. Group therapy subjects include anger management, understanding one's mental illness, and other related topics. Once mental health, psychiatric, and correctional staff determine that the offender has improved enough to function in GP, he is then transferred to a facility at his custody level.

PURPOSE OF PRESENT STUDY

The broad purpose of the project was to evaluate the psychological effects of long-term segregation on offenders, particularly those with mental illness. This study examined conditions as they existed in the Colorado prison system with respect to AS, using CSP as the AS study facility. Only males were included because females represent 2% of Colorado's AS population. We did not assign inmates to segregation, but studied those conditions as they naturally occurred. The following were the primary goals and hypotheses of the grant.

Goal 1: To determine which, if any, psychological domains are affected, and in which direction, by the different prison environments. A multitude of psychological dimensions were examined, drawing from those most often cited in the literature. The broad constructs of interest were depression/hopelessness, anxiety, psychosis, withdrawal and alienation, hostility and anger control, somatization, hypersensitivity, and cognitive impairment. We hypothesized that offenders in segregation would develop an array of psychological symptoms consistent with the SHU syndrome, with elevations across the eight constructs.

Goal 2: To assess whether offenders with mental illness decompensate differentially from those without mental illness. We were particularly interested in whether long-term segregation had a differential impact based on the presence of mental illness in offenders. We sought answers to the following questions: Does AS exacerbate symptoms in offenders with mental illness? Does AS create symptoms of mental illness in those who did not exhibit any at placement? It was hypothesized that offenders with and without mental illness would deteriorate over time, but the rate at which it occurred would be more rapid and more extreme for the mentally ill.

Goal 3: To compare the impact of long-term segregation against the general prison setting and a psychiatric care prison. In this study, the psychological and behavioral symptoms of offenders in AS were compared to similar offenders who were sent to SCCF or returned to the general prison population pursuant an AS hearing. This study used a repeated measures design over the course of a year to explore whether psychological distress was attributable to the various prison environments. It was hypothesized that inmates in segregation would experience greater psychological deterioration over time than the comparison groups.

This study also included an examination of individual characteristics such as mental health status, personality, and trauma history to determine if certain factors could predict patterns of change. The prediction analyses were exploratory in nature and we did not formulate a hypothesis about the variables that might predict differential rates of psychological decompensation.

METHOD

GROUP ASSIGNMENT

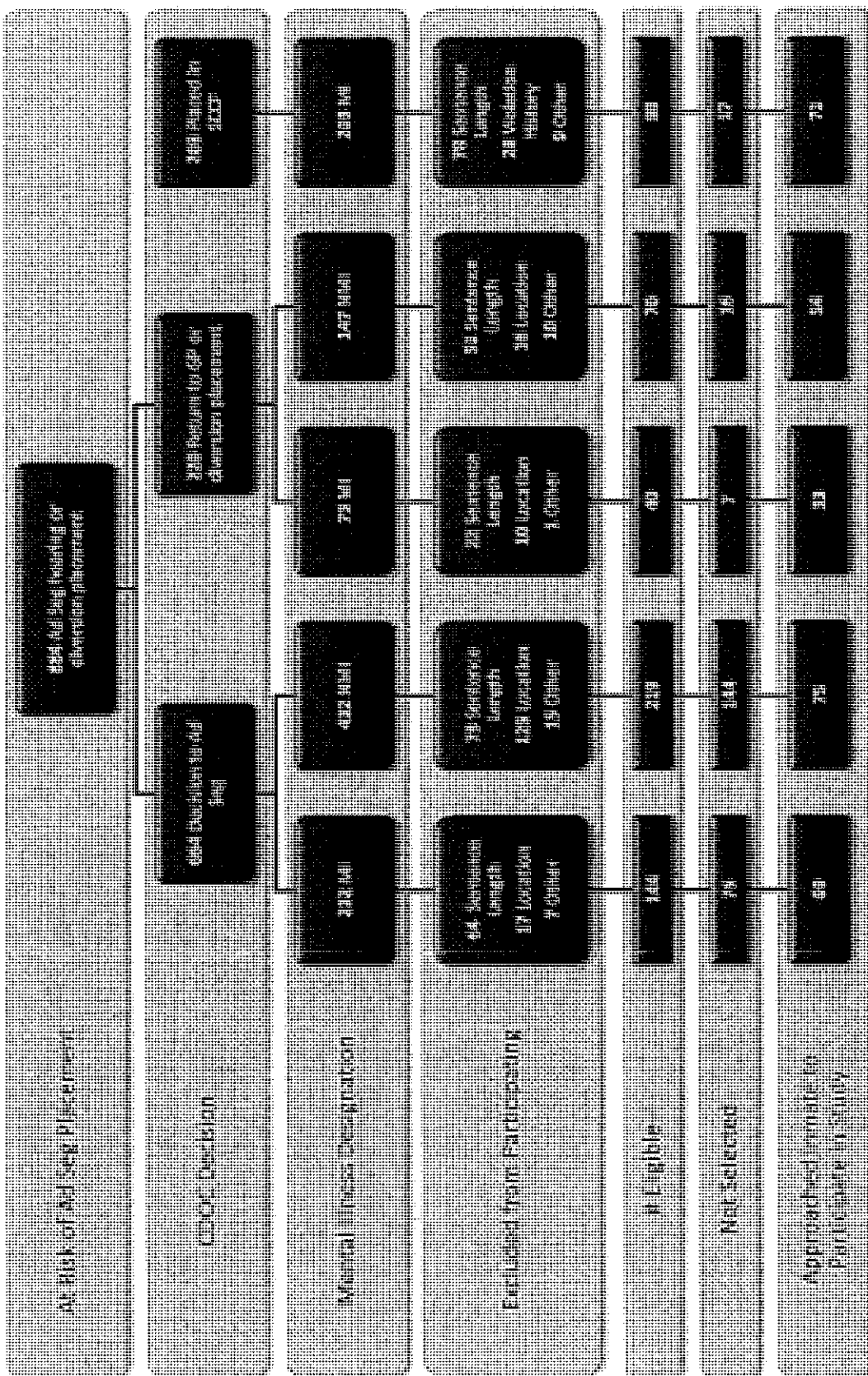
Study participants included male inmates placed in AS and comparison inmates in the GP. Placement into AS or GP conditions occurred as a function of routine CDOC operations, pending the outcome of their AS hearing, without involvement of the researchers. Inmates were identified as study candidates at the point the offenders were notified that they would have an AS hearing. Oftentimes, it was unknown whether a particular inmate would be placed in AS or returned to GP at the time of his study consent; approximately 10% of hearings do not result in AS placement. For the purposes of this study, all study participants classified to AS were waitlisted for and placed in CSP (as opposed to Sterling Correctional Facility). Inmates who returned to GP following an AS hearing were assumed to be as similar as possible to AS inmates and, therefore, comprised the comparison groups. Comparison participants also included inmates targeted for a diversionary program that identified inmates at high risk of AS placement due to their disruptive behavior. This program discontinued shortly after the study commenced, hence few participants were identified through this method.

Inmates in both of these settings (CSP, GP) were divided into two groups – offenders with mental illness (MI) and with no mental illness (NMI). There are fewer inmates with mental illness than without, but because both subgroups were of equal interest to this study, separate groups enabled over-selection of inmates with mental illness. All offenders are rated on a psychological needs level by trained clinicians upon intake into CDOC and periodically during their incarceration as warranted. The psychological needs level has a 5-point rating, where higher values indicate the need for more intensive services, and a qualifier code that indicates whether the offender has a serious and persistent mental disorder. Most inmates rated 3 through 5 have an Axis I diagnosis, although certain Axis II diagnoses may infrequently warrant this rating (e.g., borderline, schizotypal). Disorders that typically qualify as serious and pervasive are mood disorders including major depression, other depressive disorders, dysthymic, and bipolar disorders; psychotic disorders including schizophrenic, paranoid, delusional, and schizophreniform disorders; dissociative identity disorder; and posttraumatic stress disorder. In this study, inmates assessed with a psychological needs level of 3 through 5 were defined as MI and levels 1 or 2 were defined as NMI.

A third comparison group was included. This group included inmates with severe mental health problems placed in SCCF. Of the inmates placed in SCCF, only those with patterns of prison misbehavior, as measured by disciplinary violations, were included in the study. However, inmates placed into AS at SCCF were excluded because of the small number and because many had transferred from AS at CSP or Sterling Correctional Facility, where the effects of the earlier placement would be unknown. The purpose of the SCCF comparison group was to study inmates with serious mental illness and behavioral problems who were managed in a psychiatric prison setting.

Figure 1 illustrates the number of offenders who were eligible for the study and details the selection of offenders within each of the five study groups. Given that the purpose of this project was to study long-term segregation, inmates projected to release from prison before administration of the final testing session were excluded. Inmates were also excluded if they could not read English or if their reading level was not high enough (roughly eighth grade) to complete the battery of tests. SCCF inmates were excluded if they did not have significant disciplinary violations in their history. Infrequently, offenders were excluded for other reasons such as being an interstate compact offender, being the suspect in a high-profile murder investigation (as reason for AS placement), or a visual impairment prohibiting them from reading. Finally, inmates were sampled as a matter of convenience. Because this project funded only one field researcher, participants were selected based on their proximity by either timing or location to others who could be included in this study.

Figure 1. Eligibility and Selection of Study Participants



PARTICIPANTS

Figure 2 details the flow of participants through the study, including an account of how many offenders completed the testing at each interval. A total of 302 male inmates were approached to participate in the study. Thirty refused to participate. Two more offenders were considered a passive refusal and were removed for inappropriate sexual behavior towards the researcher during the first testing session. An additional 23 offenders later withdrew their consent, although the data collected to the point of their withdrawal was used. In addition to refusals and withdrawals, 10 inmates released prior to the end of the study due to discretionary releases by the Parole Board and one GP participant died of a drug overdose.

Five testing sessions were initially established at 3-month intervals, beginning with the date of consent and initial administration. Therefore, tests were scheduled at 3 months, 6 months, 9 months and 12 months after the baseline assessment. However, this schedule was problematic for the CSP groups. When the study began, there was a 3-month average wait for inmates to be transferred to CSP due to a shortage of AS beds. While on the waitlist, AS inmates were held in a punitive segregation bed at their originating facility. It was determined that the primary goal was to study inmates in a single long-term segregation facility (CSP) to limit confounding variables and that therefore the baseline measure should be collected upon placement into CSP. However, it was also recognized that significant changes could occur while inmates were held in segregation at their originating facility. Therefore, a “pre-baseline” measure was collected as close to the AS hearing as possible, which meant that the CSP groups completed six test intervals rather than five. The time between the pre-baseline and baseline measure varied according to how long the inmate was on the waitlist. The median time between pre and baseline tests was 99 days, although eight offenders were moved into CSP so quickly that they did not have a pre-baseline measure. In the analyses, tests were aligned across groups according to the test number, such that the CSP groups had an additional test at the end rather than at the beginning.

Participants’ ages ranged from 17 to 59 at the time of consent, with a mean age of 31.8 ($SD = 9.1$). The racial/ethnic breakdown of participants was 40% white, 36% Hispanic, 19% African American, 4% Native American, and 1% Asian. Of the inmates with mental illness who were included in this study, 56% were identified with a serious and pervasive disorder. Other participant characteristics are described in greater detail in the results section, including comparisons of study samples to eligible pools and comparisons of refusers to participants.

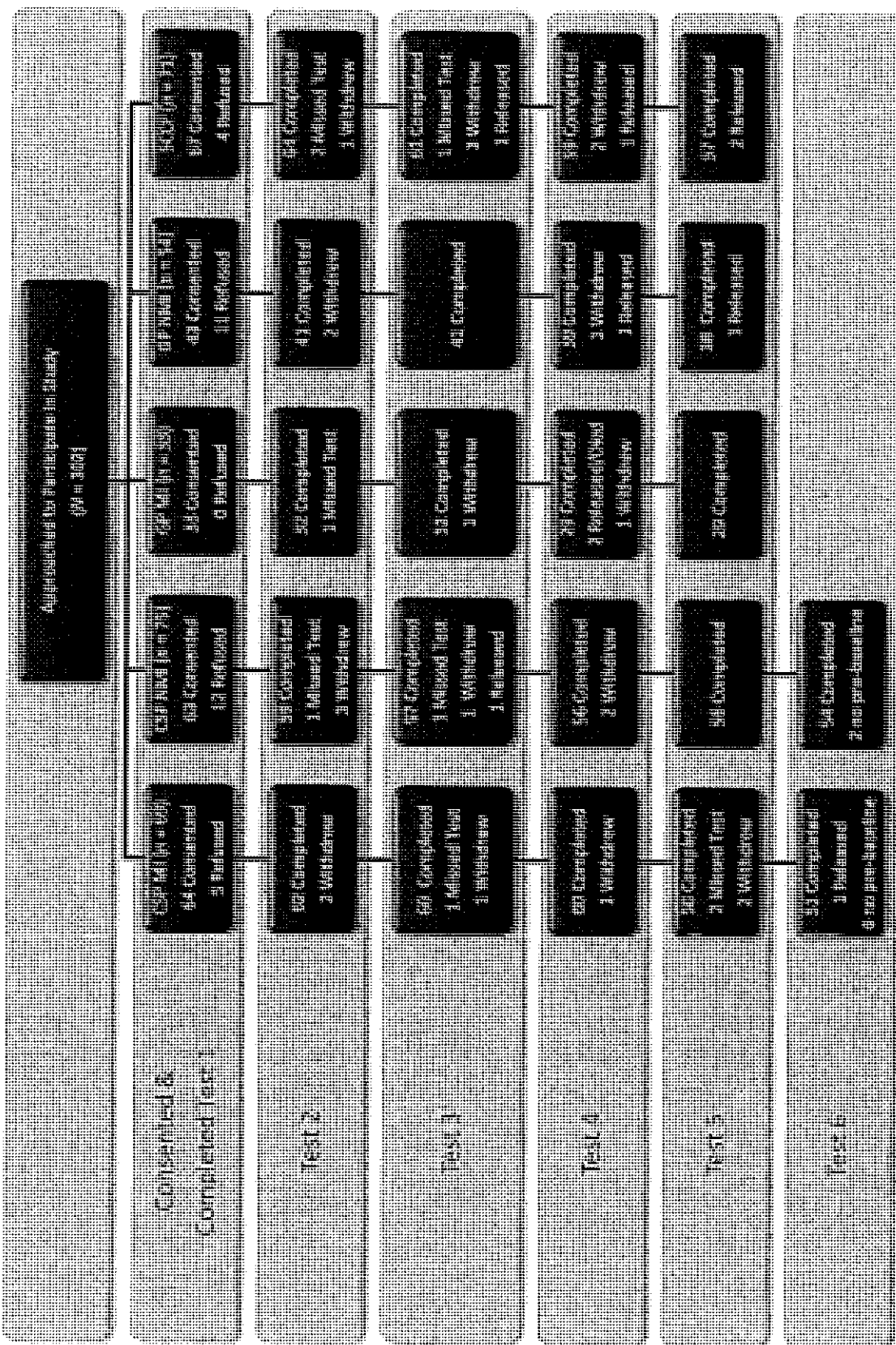
MATERIALS

Assessment tools were selected to comprehensively cover the variety of psychological constructs associated with AS (e.g., Arrigo & Bullock, 2008; Grassian, 1983; Haney, 2003). The primary constructs assessed in this study were as follows: (1) anxiety, (2) cognitive impairment, (3) depression/hopelessness, (4) hostility/anger control, (5) hypersensitivity, (6) psychosis, (7) somatization, and (8) withdrawal/alienation. Additionally, malingering, self-harm, trauma, and personality disorders were assessed.

Research materials were selected to meet the following criteria: (1) use of assessments with demonstrated reliability and validity, (2) use of multiple sources for providing information (e.g., self-report, clinician ratings, files), (3) use of multiple assessments of each construct of interest, (4) ability to use within the prison setting, and (5) ease of administration, including no specialized equipment, no physical contact, short length of time, and appropriate reading level.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions and views of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Figure 2. Flow of Participants through Study



This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

After selection of the self-report assessments was complete, there remained several areas of interest (e.g., panic disorder, hypersensitivity to external stimuli, physical hygiene) for which there was no established measure that met our criteria. In conjunction with the study advisory board, the research team developed a 39-item instrument to assess these areas. This instrument, called the Prison Symptom Inventory (PSI), is shown in Appendix A.

In addition to self-report assessments, ratings of psychological functioning were obtained from clinical staff and ratings of behavior in the housing unit were obtained from correctional staff. Official record data were also gathered from electronic and paper files. This section summarizes information for self-report assessments, staff ratings, and behavioral data. Complete descriptions of the individual measures and their known psychometric properties from past research and for the current study are provided in Appendix B. Additional analyses of the psychometric properties of the PSI are presented in Appendix C.

Data were collected directly from participants on 12 self-report assessments (ten paper-and-pencil tests, two administered by the researcher) to assess 12 different constructs. Table 1 provides a list of the assessment tools for each construct. Most assessments were collected at each testing period, although personality disorders, self-harm, and trauma history were not collected at all time periods. It was determined that personality and trauma history were relatively stable constructs that needed to be assessed only once to limit the testing burden on study participants. Also, due to the burden on already limited mental health resources, the BPRS was only administered at the first, third, and fifth testing intervals.

Table 1. Assessments and Constructs

	Anxiety	Cognitive Impairment	Depression – Hopelessness	Hostility – Anger Control	Hypersensitivity	Psychosis	Somatization	Withdrawal – Alienation	Malingering	Personality Disorder	Self-Harm	Trauma
Beck Hopelessness Scale (BHS)			I									
Brief Psychiatric Rating Scale (BPRS)												
Activity												
Anxious-Depressed	C		C									
Hostility/Suspiciousness				C								
Thought Disorder						C						
Withdrawal								C				
Brief Symptom Inventory (BSI)												
Anxiety	I											
Depression			I									
Hostility				I								
Interpersonal Sensitivity					I							
Obsessive-Compulsive	I											
Paranoid Ideation						I						
Phobic Anxiety	I											
Psychoticism						I						
Somatization								I				
Coolidge Correctional Inventory (CCI)										I		
Deliberate Self-Harm Inventory (DSHI)											I	
Personality Assessment Screener (PAS)												
Acting Out				I								
Alienation								I				

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

	Anxiety	Cognitive Impairment	Depression – Hopelessness	Hostility – Anger Control	Hypersensitivity	Psychosis	Somatization	Withdrawal – Alienation	Malingering	Personality Disorder	Self-Harm	Trauma
Anger Control				I								
Health Problems							I					
Hostile Control				I								
Negative Affect	I		I									
Psychotic Features						I						
Social Withdrawal								I				
Suicidal Thinking			I									
Prison Behavior Rating Scale (PBRs)												
Anti-Authority				O								
Anxious-Depressed	O		O									
Dull-Confused		O										
Prison Symptom Inventory (PSI)												
Panic Disorder	I											
Hypersensitivity/External Stimuli					I							
Physical Symptoms							I					
Profile of Mood States (POMS)												
Anger-Hostility				I								
Depression-Dejection			I									
Fatigue-Inertia							I					
Tension-Anxiety	I											
St Louis Univ Memory Scale (SLUMS)		R										
State-Trait Anxiety Inventory (STAI)												
State Anxiety	I											
Trait Anxiety	I											
Structured Inventory of Malingered Symptomatology (SIMS)									I			
Trail Making Test (TMT)		R										
Trauma Symptom Inventory (TSI)												I

Note. C = Clinician rating; I = Inmate self-report; O = Officer rating; R = Researcher administered. Shaded tests not administered at every testing interval.

Self-Report Assessments

A composite score was developed for seven of the eight primary constructs by standardizing scores from the scales on the self-report assessments. Standardized scores were used so that comparisons between constructs could be made more easily and to create a single measure for constructs assessed by multiple self-report assessments. Scores were standardized by centering on the mean of the entire sample at the first assessment and dividing by the standard deviation. A composite score was computed by standardizing each assessment and averaging the standardized scores across the individual assessments as the composite score. Reliabilities for these composites are presented in the discussion of each construct.

Anxiety Construct. Anxiety was measured by eight self-report variables assessed at each time period. The self-report measures used to create the anxiety composite score were the State and Trait subscales from the State-Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, & Lushene, 1970); the Obsessive-Compulsive, Anxiety, and Phobic Anxiety subscales from the Brief Symptom Inventory (BSI; Derogatis, 1993); the Negative

Affect subscale from the Personality Assessment Screener (PAS; Morey, 1997); the Tension-Anxiety subscale from the Profile of Mood States (POMS; McNair, Lorr, & Droppleman, 1992); and the Panic Disorder subscale from the PSI. The following PSI items were included on the Panic Disorder subscale: 2, 6, 10, 13, 16, 17, 20, 25, and 30.

Internal consistency reliabilities were computed for each assessment period for the entire sample. The mean Cronbach's alpha across individual anxiety measures and time periods was .87 (range = .60 to .95). The Cronbach's alphas for the composite ranged between .89 and .91 for the six time periods. Reliabilities were similar across testing intervals, and they were similar to internal consistency estimates from normative samples. Test-retest correlations between sequential time periods ranged between .49 and .86 ($M = .76$) indicating reasonable stability over 3 month assessment periods. The validity coefficients between self-assessments of the anxiety construct indicated evidence for convergent validity, with correlations between measures ranging from .36 to .85 ($M = .65$) across all time periods.

Cognitive Impairment Construct. Cognitive impairment was assessed by two individually administered tests. The Saint Louis University Memory Scale (SLUMS; Tariq, Tumosa, Chibnall, Perry, & Morely, 2006) was used to assess orientation, memory, attention, and executive function. The SLUMS is an 11-item scale and yields a single total score ranging from 0 to 30, where higher scores indicate stronger cognitive abilities. The Trail Making Test (TMT; Reitan, 1958) was used to assess attention. The time required to complete the A (connect sequential numbers) and B (connect alternating numbers and letters) tasks were collected, and the ratio of times (B/A) was used as the attention measure in subsequent analyses.

The SLUMS demonstrated low internal consistency with a mean Cronbach's alpha of .52 across groups and time periods (range = .48 to .60). We could not find comparative information on this newly developed measure. Because this is a multidimensional measure of cognitive function, internal consistency may not be the correct assessment of quality. The correlations between sequential time periods ranged from .38 to .84 ($M = .67$), indicating good test-retest reliability. The Trails B/A ratio and SLUMS total score were negatively correlated (range = -.17 to -.31), as would be expected because the tests are scaled in opposite directions; however, these correlations are fairly small, indicating that these two measures are assessing different cognitive functions. Because of the weak correlations between the SLUMS and TMT, each of these assessments was used individually to assess cognitive impairment rather than combining them to yield a composite score.

Depression-Hopelessness Construct. The depression-hopelessness construct was assessed using five self-report measures. The scales used to create this construct were the Beck Hopelessness Scale (BHS; Beck & Steer, 1993), the BSI Depression subscale (Derogatis, 1993), the PAS Negative Affect and Suicidal Thinking subscales (Morey, 1997), and the POMS Depression-Dejection subscale (McNair et al., 1992).

Internal consistency reliabilities were computed for each assessment period for the entire sample. The mean Cronbach's alpha across depression measures and time periods was .87 (range = .60 to .96). The Cronbach's alpha for the composite ranged between .71 and .77 ($M = .75$) for the six time periods. Internal consistency estimates for the subscales were similar to reliabilities from normative data. The test-retest correlations for the depression-hopelessness composite were strong ($M = .76$, range = .57 to .90) indicating good stability over time. The validity coefficients between self-assessments of the depression-hopelessness construct indicated good convergent validity with estimates ranging from .35 to .89 ($M = .60$) across all measures and time periods.

Hostility-Anger Control Construct. The hostility-anger control composite was assessed using five self-report measures: the BSI Hostility subscale (Derogatis, 1993); the Anger Control, Hostile Control, and Acting Out subscales on the PAS (Morey, 1997); and the POMS Anger-Hostility subscale (McNair et al., 1992).

Internal consistency reliabilities were computed for each scale at each assessment period for the entire sample and a mean Cronbach's alpha of .62 (range = .27 to .94) was obtained. The Cronbach's alphas for the composite ranged between .54 and .61 ($M = .57$) for the six time periods. Although these reliabilities were lower than expected, the smaller internal consistency estimates were for the scales with a small number of items (i.e., PAS subscales with two items) and these reliability estimates are similar to other literature. The correlations between sequential time periods ranged between .56 and .84 ($M = .75$) and suggest that the hostility composite is stable over 3 month periods. The validity coefficients between self-assessments of the hostility-anger control construct were quite variable with validity coefficients ranging between .11 and .84 ($M = .42$) across all measures and time periods; it was the PAS Acting Out and Hostile Control subscales that tended to have lower correlations for this composite. These lower correlations along with the lower composite internal consistency estimates suggest a potential multidimensional construct. Because the composite was stable and the different aspects of hostility-anger control were relevant to this study, all subscales were kept together for the composite measure. Analyses were conducted without the PAS subscales and the overall study results did not change substantially (results available from the authors upon request), thus all measures were included as part of the composite.

Hypersensitivity Construct. The hypersensitivity construct was measured by two self-report measures—the Hypersensitivity to External Stimulus subscale of the PSI and the Interpersonal Sensitivity subscale of the BSI (Derogatis, 1993). Items 1, 7, 31, 34, and 37 were included on the PSI Hypersensitivity to External Stimulus subscale. This composite is assessing two different aspects of hypersensitivity—environmental and interpersonal.

Internal consistency reliabilities computed for each scale at each assessment period for the entire sample indicated highly variable internal consistency estimates (.22 to .86) with a mean estimate of .56. However, examination of each scale showed that the BSI had strong internal consistency estimates (.71 to .86) whereas the PSI has low estimates (.22 to .39). The PSI was created by the researchers and its purpose was to capture variables not measured by existing measures, thus it may not be a homogeneous construct. Although internal consistency estimates of the composite were low (.47 to .61 with $M = .53$), the composite showed good test-retest reliability (.21 to .80 with a mean of .63) and the correlations between these two subscales provided evidence of convergent validity (range = .31 to .44); thus these scales were analyzed in the rest of the study as a composite variable. Analyses using each measure separately are available from the researchers upon request.

Psychosis Construct. The psychosis construct was assessed by three self-report measures. These included the Paranoid Ideation and Psychoticism subscales of the BSI (Derogatis, 1993) and the Psychotic Features subscale of the PAS (Morey, 1997).

Internal consistency estimates for the subscales ranged between .62 and .83 ($M = .77$) and estimates for the composite ranged between .73 and .80 ($M = .78$) indicating adequate internal consistency estimates for this composite and its components. Internal consistency estimates for the subscales were similar to those found with normative samples. Test-retest correlations between sequential time periods indicated strong stability over time (range .52 to .87 with a mean correlation of .71). The validity coefficients between self-

assessments of the psychosis construct provided evidence of convergent validity, ranging between .35 and .79 ($M = .63$) across all measures and time periods.

Somatization Construct. The somatization construct was measured by four self-report assessments, including the Somatization subscale of the BSI (Derogatis, 1993), the Health Problems subscale of the PAS (Morey, 1997), the POMS Fatigue-Inertia subscale (McNair et al., 1992), and the Physical Symptoms subscale of the PSI. Items 5, 8, 11, 15, 19, 24, 27, and 28 were included on the PSI Physical Symptoms subscale.

The mean Cronbach's alpha across somatization measures and time periods was .79 (range = .56 to .94) and for the composite the mean alpha was .77 (range = .73 to .79). Test-retest reliability estimates were strong with correlations ranging between .58 and .86 ($M = .76$). The correlations between the self-assessments of the somatization construct indicated good convergent validity with coefficients ranging between .38 and .67 ($M = .54$) across all measures and time periods.

Withdrawal-Alienation Construct. The withdrawal-alienation construct was assessed using two PAS subscales—Alienation and Social Withdrawal. Internal consistency reliabilities were computed for each assessment period for the entire sample and the median Cronbach's alpha across withdrawal-alienation measures and time periods was .75 (range = .69 to .83). The Cronbach's alphas for the composite ranged between .62 and .71 ($M = .67$) for the six time periods. Correlations between sequential time periods (range = .49 to .87; $M = .68$) indicated stability. Reliabilities were similar across testing intervals and were similar to reliabilities found in the normative samples. The correlations between the subscales used in the withdrawal-alienation construct indicated good convergent validity with coefficients ranging between .45 and .55 ($M = .51$) across time periods.

Malingering. The Structured Inventory of Malingered Symptomatology (SIMS; Widows & Smith, 2005) was used to assess malingering on mental health disorders. Scores on five subscales (Psychosis, Neurologic Impairment, Amnesic Disorders, Low Intelligence, and Affective Disorders) were obtained at each testing period. The SIMS was used in this study as one of the tools to determine if a participant's responses may be truthful.

The subscales of the SIMS tended to be positively correlated (range = .19 to .63; $M = .51$) with each other. The median Cronbach's alpha across malingering subscales and time periods was .76 (range = .50 to .93). Lower correlations and reliability estimates tended to be with the Affective Disorder and Low Intelligence subscales.

The SIMS manual provides cut-off scores to suggest malingering on each of the subscales as well as a total score. The cut-offs were scores greater than 1 for the Psychosis subscale and scores greater than 2 for the Neurological Impairments, Amnesic Disorders, Low Intelligence, and Affective Disorders subscales. The total SIMS scale cut-off included scores greater than 14.

Personality Disorders. The Coolidge Correctional Inventory (CCI; Coolidge, 2004) was utilized to identify personality disorders among individuals. For this study, the CCI was used to assess 14 personality disorders identified in the current and past American Psychiatric Association's (1980, 2000) Diagnostic and Statistical Manuals (DSM): Antisocial, Avoidant, Borderline, Dependent, Depressive, Histrionic, Narcissistic, Obsessive-Compulsive, Paranoid, Passive-Aggressive, Sadistic, Schizoid, and Schizotypal. The CCI also has other subscales to assess, among others, DSM Axis I variables, neuropsychological functioning, and response validity.

For this study we also used the CCI measures of Axis I and personality disorders (Axis II). These variables have been hypothesized as potential predictors of the impact of segregation on psychological distress. Because personality disorders are considered relatively stable constructs, this measure was given only at the baseline assessment period. Therefore, they were not included in the change over time measures. The median Cronbach's alpha across CCI subscales was .75. The Cronbach's alphas ranged between .46 and .88 for different subscales.

Self-Harm Construct. The Deliberate Self-Harm Inventory (DSHI; Gratz, 2001) was used to assess the deliberate self-harm history at the initial assessment. The data obtained from the DSHI was coded to provide a quantitative severity rating based on the frequency of the self harming behavior and whether or not the behavior resulted in hospitalization. This variable was considered to be a potential predictor of outcomes. The baseline assessment was used to assess lifetime history of self-harm and each harming behavior was coded as *having occurred in lifetime* or *not occurring*. Scores were summed across the 17 items for a total score. This measure is meant to be given as an interview rather than a paper-and-pencil test; we modified to fit the testing situation. We had hoped to use this assessment as a repeated measure; however, misunderstanding of instructions did not allow for integrity of the data and only the lifetime assessment of self-harm was used as a potential predictor of outcomes. The internal consistency estimate for the total score was .84 indicating that it is reasonable to sum the 17 indicators into a total score.

Trauma. The Trauma Symptom Inventory (TSI; Briere, 1995) was used to assess the ongoing impact of traumatic history. This measure was selected as a potential predictor of outcomes. This was administered once at the second assessment period. Participants use a 4-point rating scale for frequency of occurrence (0 – *never* to 3 – *often*) of 100 events (e.g., flashbacks, wanting to cry, feeling jumpy) experienced within the last 6 months. Scores are obtained on 3 validity scales and 10 clinical subscales. For this study, the total score was used. The Cronbach's alpha for the total score was .97.

Staff Ratings

Two measures were completed by prison staff to assess the constructs of interest. The Brief Psychiatric Rating Scale (BPRS; Ventura, Lukoff, Nuechterlein, Liberman, Green, & Shaner, 1993) was completed by clinical staff and the Prison Behavior Rating Scale (PBRs; Cooke, 1998) was completed by correctional officers and case managers.

Clinician Ratings. The BPRS (Overall & Gorman, 1962) is a 24-item scale most commonly used to assess patients with psychiatric disorders. It is designed to assess rapidly changing symptoms (Lukoff, Nuechterlein, & Ventura, 1986; Ventura et al., 1993). It measures positive symptoms, general psychopathology, and affective symptoms. Some items can be rated after observation of the patient; others require clinical interview to obtain the patient's self report information. Each of the 24 symptom constructs are rated on a 7-point scale of severity ranging from 1 (*not present*) to 7 (*extremely severe*).

Research has indicated that there are five factors: Thought Disorder, Withdrawal, Anxious-Depressed, Hostility-Suspiciousness, and Activity (Burger, Calsyn, Morse, Klinkenberg, & Trusty, 1997; Hedlund & Vieweg, 1980). The BPRS subscales and total scores demonstrated low internal consistency with alpha estimates ranging between .40 and .66 ($M = .55$); these estimates are lower than those found with normative samples. The correlations between sequential time periods ranged from .23 to .58 ($M = .40$), indicating moderate stability over a 6 month period. The BPRS subscales had low correlations with self-report measures of the same

underlying construct, with validity correlations ranging between .03 and .49 among the corresponding measures and the average validity coefficients of the BPRS with all self-report assessments at .28. The Anxious-Depressed subscale had the strongest correlations with the self-report measures and Withdrawal had the lowest. In general, the BPRS scales had low scores, indicating a possible floor effect (see means in results section) and impacting variability as well as relationships between measures.

Correctional Staff Ratings. The PBRS was designed to assess psychological features common to prison life (Cooke, 1998). The instrument was developed for a British prison population. Therefore, some words that are not common in the U.S. were changed to be culturally appropriate (see Appendix A). Correctional staff rated 36 behaviors using a 4 point rating scale (0 – *never/rarely*; 1 – *sometimes*, 2 – *often*, 3 – *most of the time*) at each of the six time periods. There are three scales: Anti-Authority, Anxious-Depressed, and Dull-Confused. All items were summed to provide a total score. Internal consistencies were good for the PBRS scales with a mean Cronbach’s alpha across groups and times of .93 (range = .90 to .95) for the Anti-Authority scale, .91 (range .90 to .95) for the Anxious-Depressed scale, and .83 (range = .78 to .87) for the Dull-Confused scale. Total score internal consistency estimates ranged from .94 to .95. Test-retest reliabilities were highly variable with correlations ranging between .08 and .50. Correlations between testing periods were lowest from first to second assessments and tended to increase over time, which might be a function of familiarity. Correlations with self-report measures tended to be highly variable and mostly small (-.06 to .46), as they were with clinician ratings (.08 to .16).

Official Records Data

Data from official records were collected primarily from the Department of Corrections Information System (DCIS), which is an administrative database of offender data. Offender characteristics to include demographic history, criminal history and offense data, institutional behavior, and needs levels were electronically downloaded. Inmates are routinely processed through the diagnostic unit upon intake into prison, and data are gathered through various sources including arrest and pre-sentence investigation records, diagnostic interview, and pencil-and-paper tests.

Two standardized tests administered to all inmates at the diagnostic unit were included in this study to describe the population. These were the Level of Service Inventory – Revised (LSI-R; Andrews & Bonta, 1995, 2003) and the Tests of Adult Basic Education (TABE; CTB/McGraw-Hill, 1994). The LSI-R is a semi-structured interview tool that assesses criminal risk, with information verified through official records. The LSI-R total score ranges from 0 to 54 and is used to assign the level of supervision for community-based offenders and to determine allocation of services (Motiuk, Motiuk, & Bonta, 1992). The LSI-R showed moderately strong predictive validity ($r = .31$) for 1-year recidivism rates with Colorado parolees (O’Keefe, Klebe, & Hromas, 1998). The TABE is designed to measure adult proficiency in reading, mathematics, language, and spelling. It gives the information needed to place learners in the appropriate lessons for their particular skill deficiencies. Final scoring of the tests can yield grade equivalent scores. The correlation between the TABE total battery score and the GED average score was .63 (CTB/McGraw-Hill, 2004).

Resulting from the diagnostic assessment process are ratings across different needs levels, including academic, vocational, sex offender, substance abuse, medical, psychological, intellectual disabilities, assaultiveness, and self-destruction. Each level is rated on a 5-point scale, where scores of 3 through 5 indicate problem areas. Similar to the other scales, a psychological rating of 3 or greater indicates the need for mental health services. Levels may be reevaluated during an offender’s incarceration.

Institutional behavior, such as disciplinary violations and involvement in gangs, are recorded electronically over the course of an offender's incarceration. Disciplinary violations are grouped into three categories according to their seriousness. Because patterns were similar when analyzing either violation type or total violations, only totals are reported. There are three levels of gang involvement: member, associate, and suspect. Levels are ascertained by field intelligence officers who rate offenders' involvement across 11 items (e.g., self admission, moniker, gang tattoos, identification by law enforcement). Each item carries a weight ranging from 5 to 20 points, and summative scores determine the degree of gang membership or involvement. To clearly delineate offenders actively involved in gangs, only those scored as gang member were considered to have gang involvement in the following analyses.

Certain data elements were collected only for study participants during the course of their participation in the study. The following were collected and coded for the period of time between each testing interval for each participant: the amount of time spent in various settings (e.g., segregation, GP, hospital), phone records, and mental health crisis data. Additionally, activity logs from paper files for the CSP participants were collected and coded.

Phone records were received electronically from the Colorado Inmate Phone System (CIPS). From these records, researchers coded the number of phone calls attempted, the number of calls completed, and the time spent on the phone across all calls.

Mental health staff is required to make a written report in DCIS following any unscheduled mental health visit or crisis. All reports completed for participants during their participation in the study were reviewed and coded by researchers on a 3-point self-harm scale (1 – *ideation*, 2 – *self-harm behavior*, 3 – *attempted suicide*) and whether or not there was a report of a psychotic symptom during the crisis.

Pod activity records are kept by CSP correctional staff and are updated on a daily basis to provide information on an offender's time outside of his cell for shower, exercise, and porter duties. These forms also track the number of times each offender refused or was not offered these activities. Data were coded to reflect a refusal or an activity not offered on a specific day as well as the actual amount of time the offender spent participating in the activity. When the records were unclear or no information was recorded on a specific day, it was coded as unknown. Researchers coded the pod activity sheets for each offender between each testing interval and summed for the number of refusals, days an activity was not offered, unknowns, and average time spent for shower and exercise activities.

PROCEDURES

Study enrollment began July 2007 and ended March 2009, with final testing of all participants completed in March 2010. The project operated under the approval of the institutional review board at the University of Colorado at Colorado Springs (UCCS).

The research team was notified of AS hearings by the case management supervisor at each facility and of SCCF placements by the clinician who scheduled the facility transfers. Notification typically occurred before the hearings or SCCF placement to give the field researcher maximal lead time. Researchers reviewed electronic records to screen inmates for study eligibility.

Per the UCCS institutional review board, a stipulation was added to provide greater protection to inmates with mental illness. Before consenting them, researchers were required to contact mental health staff, who

in turn were asked to assess whether the offender would be able to understand the consent form and to weigh the study risks against the benefits. Other than the SCCF group, there were rarely issues. However, it was not uncommon for the SCCF clinician to wait several days or even occasionally weeks for a new arrival to stabilize prior to giving researchers approval to consent participants; these inmates were then included in the study. Two inmates were excluded from the study because clinicians did not believe they had the capacity to fully understand the consent process; both were SCCF inmates.

The field researcher was a female university employee who completed the full CDOC training academy and had a CDOC badge that permitted her unescorted access to the facilities. In advance of each visit, the field researcher contacted prison security to arrange visits with specific inmates. All inmates were escorted by security staff to the visiting room, which contained a noncontact booth for inmates in AS or punitive segregation conditions. The field researcher met individually with each inmate to review the consent form, which included the general purpose of the study, voluntary nature of participation, risks and benefits, and remuneration. Inmates were advised that the purpose of the study was to learn about their adjustment to prison and offenders in prisons across the state were being included in this study. Inmates who agreed to participate were given \$10 for each testing interval. Although this amount may at first seem high for AS inmates who do not have an opportunity to earn income, it was important that AS inmates were compensated at the same rate as GP inmates since the activities were exactly the same. Additionally, all deposits into inmate bank accounts were subject to a 30% restitution recovery fee and deposits to inmates with negative balances (common among AS inmates) were subject to a 50% reduction of the deposited amount. Therefore, actual payments ranged from \$2 to \$7.

Inmates were screened for their native language and reading abilities. Although this was done when determining study eligibility, the field researcher further assessed them at the time of consent. The testing battery was not available in alternative languages and it was determined that using interpreters could negatively impact the validity of the tests. However, the field researcher attempted to include inmates with borderline English or reading skills by helping them to understand difficult words. Eighteen inmates across the five study groups were specifically excluded from the study for lacking adequate language or reading abilities.

At the time of consent, the initial test battery was administered. The field researcher instructed participants to read the directions for each test. Instructions were highlighted by researchers when there was an indication on the test to respond with respect to a certain timeframe (e.g., in the past week). The field researcher administered the timed TMT and the SLUMS tests, and she assisted if they had questions, most frequently with the definition of a word. The researcher collected the test packet immediately following its completion, so it was not ever handled by security staff. At the same time, she visually scanned the packet before the inmate was returned to his cell to ensure that he had not inadvertently skipped a test or section of items.

Prior to leaving the facility, the researcher conducted a further review of inmates' responses for indications of intent to harm self or others. There were no items that assessed intent to harm others, but numerous items were identified as potential indicators of suicide ideation. Participants were notified at the time of consent that confidentiality would be broken if they responded affirmatively to any of these items. When participants endorsed a suicidal item, the field researcher notified mental health staff and the principal investigator immediately. Mental health staff then followed up with each case following notification to assess the seriousness or intent to self-harm. There were no participant suicides during the course of the study.

The field researcher distributed the PBRS to housing staff at each testing interval and collected the completed forms upon return visits to the facility. Mental health clinicians were generally notified that a BPRS was needed a couple weeks prior to the researcher testing to give them time to complete the assessment.

In the CSP groups, 18 out of 127 participants were consented and tested prior to their AS hearing. On average, CSP participants completed their initial test 7 days ($SD = 7.3$) after their AS hearing. Thirteen participants in the GP groups were selected from the diversion program (for being at risk of AS placement) and seven were tested prior to an AS hearing. On average, however, GP participants were tested 16 days ($SD = 18.9$) after their hearing or placement into the diversion program. At the time of consent and the initial testing, 43% of inmates had been confined in segregation (40% in AS groups and 3% in GP groups) for an average of 18.2 days ($SD = 18.1$). SCCF participants were tested within 13 days of placement on average ($SD = 8.9$).

Participants' data were kept in two separate databases. The eligibility database tracked the eligible pool of offenders, such as identifying information, current location, date of AS hearing or SCCF placement, expected release date, psychiatric status and clinician approval, selection into study or reason for exclusion, and date of consent or refusal. A testing schedule for study participants was incorporated into the database, which also had reporting capabilities in order to manage the project. A separate database tracked participants' responses to the standardized tests; no identifying information was included in this database other than a secure researcher-assigned identification number. Both databases were stored on a secured server with access restricted to the project researchers.

RESULTS

DATA ANALYSIS PLAN

We first present results that speak to the quality of the research design, addressing issues concerning sampling and group representativeness; comparing those who participated fully, partially, or refused to participate; evaluating the fidelity to confinement conditions; and examining the validity of self-report responses. Following these analyses, we present results addressing the hypotheses of interest. This study had three goals and related hypotheses:

- To determine which, if any, psychological domains are affected, and in which direction, by the different prison environments; it was hypothesized that offenders in segregation would develop an array of psychological symptoms consistent with the SHU syndrome, characterized by elevations across the eight constructs.
- To assess whether offenders with mental illness decompensate differentially from those without mental illness in AS by testing the hypothesis that both groups will get worse over time but that the rate of deterioration would be greater for the mentally ill.
- To compare the impact of AS against other prison conditions by testing the hypothesis that inmates in segregation experience greater psychological deterioration over time compared to inmates in other confinement conditions.

To test the first hypothesis, one sample t-tests are completed to see if study groups are significantly different from normative data on the study measures at each time period.

To test the second hypothesis, analysis of variance statistical techniques are used to assess if the AS groups have differential change over time. Comparisons are made on mean change over time for each construct for the mentally ill and non-mentally ill groups in AS confinement conditions.

To test the third hypothesis, analysis of variance statistical methods are used to assess mean change over time and groups for each construct of interest. In particular, it is of interest to determine whether there is a significant interaction between time and group to indicate that there is differential change over time depending on condition of confinement. An analysis is completed for those with different mental illness status. That is, the mentally ill group in AS is compared to the mentally ill groups in the general prison and in the psychiatric prison, whereas the non-mentally ill in AS are compared to the non-mentally ill in the general prison population.

Mean difference statistical results for all three analyses are supplemented with effect size measures assessing proportion of variance accounted for by the time and group variables. Significant main effects will be further investigated using pairwise comparisons to explore group differences and comparisons between means at consecutive time periods to explore time effects. If there is a statistically significant interaction, simple main effects exploring change over time for each group will be completed. All statistical tests are completed at the .05 significance level.

In addition to the analysis of variance methods to explore mean change over time, regression analysis is used to predict change over time using individual variables as potential predictors. Change over time is assessed by computing an individual slope estimate for each person on each construct of interest. Predictors

include demographic variables, criminal history variables, personality variables, and confinement conditions. Within each section we describe the data analytical tools used to complete the analyses.

SAMPLING

Group Representativeness

Because random assignment procedures were not engaged, comparisons of offender characteristics across variables routinely collected in DCIS were conducted between eligible inmates and study participants to determine the study sample’s representativeness (see Table 2). Some data are dynamic and, therefore, these data represent those that were current for each offender at the point of his eligibility for the study. The institutional behavior measures of disciplinary violations and prior AS placement were collected over their entire incarceration up to study eligibility.

Table 2. Representativeness of CSP Study Groups to Eligible Pool

	CSP MI			CSP NMII		
	Sample (n = 64)	Eligible Pool (n = 232)	p	Sample (n = 63)	Eligible Pool (n = 432)	p
Demographics						
Mean age (SD)	31.2 (9.7)	32.1 (9.2)	n.s.	30.0 (9.9)	30.4 (8.5)	n.s.
Ethnicity/Race			n.s.			n.s.
White	41%	45%		19%	27%	
Hispanic	33%	32%		54%	55%	
African American	19%	17%		22%	16%	
Other	8%	6%		5%	2%	
High school achievement			n.s.			n.s.
HS diploma	12%	12%		10%	13%	
HS equivalency	51%	45%		54%	58%	
Neither	37%	43%		36%	30%	
Test of Adult Basic Education						
Mean reading score (SD)	8.7 (3.6)	8.0 (3.6)	n.s.	7.8 (3.3)	8.6 (3.6)	n.s.
Mean math score (SD)	6.7 (2.5)	6.3 (2.8)	n.s.	6.7 (2.5)	7.2 (3.0)	n.s.
Mean language score (SD)	7.7 (4.0)	7.1 (4.1)	n.s.	7.2 (3.8)	7.6 (3.9)	n.s.
Mean total score (SD)	7.7 (3.5)	7.1 (3.5)	n.s.	7.4 (3.4)	7.8 (3.5)	n.s.
Sentence and Criminal History						
Mean prior incarcerations (SD)	.5 (0.9)	.5 (0.9)	n.s.	.4 (0.8)	.4 (0.7)	n.s.
Mean felony class 1 – 6 (SD)	3.4 (1.1)	3.7 (1.1)	.02	3.2 (1.1)	3.5 (1.1)	n.s.
Mean LSI-R (SD)	35.3 (7.4)	34.8 (6.9)	n.s.	33.1 (5.8)	33.0 (6.6)	n.s.
% Sentenced for violent crime	67%	54%	.03	70%	59%	n.s.
Institutional Behavior						
Mean # disc. violations (SD)	22.0 (27.5)	20.7 (20.1)	n.s.	13.2 (10.8)	13.9 (14.1)	n.s.
% Prior AS placement	38%	38%	n.s.	32%	29%	n.s.
% Gang member	30%	33%	n.s.	43%	45%	n.s.
Need Levels (% scored 3-5)						
% Academic	42%	45%	n.s.	41%	37%	n.s.
% Vocational	83%	82%	n.s.	87%	83%	n.s.
% Medical	23%	17%	n.s.	10%	9%	n.s.
% Substance abuse	83%	80%	n.s.	71%	81%	n.s.
% Sex offender	44%	33%	n.s.	30%	22%	n.s.
% Intellectual disability	11%	10%	n.s.	3%	3%	n.s.
% Anger	69%	61%	n.s.	70%	64%	n.s.
% Self-destruction	34%	25%	n.s.	10%	9%	n.s.

Nonparametric chi-square and *t* test analyses were conducted for both sets of group comparisons. There were no differences between the CSP NMI study sample and eligible pool. The only difference for the CSP MI group was that study participants had a more serious felony offense, as measured by felony class (class 1 is most serious and class 6 is least) and percent with a violent crime, than individuals in the eligible pool.

Refusals

The field researcher asked inmates who refused to participate or who withdrew their consent for their reasons. Half of them gave no reason for doing so. Of those who listed their reasons, 10 inmates stated general disinterest, 6 were skeptical of the research, 4 feared retaliation from their gang or other inmates for participating, 3 listed monetary reasons, and 3 expected imminent release due to an appeal. Chi-square analyses and *t*-tests were conducted between study participants and inmates who refused to participate in the study or withdrew their consent to determine if significant differences existed (see Table 3). The only measured difference between the two groups was that participants had higher LSI-R scores, which indicates higher recidivism risk.

Table 3. Comparison of Refusals to Study Participants

	Refusals (<i>n</i> = 55)	Participants (<i>n</i> = 247)	<i>p</i>
Demographics			
Mean age (<i>SD</i>)	33.2 (10.5)	31.7 (8.9)	n.s.
Ethnicity/Race			
White	44%	40%	
Hispanic	33%	37%	
African American	20%	18%	
Other	4%	5%	
High school achievement			
HS diploma	23%	13%	
HS equivalency	42%	56%	
Neither	36%	32%	
Test of Adult Basic Education			
Mean reading score (<i>SD</i>)	8.2 (3.6)	8.5 (3.5)	n.s.
Mean math score (<i>SD</i>)	7.1 (3.1)	6.6 (2.8)	n.s.
Mean language score (<i>SD</i>)	7.9 (4.0)	7.3 (3.9)	n.s.
Mean total score (<i>SD</i>)	7.8 (3.4)	7.5 (3.5)	n.s.
Sentence and Criminal History			
Mean prior incarcerations (<i>SD</i>)	.3 (0.6)	.5 (0.8)	n.s.
Mean felony class 1 – 6 (<i>SD</i>)	3.2 (1.1)	3.4 (1.1)	n.s.
Mean LSI-R (<i>SD</i>)	31.4 (7.4)	34.3 (7.3)	.02
% Sentenced for violent crime	70%	60%	n.s.
Institutional Behavior			
# Disciplinary violations	16.9 (24.0)	15.8 (16.8)	n.s.
% Prior AS placement	26%	23%	n.s.
% Gang member	22%	28%	n.s.
Need Levels (% scored 3-5)			
% Academic	42%	39%	n.s.
% Vocational	87%	83%	n.s.
% Medical	13%	19%	n.s.
% Substance abuse	70%	79%	n.s.
% Sex offender	35%	31%	n.s.
% Intellectual disability	15%	9%	n.s.
% Anger	72%	60%	n.s.
% Self-destruction	24%	23%	n.s.

Complete and Incomplete Testers

There were 222 participants who completed all testing sessions of the study; thus only 18% of participants did not have all self-report assessments for every time period. There were a number of reasons why participants did not complete all testing periods: some withdrew their consent, others were paroled before the end of the study, and some were not available for a specific testing interval (e.g., out to court for extended trial). There was not a significant differential incompleteness rate across the five study groups, $\chi^2(4, N = 270) = 3.71, p = .45$, with incompleteness rates of 25% for CSP MI, 16% for CSP NMI, 18% for GP NMI, 15% for GP MI, and 18% for SCCF. Comparisons were made between those who did and did not complete all assessments on demographic, background variables, and the dependent variables. To compute a score on the dependent variables, the mean across scores for available time periods was computed for the self-report composites and cognitive variables. Table 4 provides information on these comparisons. Participants who did not complete the entire study had significantly higher self-destruction needs, higher mean hostility composite scores, and lower cognitive function as demonstrated by significantly lower scores on both measures of cognitive performance (SLUMS, Trails B/A).

Table 4. Comparison of Incomplete to Complete Testers

	Incomplete (n=48)	Complete (n=222)	p
Demographics			
Mean age (SD)	31.4 (8.7)	34.1 (10.6)	n.s.
Ethnicity/Race			n.s.
White	49%	39%	
Hispanic	34%	36%	
African American	13%	20%	
Other	4%	5%	
High school achievement			n.s.
HS diploma	12%	19%	
HS equivalency	57%	44%	
Neither	31%	37%	
Test of Adult Basic Education			
Mean reading score (SD)	8.6 (3.5)	7.7 (3.7)	n.s.
Mean math score (SD)	6.6 (2.8)	7.1 (3.3)	n.s.
Mean language score (SD)	7.4 (3.9)	7.5 (4.1)	n.s.
Mean total score (SD)	7.6 (3.5)	7.4 (3.6)	n.s.
Sentence and Criminal History			
Mean prior incarcerations (SD)	0.5 (0.8)	0.4 (0.8)	n.s.
Mean felony class 1 – 6 (SD)	3.5 (1.1)	3.2 (1.2)	n.s.
Mean LSI-R (SD)	34.0 (7.5)	33.3 (7.5)	n.s.
% Sentenced for violent crime	65%	61%	n.s.
Institutional Behavior			
# Disciplinary violations	15.1 (15.4)	22.5 (28.8)	n.s.
% Prior AS placement	22%	27%	n.s.
% Gang member	27%	29%	n.s.
Need Levels (% scored 3-5)			
% Academic	37%	43%	n.s.
% Vocational	84%	85%	n.s.
% Medical	18%	15%	n.s.
% Substance abuse	78%	72%	n.s.
% Sex offender	32%	30%	n.s.
% Intellectual disability	9%	13%	n.s.
% Anger	60%	64%	n.s.
% Self-destruction	21%	36%	.04

	Incomplete	Complete	<i>p</i>
Composites and Cognitive Measures			
Mean Anxiety (<i>SD</i>)	-.18 (.72)	-.08 (.80)	n.s.
Mean Depression-Hopelessness (<i>SD</i>)	-.16 (.73)	-.01 (.76)	n.s.
Mean Hostility-Anger Control (<i>SD</i>)	-.16 (.57)	.02 (.65)	.05
Mean Hypersensitivity (<i>SD</i>)	-.13 (.68)	-.12 (.77)	n.s.
Mean Psychosis (<i>SD</i>)	-.18 (.75)	-.06 (.73)	n.s.
Mean Somatization (<i>SD</i>)	-.15 (.69)	.06 (.82)	n.s.
Mean Withdrawal-Alienation (<i>SD</i>)	-.22 (.63)	-.03 (.70)	n.s.
Mean SLUMS (<i>SD</i>)	23.18 (3.36)	22.02 (3.76)	.03
Mean Trails (<i>SD</i>)	2.82 (0.75)	3.18 (1.12)	.01

The amount of complete data was higher on self-report assessments (85%) than clinician ratings (76%) and correctional staff ratings (57%). There were significant differences in groups' completion rates for clinician ratings with the MI groups having more complete data (CSP MI 70%, GP MI 76%, SCCF 75%) than the NMI groups (CSP NMI 56%, GP NMI 58%). There were also significant differences between completion rates for the correctional staff ratings with the GP groups having less complete data (GP NMI 47%, GP MI 49%) than the other three groups (CSP MI 64%, CSP NMI 75%, SCCF 61%).

Group Fidelity to Conditions of Confinement

Participants remained in their assigned group regardless of later placements throughout the prison system. Table S summarizes the locations of study participants by group and by testing interval (each interval is the period of time between two assessment periods).

One of the challenges of applied research is the researchers' lack of control over the independent variable, which in this case is the condition of confinement. Therefore, all offenders in AS were not confined in segregation for their entire period of participation in the study. Over the course of the study, 15 offenders in the CSP MI group were placed in the specialized OMI program; most completed at least three tests prior to the transfer. Some of the inmates placed in CSP were taken to county jail for a court appearance. Conversely, inmates in the GP groups may have at some time during their study participation been placed in punitive segregation or even AS. There were five GP MI and four GP NMI participants who were placed in AS during the course of their segregation; the remainder of GP inmates who had time in segregation were in punitive segregation. Seven of the nine GP inmates were reclassified to AS primarily between the third and fourth assessment periods, one was reclassified after only two tests, and one was reclassified two days prior to his final test.

Due to the contamination across groups, a separate set of analyses were conducted using only the "pure" cases, which included those who only experienced a single condition of confinement during their entire study participation. There were 26 pure cases in the CSP MI group, 39 in the CSP NMI group, 13 in the GP MI, and 11 in the GP NMI. The *p* values and partial eta-squares for the self-report composites were compared for these pure cases and the original study groups. The SCCF group was not included because those participants were expected to transfer from SCCF once stabilization occurred. A result would be considered different if both the *p* value changed significance and the effect size was not of the same magnitude. Because of the smaller sample size in the pure group analysis, it might be possible for an effect of the same magnitude to no longer be statistically significant, thus we did not count this as a different result. The same pattern of results was found for both samples (total vs. pure) except on the hypersensitivity composite. For

this variable, there was a significant time effect for the entire sample ($p = .001$, $\eta^2 = .026$) demonstrating higher scores at the first assessment compared to all other periods but no significant time effect for the pure sample ($p = .56$, $\eta^2 = .009$). As it does not appear that changing locations was a major explanation for the results, subsequent analyses included all offenders who participated in the study. (Complete statistical results are available upon request from the authors.)

Table 5. Number of Days by Location for each Group at each Testing Interval

Group	Location	Interval 1		Interval 2		Interval 3		Interval 4		Interval 5	
		<i>n</i>	<i>M</i> (days)	<i>n</i>	<i>M</i> (days)	<i>n</i>	<i>M</i> (days)	<i>n</i>	<i>M</i> (days)	<i>n</i>	<i>M</i> (days)
CSP MI (<i>n</i> = 64)	CSP	62	19.8	60	78.3	57	91.2	50	88.2	44	83.0
	Other seg	56	88.9	2	4.5	2	9.5	0	--	2	10.5
	SCCF	1	89.0	0	--	0	--	1	69.0	1	71.0
	GP	4	43.8	3	31.0	7	63.0	8	74.6	15	75.7
	Other ^a	6	32.0	5	12.4	9	8.9	5	20.2	5	14.2
CSP NMI (<i>n</i> = 63)	CSP	59	14.2	57	82.7	56	92.2	56	92.7	54	91.4
	Other seg	57	90.3	2	3.5	2	43.0	0	--	0	--
	SCCF	0	--	0	--	0	--	0	--	0	--
	GP	4	5.0	0	--	0	--	0	--	3	62.3
	Other ^a	5	39.0	5	8.4	2	15.5	4	1.8	1	1.0
GP MI (<i>n</i> = 33)	CSP	0	--	0	--	4	24.3	4	80.5	--	--
	Other seg	9	12.4	7	23.0	9	34.2	6	30.7	--	--
	SCCF	0	--	0	--	0	--	0	--	--	--
	GP	32	87.2	32	89.9	27	84.8	25	88.8	--	--
	Other ^a	0	--	3	1.0	1	1.0	1	11.0	--	--
GP NMI (<i>n</i> = 43)	CSP	0	--	0	--	0	--	0	--	--	--
	Other seg	10	39.8	9	13.3	5	39.6	15	37.9	--	--
	SCCF	0	--	0	--	0	--	1	46.0	--	--
	GP	41	79.2	41	89.1	39	85.3	35	83.7	--	--
	Other ^a	2	23.5	4	11.5	2	12.5	2	2.5	--	--
SCCF (<i>n</i> = 67)	CSP	0	--	0	--	0	--	1	53.0	--	--
	Other seg	1	37.0	4	7.5	4	10.3	8	24.0	--	--
	SCCF	64	77.3	54	79.1	40	92.4	34	80.6	--	--
	GP	7	26.6	21	66.2	24	77.3	29	75.6	--	--
	Other ^a	5	12.0	6	32.5	3	39.7	6	9.7	--	--

Note. Individuals may have multiple locations within a study period, so the *n*'s within a group and interval can be larger than the group sample size.

^a Other included out to court (county jail), in custody of US Marshall, hospital or external medical, community placement, and time in transport.

VALIDITY OF RESPONSES

Most of the assessments used in this study were self-report measures, which always carry the risk of not being completed accurately by the participant. Because of this risk, several measures were collected to assess the validity of individual responses.

During data collection and data entry, responses were scanned for abnormal pattern of responses (e.g., the same response selected for all items). Each person's pattern of response was coded as potentially questionable or not. If the pattern was noticed during data collection, then the participant was questioned about his response pattern and asked to redo the test if he admitted to not being truthful. If the participant said he was being honest and the researcher still did not believe him, she marked the test as questionable. Overall, 12% of participants had a questionable response pattern on any measure at any time period (see Table 6);

however, there were no differences between the number of questionable response patterns across groups, $\chi^2(4, N = 270) = 3.87, p = .42$.

Table 6. Percentage of Participants with Questionable Response Patterns

	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Time 1	5%	0%	6%	0%	2%	2%
Time 2	6%	7%	0%	0%	3%	4%
Time 3	3%	5%	3%	5%	0%	3%
Time 4	5%	5%	0%	3%	0%	3%
Time 5	4%	7%	0%	10%	2%	5%
Time 6	8%	7%	NA	NA	NA	8%

Potential malingering was assessed using the SIMS, a 75-item screening measure for detecting feigned symptoms of psychopathology and cognitive functioning in clinical and forensic settings. A total score and scores on five subscales (i.e., Psychosis, Neurological Impairment, Amnesic Disorders, Low Intelligence, Affective Disorders) were obtained. The SIMS assesses whether respondents endorse atypical, improbable, inconsistent, or illogical symptoms. Scores above the cutoff suggest malingering but may also suggest genuine psychopathology. Eighty-five percent of participants had at least one elevated score across the different subscales of the SIMS (see Table 7). The percentage of participants with elevated scores was significantly different across groups, $\chi^2(4, N = 270) = 56.82, p < .001$, with the MI groups (CSP MI 92%, SCCF 96%, GP MI 97%) demonstrating more elevated scores than the NMI groups and with the CSP NMI group (86%) showing more elevations than the GP NMI group (49%). We also considered using a rule of removing participants if they were elevated on multiple scales; however, multiple elevations within a time period were still high among the mentally ill groups (47% to 62%). Because elevated scores may actually reflect psychopathology, we did not eliminate anyone from the study based on this measure. The SIMS was administered to detect potential malingering and was not intended as an outcome measure, thus there are no further analyses with this measure (Appendix B provides summary statistics for the sample).

We further examined response patterns within the main constructs of interest. Because multiple measures were used for each construct, we computed variability across standardized measures of the same construct in order to see if a person was responding in an inconsistent fashion (see Table 8). For example, inconsistent responses within the depression construct might entail a high score on the BHS but a low score on the BSI Depression scale, where one might expect the pattern of scores to be similar. If the variability score for a participant was greater than two standard deviations from the mean on any composite, responses were examined. Approximately 17% of participants had a value greater than this cutoff. Different rates of inconsistent responses were found across the groups, $\chi^2(4, N = 270) = 10.09, p = .04$, with the lowest incidence of inconsistency for GP MI (9%), CSP NMI (10%), and GP NMI (12%) groups, and higher incidences for CSP MI (20%) and SCCF (27%) groups.

To explore if results were influenced by participants with inconsistent or questionable responses, three sets of analyses comparing group differences on composite variables were completed using (1) all participants, (2) those who did not have a questionable response, and (3) those who did not have inconsistent responses. Removal of persons with questionable or inconsistent responses did not change the overall effects and results, so all participants are used in the analyses for this report (full statistical results are available from the authors upon request).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 7. Elevation on SIMS Scales

	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Affective Disorders						
Time 1	59%	35%	55%	16%	72%	49%
Time 2	60%	28%	58%	12%	69%	47%
Time 3	68%	36%	62%	17%	69%	52%
Time 4	58%	39%	59%	13%	59%	47%
Time 5	41%	29%	55%	13%	65%	41%
Time 6	61%	37%	NA	NA	NA	48%
Neurological Impairment						
Time 1	52%	32%	48%	21%	67%	46%
Time 2	42%	30%	42%	17%	58%	40%
Time 3	42%	33%	56%	12%	56%	40%
Time 4	48%	41%	45%	10%	51%	41%
Time 5	43%	25%	45%	16%	46%	35%
Time 6	41%	24%	NA	NA	NA	32%
Psychosis						
Time 1	47%	29%	46%	7%	76%	43%
Time 2	50%	22%	36%	12%	69%	41%
Time 3	53%	24%	38%	15%	64%	41%
Time 4	52%	23%	34%	10%	61%	39%
Time 5	39%	18%	28%	10%	60%	33%
Time 6	39%	18%	NA	NA	NA	29%
Low Intelligence						
Time 1	11%	5%	3%	2%	12%	7%
Time 2	10%	7%	12%	2%	11%	8%
Time 3	17%	9%	3%	2%	13%	10%
Time 4	13%	9%	10%	8%	10%	10%
Time 5	12%	11%	3%	0%	9%	8%
Time 6	16%	11%	NA	NA	NA	12%
Amnestic Disorders						
Time 1	20%	2%	21%	2%	33%	16%
Time 2	26%	5%	15%	2%	40%	20%
Time 3	23%	5%	12%	0%	31%	16%
Time 4	22%	7%	24%	3%	30%	18%
Time 5	16%	5%	17%	0%	26%	14%
Time 6	16%	4%	NA	NA	NA	9%
Total Score						
Time 1	50%	29%	46%	7%	78%	44%
Time 2	50%	22%	48%	10%	72%	42%
Time 3	57%	26%	53%	7%	60%	42%
Time 4	52%	27%	48%	10%	61%	41%
Time 5	41%	27%	41%	8%	53%	35%
Time 6	45%	24%	NA	NA	NA	34%

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 8. Potentially Inconsistent Responses within a Composite Score

	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Anxiety						
Time 1	0%	0%	0%	0%	0%	0%
Time 2	0%	2%	0%	0%	2%	1%
Time 3	0%	0%	0%	0%	0%	0%
Time 4	0%	0%	0%	0%	0%	0%
Time 5	0%	0%	0%	0%	2%	0%
Time 6	0%	0%	NA	NA	NA	0%
Depression-Hopelessness						
Time 1	0%	0%	0%	0%	0%	0%
Time 2	0%	0%	0%	0%	2%	0%
Time 3	0%	0%	0%	0%	0%	0%
Time 4	2%	0%	0%	0%	0%	0%
Time 5	2%	0%	0%	0%	0%	0%
Time 6	0%	0%	NA	NA	NA	0%
Hostility-Anger Control						
Time 1	5%	0%	0%	0%	2%	2%
Time 2	6%	2%	0%	0%	3%	3%
Time 3	3%	4%	0%	0%	3%	2%
Time 4	5%	2%	0%	0%	5%	3%
Time 5	5%	4%	0%	0%	2%	2%
Time 6	0%	2%	NA	NA	NA	1%
Hypersensitivity						
Time 1	8%	3%	9%	5%	12%	7%
Time 2	6%	2%	9%	0%	8%	5%
Time 3	5%	4%	9%	2%	12%	6%
Time 4	9%	7%	7%	3%	12%	8%
Time 5	4%	2%	10%	5%	9%	6%
Time 6	12%	2%	NA	NA	NA	7%
Psychosis						
Time 1	5%	2%	3%	0%	6%	3%
Time 2	2%	0%	0%	0%	0%	0%
Time 3	0%	0%	0%	2%	0%	0%
Time 4	2%	0%	0%	0%	0%	0%
Time 5	2%	4%	0%	0%	2%	2%
Time 6	4%	2%	NA	NA	NA	3%
Somatization						
Time 1	2%	0%	0%	2%	2%	1%
Time 2	0%	2%	0%	0%	0%	<1%
Time 3	7%	2%	0%	0%	0%	2%
Time 4	3%	2%	0%	0%	0%	1%
Time 5	2%	2%	3%	0%	0%	1%
Time 6	0%	0%	NA	NA	NA	0%
Withdrawal-Alienation						
Time 1	2%	2%	0%	5%	4%	3%
Time 2	2%	3%	0%	5%	3%	12%
Time 3	2%	4%	6%	2%	2%	3%
Time 4	0%	2%	0%	3%	2%	1%
Time 5	4%	2%	0%	3%	2%	2%
Time 6	6%	4%	NA	NA	NA	5%

OFFICIAL RECORD DATA

Several sets of official record data were gathered and coded to use as potential outcomes or predictors of change. It was expected that CSP inmates might experience varying levels of isolation based on the amount of time spent at the different QOL levels, the amount of visits and phone contacts, and out of cell time for showers and recreation. In gathering QOL levels, however, data were obtained from two different sources that had conflicting information. Because of the quality of this data, it was not possible to code or use in this study. Following is a discussion of the other official record data gathered and coded.

CIPS Data

CIPS data were collected on the five study groups by testing interval in order to examine amounts of phone contact. A testing interval consisted of the day the offender tested on their self-report measures through the day before the next battery, generally three months. The following data were collected on each offender for each interval: 1) total number of calls attempted, 2) total calls completed (i.e., offender was able to reach another person), and 3) the average duration in minutes per week of all completed calls. A total of 75 offenders did not have any calls during at least one time period. Though the CSP groups had one more testing interval than the other groups, summary statistics are presented only for the four common intervals for each group. Table 21 provides the mean number (and standard deviation) of total calls attempted, total calls completed, and average duration (minutes/week).

Table 21. Mean (and SD) for Phone Call Data for each Time Interval by Study Group

Group	Measure	Time 1 to 2	Time 2 to 3	Time 3 to 4	Time 4 to 5
CSP MI (n = 55)	Attempted	30.80 (57.72)	9.13 (10.11)	12.06 (14.53)	17.62 (20.94)
	Completed	6.24 (13.11)	2.56 (2.65)	3.58 (4.40)	5.47 (7.89)
	Avg mins/wk	6.86 (15.17)	3.77 (4.17)	4.61 (5.89)	6.92 (3.20)
CSP NMI (n = 56)	Attempted	41.52 (80.14)	14.55 (15.94)	23.46 (76.43)	20.75 (22.25)
	Completed	11.11 (27.52)	3.68 (3.93)	6.86 (7.66)	6.00 (6.38)
	Avg mins/wk	12.54 (32.11)	5.26 (5.46)	9.22 (10.60)	7.96 (8.61)
GP MI (n = 29)	Attempted	86.03 (95.49)	129.45 (312.03)	102.31 (202.69)	432.17 (641.45)
	Completed	22.59 (42.72)	26.62 (59.33)	18.17 (28.94)	97.69 (143.62)
	Avg mins/wk	23.40 (38.42)	31.01 (75.21)	18.82 (32.40)	104.00 (159.71)
GP NMI (n = 38)	Attempted	122.40 (131.74)	118.87 (137.45)	105.34 (136.08)	496.63 (495.10)
	Completed	27.08 (40.75)	24.03 (27.15)	18.08 (33.38)	98.71 (119.22)
	Avg mins/wk	29.78 (47.09)	26.38 (30.26)	21.36 (39.62)	105.40 (131.08)
SCCF (n = 56)	Attempted	59.11 (80.98)	59.29 (76.43)	56.71 (88.29)	241.13 (325.30)
	Completed	14.41 (27.22)	15.36 (30.63)	13.07 (30.22)	58.12 (125.70)
	Avg mins/wk	15.85 (27.93)	13.94 (26.11)	11.39 (27.67)	52.87 (111.57)

During the course of the project, important changes were made to the CIPS program. On July 1, 2008, the CIPS pricing was changed so that all offender calls generated from any CDOC facility dialing someone within the continental United States cost the same price. Previously, it was more costly for an offender to make a phone call to someone located outside of Colorado. Additionally, in July of 2007, one trial pod at CSP changed how offenders were able to access the phone system by providing cordless phones that inmates were able to use in their cells. After the trial period ended, the remainder of the facility transitioned to the cordless phone system in July 2009. This change allowed prisoners at CSP to access phones more frequently. Prior to the introduction of the cordless phones, inmates were required to be escorted by two staff members from their cells to the day hall where they would be tethered near the phone. This method is highly

staff intensive and because of other required staff duties, staff were not always able to escort offenders to the phone when requested.

The changes in the CIPS program makes interpretation of change over time difficult; it also impacts group comparisons because group assessments did not occur evenly over the study period (e.g., because there were a smaller number of potential CSP MI participants, we started data collection activities earlier for that group). Thus, further statistical analyses were not completed on these data because it is difficult to know to what causes any potential findings could be attributed.

Pod Activity Data

Pod activity data were collected on the two CSP groups. Pod activity data were gathered from records that housing staff at CSP keep on every offender at the facility to track offenders' exercise and shower habits. Researchers were provided access to this data for use in the study. Data were coded by testing interval (i.e., activities that occurred from one testing period until the next testing period). The following data were collected on each offender: 1) number of days each offender refused an offer to exercise and/or shower, 2) number of days each offender was not offered exercise and/or shower, 3) average number of hours per week an offender participated in exercise and/or shower, 4) the number of days the prisoner participated in exercise and/or shower, and 5) the number of unknowns for recreation and/or shower in that time period. Pod activity data also track an inmate's work record, but because so few participants held jobs during the study period, this data was not included. Table 22 provides the mean number (and standard deviation) for each variable by group and testing interval.

Table 22. Summary Statistics for Pod Activities (Exercise and Showers)

Activity	Interval:	Time 2 to 3		Time 3 to 4		Time 4 to 5		Time 5 to 6		
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Exercise	Activity									
	CSP MI (<i>n</i> = 61)	Days refused	26.02	12.43	29.52	16.38	30.18	17.47	29.74	19.38
		Days not offered	1.85	2.06	2.39	1.92	2.92	2.50	3.28	3.54
		Hours per week	0.71	1.02	0.98	1.17	0.98	1.26	1.00	1.19
		Days of activity	9.23	12.33	13.41	15.61	14.10	16.99	14.54	17.17
Unknown		43.95	12.08	47.69	12.59	46.30	15.34	46.1	23.34	
CSP NMI (<i>n</i> = 61)	Days refused	22.10	13.35	23.28	14.98	28.25	15.73	29.48	16.08	
	Days not offered	2.16	2.05	2.80	1.78	2.69	2.16	4.59	5.27	
	Hours per week	1.41	1.05	1.41	1.22	1.48	1.16	1.68	1.24	
	Days of activity	18.00	13.23	20.69	16.46	22.07	16.12	23.66	17.39	
	Unknown	41.49	9.80	44.20	12.29	39.82	6.99	37.39	9.47	
Shower	Activity									
	CSP MI (<i>n</i> = 61)	Days refused	9.90	10.36	12.74	12.76	12.61	12.55	12.28	12.97
Days not offered		1.30	1.54	1.61	1.57	1.90	2.01	2.34	3.14	
Hours per week		1.41	0.59	1.43	0.65	1.37	0.68	1.22	0.64	
Days of activity		36.07	14.61	39.69	17.01	39.49	19.06	36.44	18.91	
Unknown		33.79	14.02	38.98	15.78	39.49	17.38	42.59	26.65	
CSP NMI (<i>n</i> = 61)	Days refused	6.82	9.74	7.02	9.35	10.00	12.20	10.36	10.69	
	Days not offered	1.44	1.36	1.77	1.43	1.93	1.97	3.52	5.09	
	Hours per week	1.75	0.42	1.75	0.51	1.70	0.53	1.73	0.58	
	Days of activity	44.07	9.77	48.95	12.58	48.70	12.80	50.13	14.92	
	Unknown	31.41	6.13	33.23	8.46	32.18	5.02	31.1	7.19	

Due to problems with coding this data, it was not felt that further analyses could be run. As was previously noted, the data were collected from records that CSP staff were already keeping on offenders and it was often difficult to decipher and/or interpret the records. For example, if a variable was left blank it was not known if that meant the offender was not offered the activity or if he refused to participate resulting in much of the data being coded as unknown.

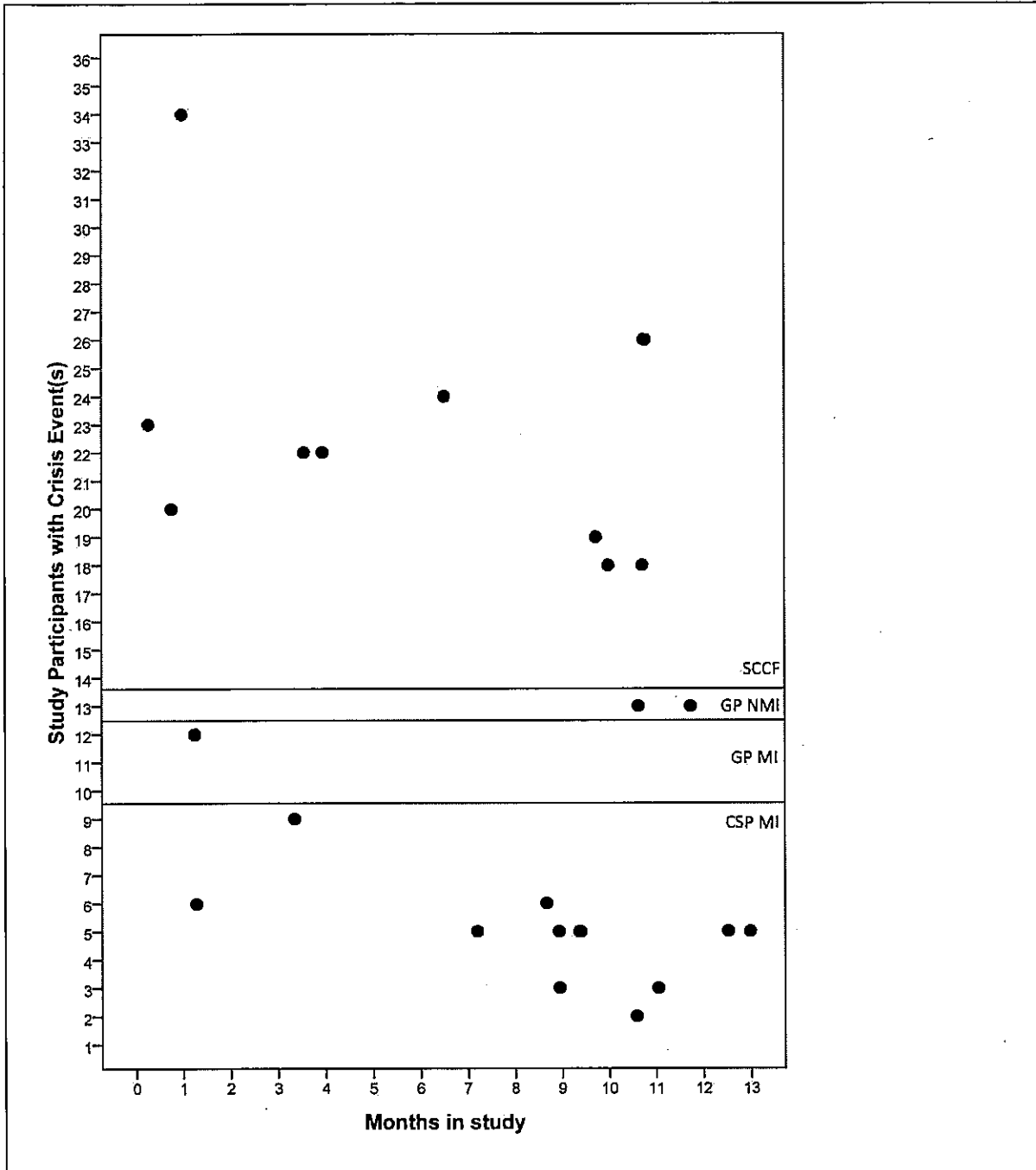
Mental Health Crisis Data

Any situation that is not a scheduled appointment and requires immediate psychological intervention is considered a crisis event; crisis events are documented by clinicians in DCIS. For this study, these events were reviewed and coded for whether there was self-harming ideation or behavior and whether there was a report of an inmate experiencing at least one symptom commonly associated with psychosis. A total of 36 participants had a self-harming ideation or behavior or a report of altered thought patterns commonly associated with psychosis (see Figures 29 and 30). The self-harm data were coded into three categories to indicate a range from self-harm ideation to suicide attempt. Psychotic symptoms were reported as a single category, but the researchers used a low standard in coding psychotic symptoms. If there was any mention of hallucinations or delusional thoughts, even if not observed by the clinician or if denied by the offender, the crisis event was coded as having a psychotic symptom. For example, one participant had threatened self-injurious behavior but also reported the presence of visual hallucinations in the past; this was also coded as a psychotic symptom even though the clinician stated that there was no evidence for psychosis. It should be noted that some events involved both symptoms of psychosis and self-harming ideation/behaviors; therefore case numbers represent the same person on both graphs.

There are several limitations of these data. These include that self-harming ideation/behavior or psychotic symptoms could have occurred without staff's knowledge, offenders may have discussed or exhibited thoughts or behaviors on these dimensions during regularly scheduled mental health appointments, and offenders' self-harming histories prior to study entry were unknown. For example, it was clear from the crisis notes that an individual with numerous crisis events had a long history of self-harming behavior and SCCF placements prior to his enrollment in the study; this is not reflected in the data. Furthermore, the reason for the self-harming ideation/behavior is not captured in the graphed data, but the reasons vary widely. As an example, one offender threatened suicide because he did not want to be removed from CSP to be placed in a new program for offenders with mental illness located at a lower security facility. In another example, one person reported self-harming behavior due to a recent automobile accident where several family members died. Therefore, without more information, it is not possible to attribute the reasons for their mental health crisis to their confinement setting.

We were interested in including the crisis data as an outcome measure in the change over time analyses in order to determine if the occurrence of crises was impacted by confinement conditions, mental status, and time. Because the number of participants who experienced a crisis event was so small, it was not possible to include this variable as an outcome measure in the change over time analyses. These data raise more questions than they provide answers; it was determined that further case study of participants' mental health histories was outside the scope of the current research.

Figure 30. Crisis Events by Individual Participants who had a Psychotic Symptom



GROUP COMPARISONS*Offender Characteristics*

Study groups differed from each other at entry into the study in a number of statistically significant ways (see Table 9). Some differences were consistent with their AS placement or the mental health needs of the groups. The two CSP groups were more likely to have a prior AS placement and to have higher anger needs. The three mentally ill groups (CSP MI, GP MI, SCCF) had higher needs for medical and intellectual disability services. The CSP MI and SCCF groups had higher self destruction needs and were less likely to be gang members. The two GP groups had the lowest rates of sex offender treatment needs. Finally, ethnic/racial composition was different for each of the groups, with more whites in the GP MI and SCCF groups and more Hispanics in the CSP NMI group.

Table 9. Study Group Comparisons on Offender Characteristics

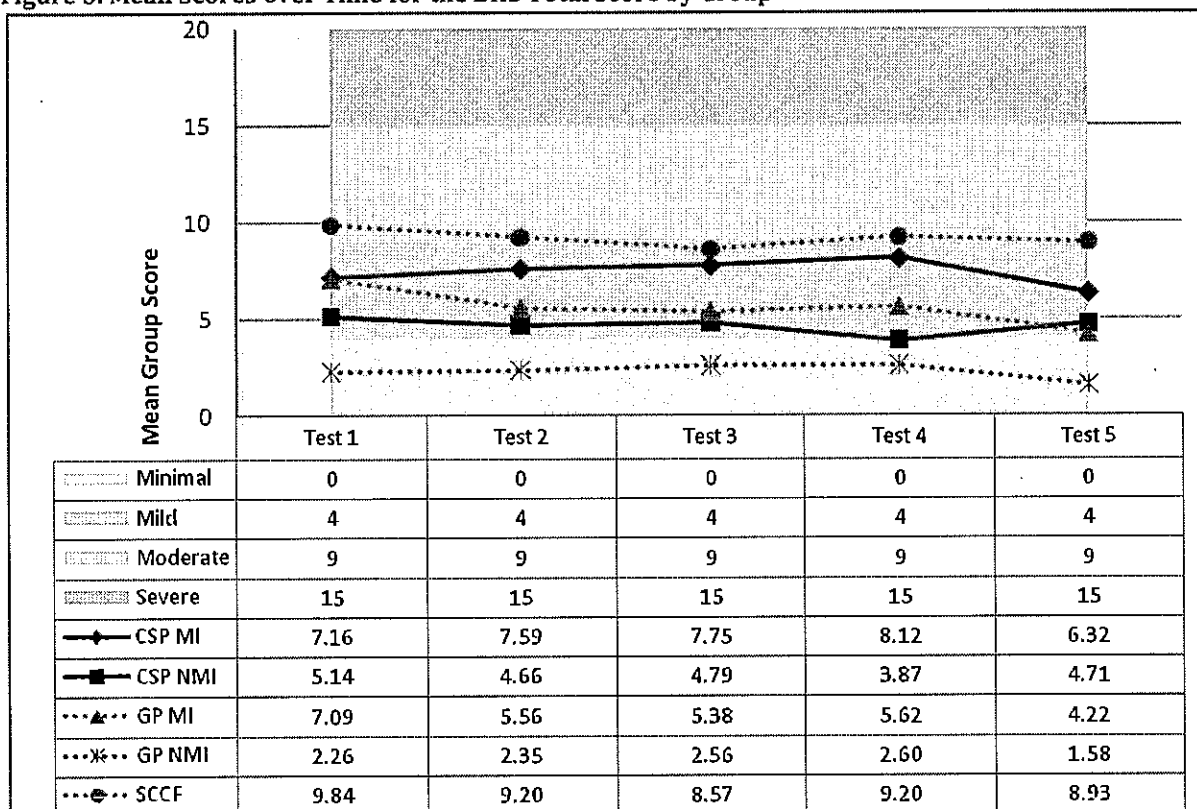
	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	<i>p</i>
Demographics						
Mean age (<i>SD</i>)	31.2 (9.7)	30.0 (9.9)	30.2 (7.8)	33.5 (7.5)	33.9 (8.7)	n.s.
Ethnicity/Race						.01
White	41%	19%	52%	40%	55%	
Hispanic	33%	54%	39%	33%	24%	
African American	19%	22%	9%	26%	15%	
Other	8%	5%	0%	2%	6%	
High school achievement						n.s.
HS diploma	12%	10%	7%	18%	19%	
HS equivalency	51%	54%	58%	63%	50%	
Neither	37%	36%	36%	20%	31%	
Test of Adult Basic Education						
Mean reading score (<i>SD</i>)	8.7 (3.6)	7.8 (3.3)	8.7 (3.0)	10.2 (3.0)	7.7 (3.7)	.01
Mean math score (<i>SD</i>)	6.7 (2.5)	6.7 (2.5)	6.9 (2.6)	7.1 (3.3)	6.5 (3.4)	n.s.
Mean language score (<i>SD</i>)	7.7 (4.0)	7.2 (3.8)	7.9 (3.5)	7.8 (4.1)	7.0 (4.2)	n.s.
Mean total score (<i>SD</i>)	7.7 (3.5)	7.4 (3.4)	7.8 (3.0)	8.3 (3.8)	7.1 (3.7)	n.s.
Sentence and Criminal History						
Mean prior incarcerations (<i>SD</i>)	0.5 (0.9)	0.4 (0.8)	0.4 (0.6)	0.5 (0.7)	0.7 (0.9)	n.s.
Mean felony class 1 – 6 (<i>SD</i>)	3.4 (1.1)	3.1 (1.1)	3.6 (1.1)	3.5 (1.0)	3.6 (1.0)	n.s.
Mean LSI-R (<i>SD</i>)	35.3 (7.4)	33.1 (5.8)	35.4 (8.6)	31.8 (7.0)	33.7 (8.5)	n.s.
% Sentenced for violent crime	67%	70%	61%	54%	54%	n.s.
Institutional Behavior						
# Disciplinary violations	22.0 (27.5)	13.2 (10.8)	17.2 (15.3)	16.0 (15.5)	14.2 (16.6)	n.s.
% Prior AS placement	38%	32%	27%	19%	0%	<.001
% Gang member	30%	43%	21%	33%	10%	<.001
Need Levels (% scored 3-5)						
% Academic	42%	41%	39%	26%	40%	n.s.
% Vocational	83%	87%	88%	77%	85%	n.s.
% Medical	23%	10%	18%	7%	27%	.02
% Substance abuse	83%	71%	91%	78%	71%	n.s.
% Sex offender	44%	30%	24%	14%	37%	.02
% Intellectual disability	11%	3%	9%	0%	20%	.01
% Anger	69%	70%	56%	51%	52%	n.s.
% Self-destruction	34%	10%	16%	10%	39%	<.001

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or views expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

NORMATIVE COMPARISONS

Because we used standardized assessments, normative data were available to compare to the study groups. Normative means were based on non-clinical male samples when available; a general adult sample was used if a male sample was not available. If only clinical sample normative data were available then those were used. Figures 3 to 10 provide the means over time for the measures with total scores for each study group along with highlighted cutoff score ranges. Each graph shows the possible range of scores on the y axis. Figures 3 to 10 are presented for visual reference only; analyses are conducted in later sections. Normative comparisons for subscales used in this study are available in Appendix B.

Figure 3. Mean Scores over Time for the BHS Total Score by Group



This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Figure 4. Mean Scores over Time for the BSI Global Symptom Index by Group

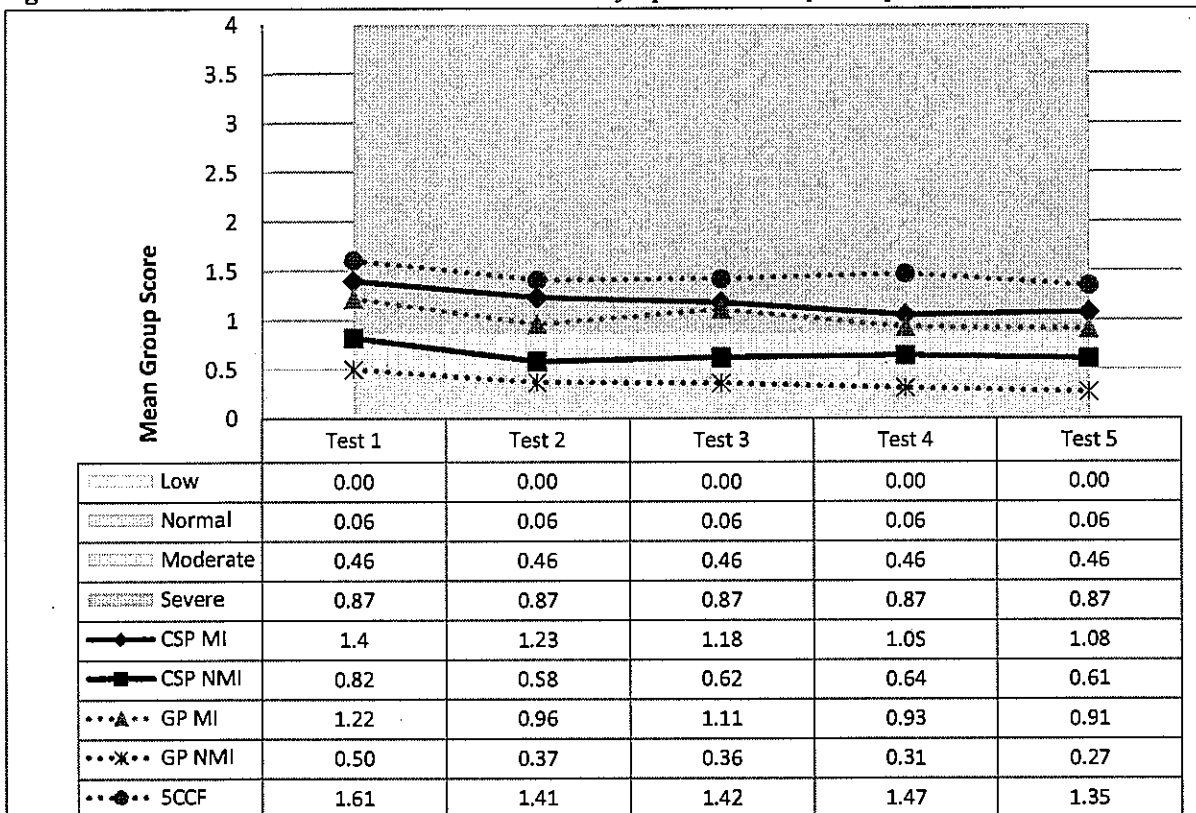
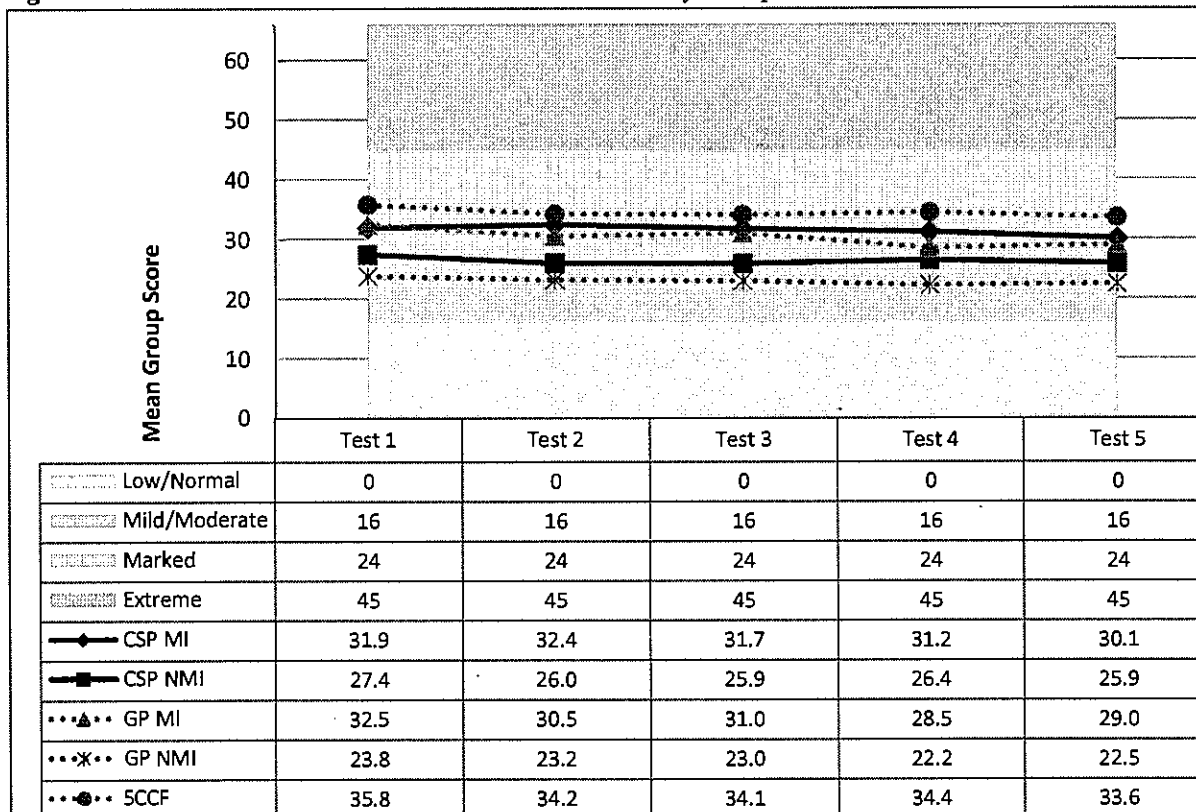


Figure 5. Mean Scores over Time for the PAS Total Score by Group



This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions of points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Figure 6. Mean Scores over Time for the POMS Total Score by Group

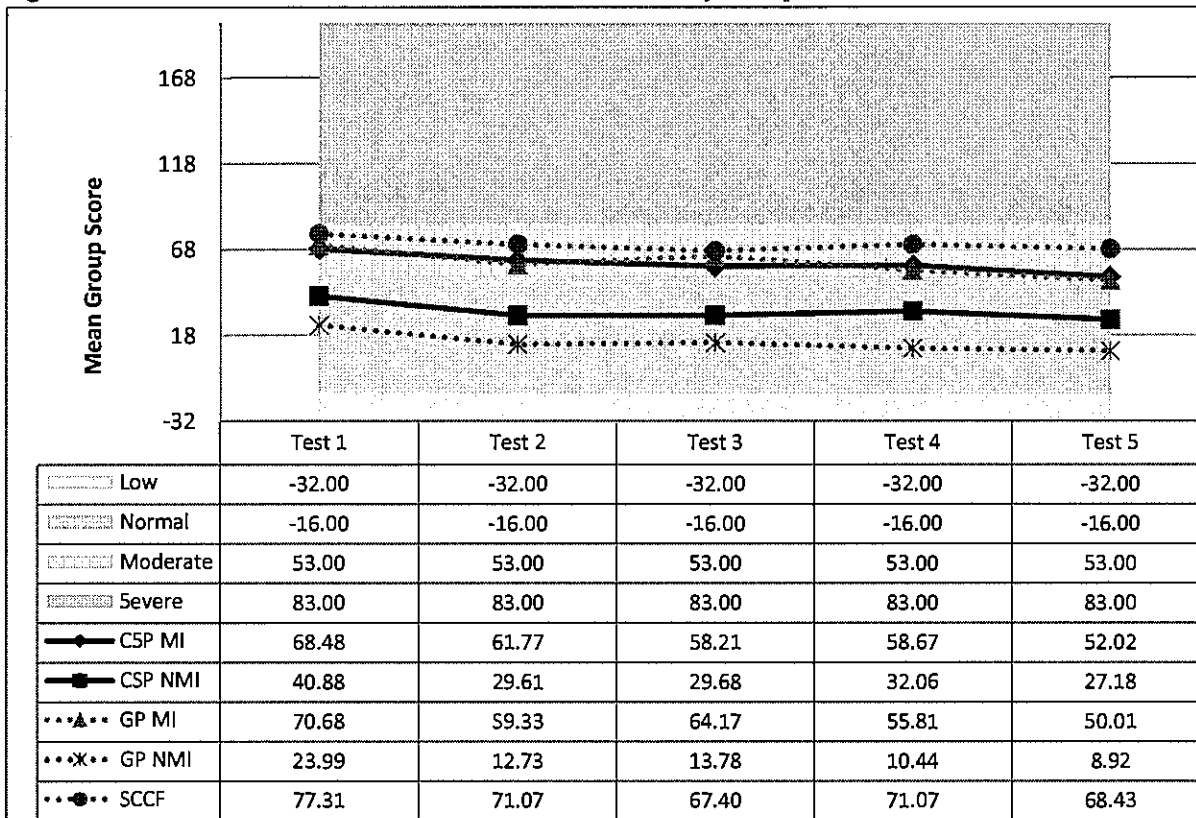
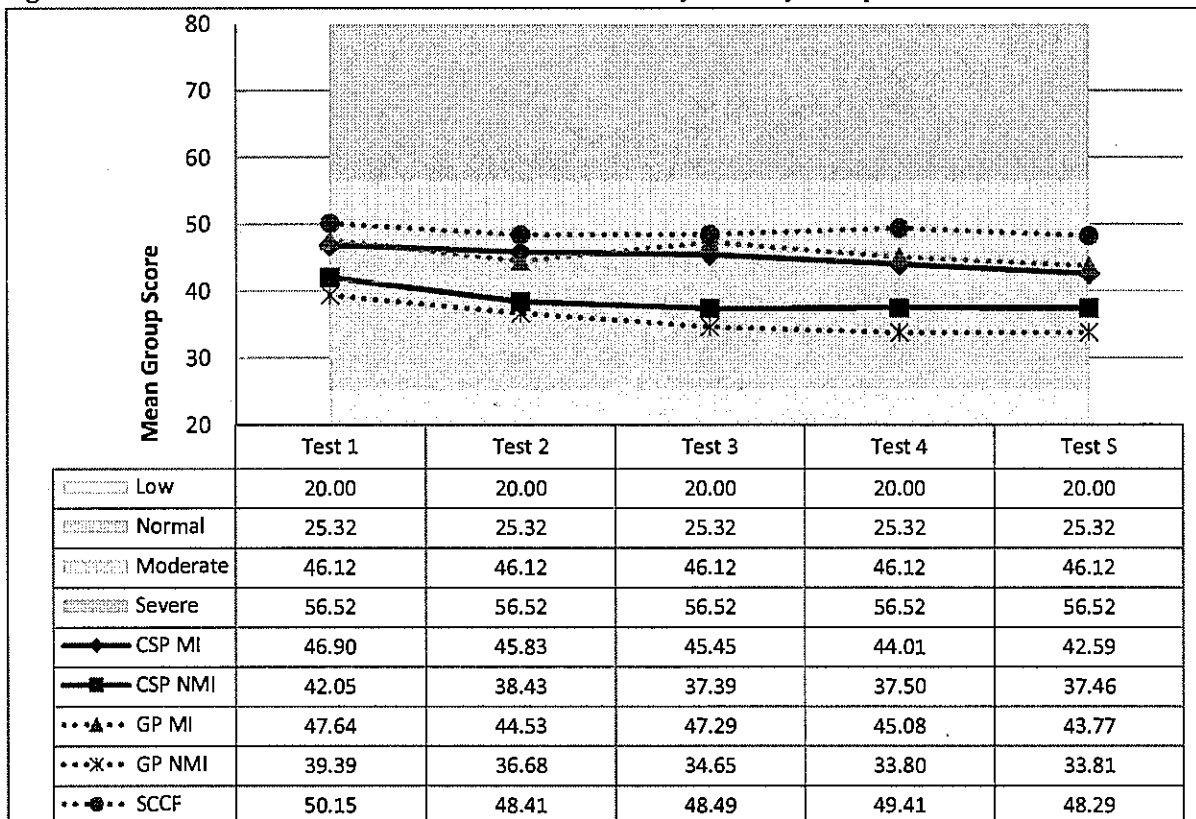


Figure 7. Mean Scores over Time for the STAI State Anxiety Score by Group



This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Figure 8. Mean Scores over Time for the STAI Trait Anxiety Score by Group

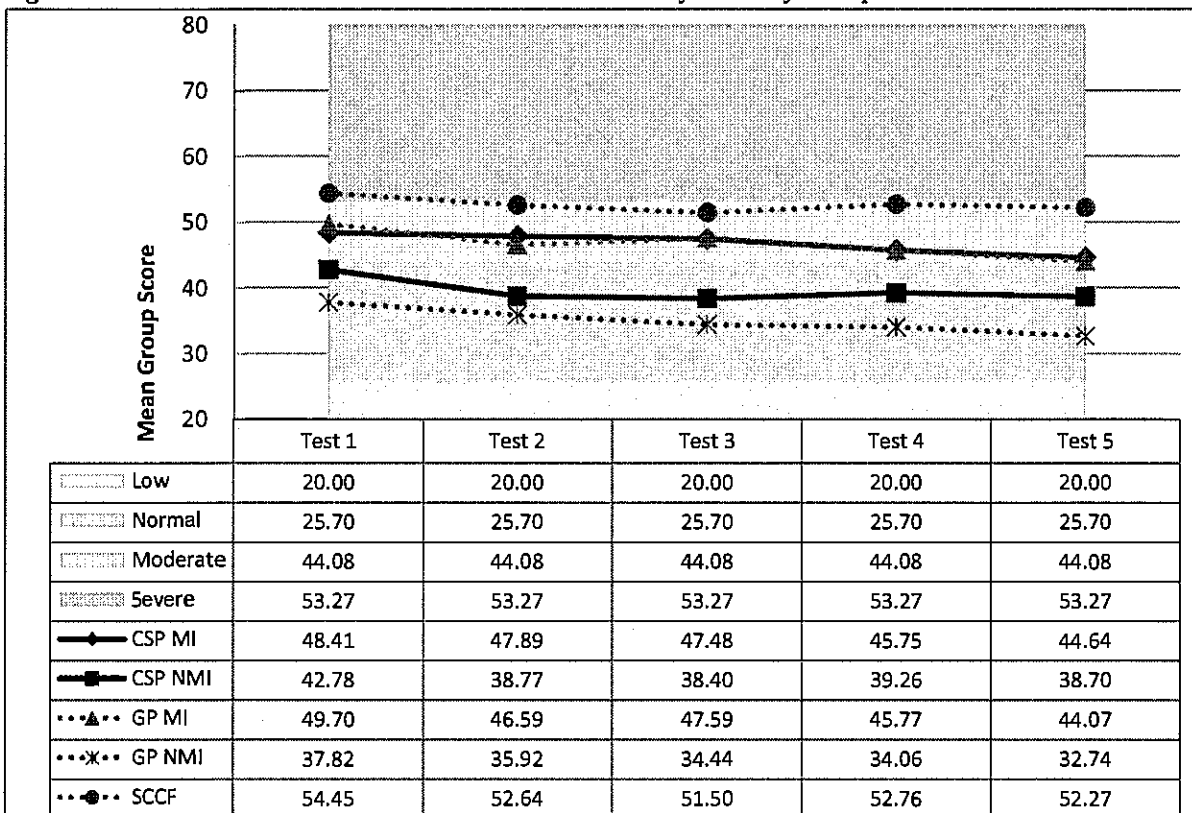
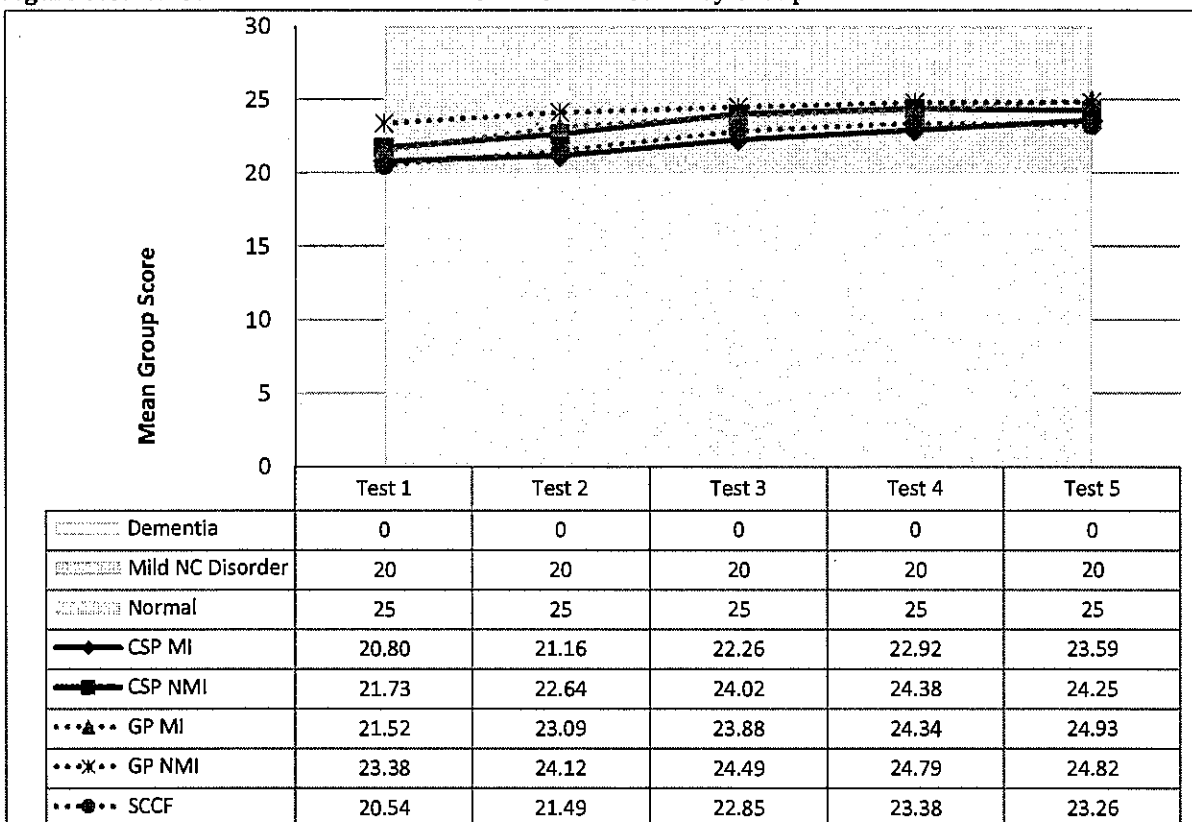
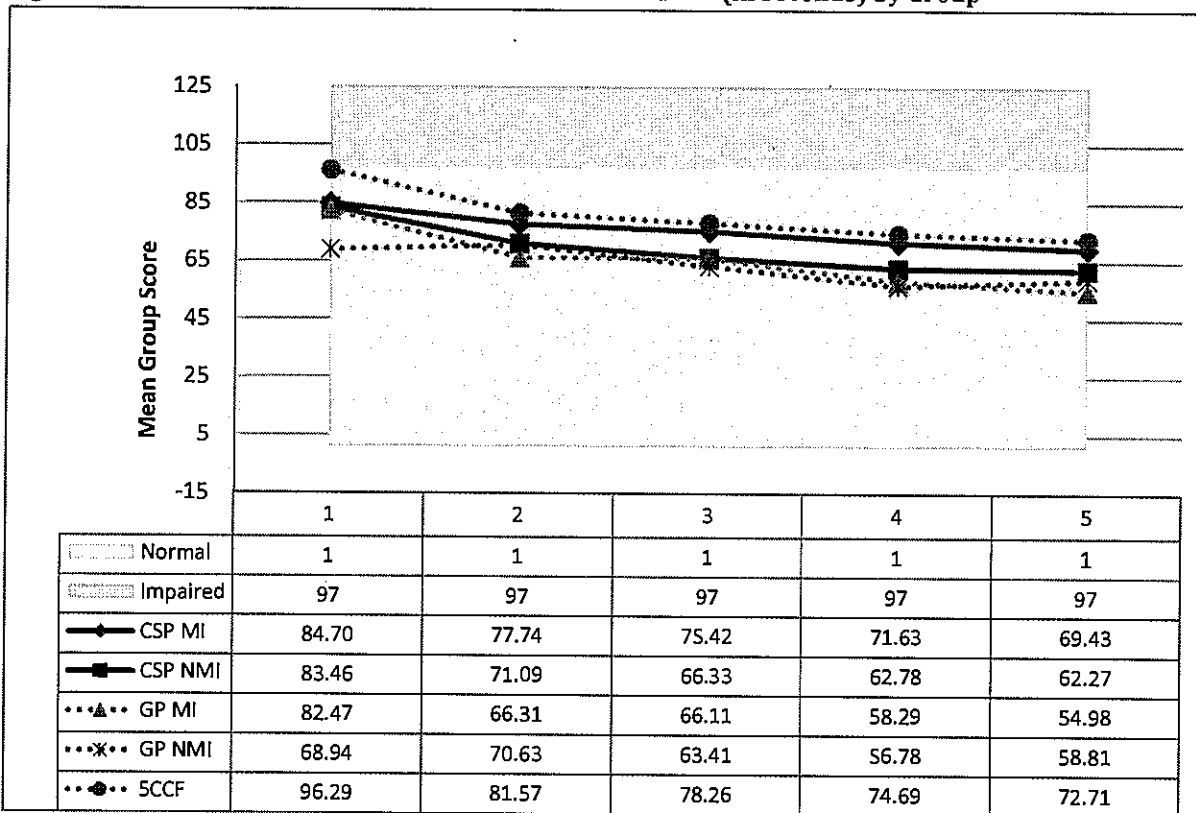


Figure 9. Mean Scores over Time for the SLUMS Total Score by Group



This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions of points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Figure 10. Mean Scores over Time for the Trails Task B Time (in Seconds) by Group



CHANGE OVER TIME

To compare change over time on the psychological measures, three sets of analyses comparing mean change over time were completed—comparisons between the two AS groups (CSP MI vs. CSP NMI) across the six time periods, comparisons of the NMI groups (CSP NMI vs. GP NMI) across the five common time periods, and comparisons of the MI groups (CSP MI vs. SCCF vs. GP MI) across the five common time periods. Mixed design analysis of variance was used to analyze the data for all participants who had data on the composite scores. Huyn-Feldt correction factors were used to adjust the degrees of freedom due to lack of sphericity for the within subject factors. Partial eta-square, providing the percentage of variance explained, was used as an effect size measure (represented by η^2 in the tables). An effect was considered small if it accounted for 1% to 5% of the variance, medium if it accounted for 6% to 14% of the variance, and large if it accounted for 15% or more of the variance. A significance level of .05 was used to determine a statistically significant effect. In addition to mean comparisons over time, a slopes analysis was completed in which slopes were computed for each individual to represent rate of change over time and then comparisons were made between groups.

As a reminder, higher scores on self-report composites, Trails derived scores, correctional staff ratings, and clinician ratings indicate worse performance (e.g., more depression, more anxiety), whereas higher scores indicate better cognitive performance on the SLUMS. Composites are standardized scores and indicate deviation from the first assessment period scores.

Comparisons between CSP Groups

A key purpose of the study was to compare segregated inmates with mental illness to those without mental illness to determine if AS has a differential impact on participants with different mental health needs. Participants were compared across the six time periods. The first assessment was completed while participants were awaiting a hearing for potential placement in AS. The second assessment occurred within 2 weeks of being placed in CSP. The third through sixth assessments were completed approximately every three months following placement in CSP, with the sixth assessment at one year post-placement in CSP.

Comparisons on Self-Report Measures. Comparisons between the two CSP groups were made on the seven composite scores and the two cognitive measures. The summary statistics (mean and standard deviation) for each group at each time period on the composites and cognitive measures are given in Table 12 and the inferential statistics (F values and partial eta-squared) are given in Table 13. Across all seven mental health composites, the MI group scored statistically higher than the NMI group indicating that there was more psychological distress for the MI groups. The effect sizes for the differences between groups vary across composites with large effects for anxiety, depression-hopelessness, and somatization composites. The NMI group had significantly higher average scores on the SLUMS measure, although this was a small effect. There was no significant difference between the groups on the Trails derived score.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 12. Summary Statistics on Self-Report Measures across 6 Time Periods for the Two CSP Groups

Variable	<i>M</i> ₁	<i>SD</i> ₁	<i>M</i> ₂	<i>SD</i> ₂	<i>M</i> ₃	<i>SD</i> ₃	<i>M</i> ₄	<i>SD</i> ₄	<i>M</i> ₅	<i>SD</i> ₅	<i>M</i> ₆	<i>SD</i> ₆
Anxiety												
CSP MI (<i>n</i> = 48)	.30	.84	.19	.83	.20	.88	.03	.89	-.05	.88	-.17	.78
CSP NMI (<i>n</i> = 53)	-.43	.53	-.58	.56	-.55	.61	-.54	.60	-.58	.58	-.60	.59
Both Groups	-.06	.78	-.19	.80	-.18	.84	-.25	.80	-.32	.78	-.38	.72
Depression-Hopelessness												
CSP MI (<i>n</i> = 48)	.20	.76	.08	.72	.09	.82	.07	.87	-.09	.82	-.17	.73
CSP NMI (<i>n</i> = 53)	-.37	.53	-.46	.58	-.46	.59	-.48	.56	-.50	.57	-.56	.55
Both Groups	-.09	.71	-.19	.70	-.18	.76	-.21	.77	-.30	.73	-.37	.67
Hostility-Anger Control												
CSP MI (<i>n</i> = 48)	.20	.64	-.01	.63	-.01	.68	-.07	.74	-.07	.73	-.19	.74
CSP NMI (<i>n</i> = 53)	-.15	.60	-.35	.52	-.25	.59	-.24	.59	-.25	.62	-.23	.68
Both Groups	.03	.64	-.18	.60	-.13	.64	-.15	.69	-.16	.67	-.21	.70
Hypersensitivity												
CSP MI (<i>n</i> = 48)	.24	.84	.15	.88	.23	.94	.02	.89	.02	.84	-.07	.80
CSP NMI (<i>n</i> = 53)	-.21	.62	-.53	.67	-.42	.74	-.30	.75	-.37	.72	-.40	.65
Both Groups	.01	.76	-.19	.84	-.10	.90	-.14	.83	-.18	.80	-.23	.74
Psychosis												
CSP MI (<i>n</i> = 48)	.18	.78	.06	.78	.17	.96	-.10	.88	.06	.97	-.09	.91
CSP NMI (<i>n</i> = 53)	-.26	.71	-.50	.61	-.46	.64	-.40	.70	-.42	.75	-.46	.64
Both Groups	-.04	.78	-.22	.75	-.15	.89	-.25	.80	-.18	.89	-.27	.80
Somatization												
CSP MI (<i>n</i> = 48)	.23	.80	.12	.79	.22	.76	.05	.77	-.03	.83	-.16	.69
CSP NMI (<i>n</i> = 53)	-.45	.62	-.54	.64	-.48	.59	-.48	.55	-.53	.61	-.55	.60
Both Groups	-.11	.78	-.21	.78	-.13	.76	-.21	.71	-.28	.76	-.36	.67
Withdrawal-Alienation												
CSP MI (<i>n</i> = 48)	.12	.85	.37	.89	.29	.81	.33	.81	.14	.79	.18	.85
CSP NMI (<i>n</i> = 53)	-.31	.59	-.15	.84	-.17	.81	-.02	.82	-.08	.85	-.04	.78
Both Groups	-.10	.76	.11	.90	.06	.84	.16	.83	.03	.82	.07	.82
SLUMS												
CSP MI (<i>n</i> = 48)	20.75	5.59	20.88	4.91	22.35	4.72	22.75	4.50	23.60	4.07	23.85	4.58
CSP NMI (<i>n</i> = 53)	21.85	3.49	22.55	3.64	24.04	3.28	24.40	2.94	24.34	3.23	25.26	2.90
Both Groups	21.30	4.62	21.70	4.35	23.20	4.10	23.57	3.83	23.97	3.65	24.56	3.84
Trails B/A												
CSP MI (<i>n</i> = 48)	2.98	.95	2.61	1.07	2.61	.88	2.44	.98	2.58	.60	2.34	.74
CSP NMI (<i>n</i> = 53)	3.10	1.54	2.78	.79	2.94	1.02	2.44	.98	2.58	.60	2.70	.88
Both Groups	3.04	1.30	2.70	.92	2.77	.97	2.64	1.05	2.59	.74	2.52	.84
PBRs Anti-authority												
CSP MI (<i>n</i> = 43)	7.12	7.22	5.68	5.46	6.67	7.78	3.40	5.04	4.68	6.68	3.70	5.02
CSP NMI (<i>n</i> = 49)	8.06	7.44	5.56	6.41	4.00	4.95	4.41	4.87	3.20	4.45	2.52	4.46
Both Groups	7.59	7.32	5.62	5.96	5.34	6.53	3.91	4.95	3.94	5.62	3.11	4.74
PBRs Anxious-Depressed												
CSP MI (<i>n</i> = 41)	6.03	6.21	3.05	3.95	3.44	4.53	2.66	3.37	2.67	3.83	2.90	3.75
CSP NMI (<i>n</i> = 49)	3.34	4.36	2.00	3.06	1.49	2.51	1.61	3.45	1.77	3.31	1.51	2.96
Both Groups	4.68	5.42	2.52	3.51	2.46	3.68	2.14	3.44	2.22	3.56	2.21	3.93
PBRs Dull-Confused												
CSP MI (<i>n</i> = 42)	4.45	4.46	2.14	2.93	2.71	3.15	2.82	3.70	2.24	3.03	2.42	3.03
CSP NMI (<i>n</i> = 47)	1.91	2.42	1.32	2.16	.87	1.21	1.02	1.50	1.05	1.59	1.38	2.68
Both Groups	3.18	3.74	1.73	2.57	1.79	2.49	1.92	2.89	1.64	2.44	1.90	2.88
PBRs Total												
CSP MI (<i>n</i> = 41)	17.73	16.22	11.32	10.35	13.21	13.72	9.42	9.94	9.60	11.34	8.57	9.80
CSP NMI (<i>n</i> = 49)	13.33	12.42	9.04	10.22	6.47	7.63	7.13	7.63	6.32	7.41	5.42	8.31
Both Groups	15.53	14.36	10.18	10.29	9.84	11.29	8.28	8.78	7.96	9.49	6.99	9.10

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

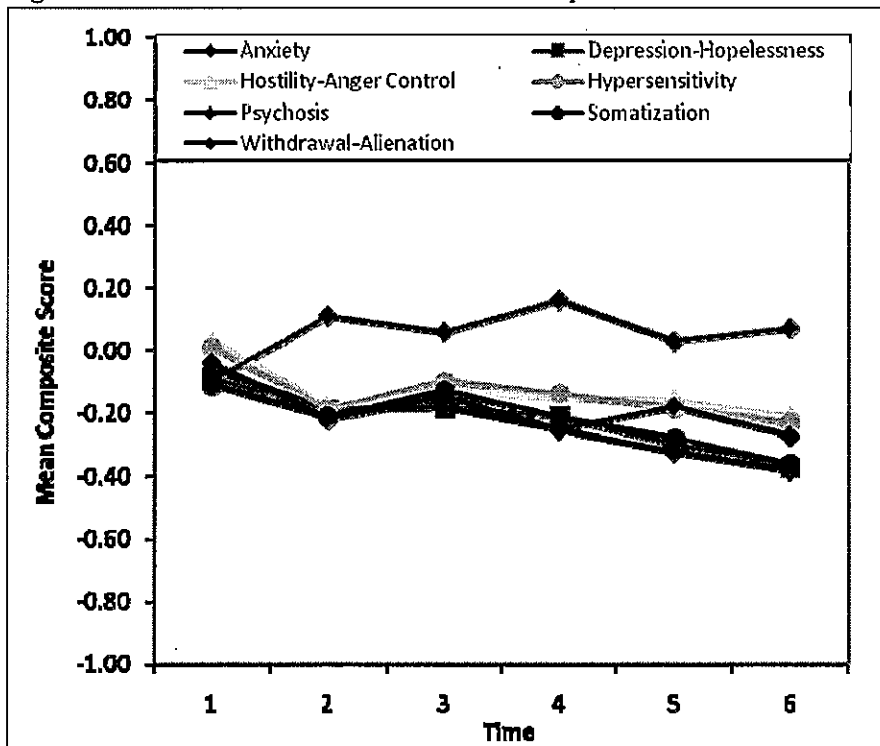
Table 13. F Statistics and Partial η^2 Comparing AS Groups across 6 Time Periods

Variable	Group Main Effect	Time Main Effect	Interaction Effect
Self-Report			
Anxiety	$F(1, 99) = 25.85, p < .001, \eta^2 = .21$	$F(3.98, 393.87) = 8.13, p < .001, \eta^2 = .08$	$F(3.98, 393.87) = 2.97, p = .02, \eta^2 = .03$
Depression-Hopelessness	$F(1, 99) = 18.86, p < .001, \eta^2 = .16$	$F(4.10, 405.75) = 6.21, p < .001, \eta^2 = .06$	$F(4.10, 405.75) = 1.12, p = .35, \eta^2 = .01$
Hostility-Anger Control	$F(1, 99) = 4.08, p = .05, \eta^2 = .04$	$F(4.08, 403.72) = 4.58, p = .001, \eta^2 = .04$	$F(4.08, 403.72) = 2.37, p = .05, \eta^2 = .02$
Hypersensitivity	$F(1, 99) = 14.03, p < .001, \eta^2 = .12$	$F(4.81, 476.08) = 2.91, p = .02, \eta^2 = .03$	$F(4.81, 476.08) = 2.50, p = .03, \eta^2 = .02$
Psychosis	$F(1, 99) = 13.51, p < .001, \eta^2 = .12$	$F(4.34, 430.18) = 2.79, p = .02, \eta^2 = .03$	$F(4.34, 430.18) = 1.49, p = .20, \eta^2 = .02$
Somatization	$F(1, 99) = 23.63, p < .001, \eta^2 = .19$	$F(4.34, 429.69) = 6.04, p < .001, \eta^2 = .06$	$F(4.34, 429.69) = 2.84, p = .02, \eta^2 = .03$
Withdrawal-Alienation	$F(1, 99) = 7.10, p = .01, \eta^2 = .07$	$F(4.74, 469.56) = 3.62, p = .004, \eta^2 = .04$	$F(4.74, 469.56) = 1.93, p = .09, \eta^2 = .02$
SLUMS	$F(1, 99) = 3.99, p = .05, \eta^2 = .04$	$F(4.56, 451.82) = 31.78, p < .001, \eta^2 = .24$	$F(4.56, 451.82) = .71, p = .60, \eta^2 = .01$
Trails B/A	$F(1, 91) = 2.74, p = .10, \eta^2 = .03$	$F(4.14, 376.44) = 4.91, p = .001, \eta^2 = .05$	$F(4.14, 376.44) = .81, p = .52, \eta^2 = .01$
Staff Report			
PBRs Anti-Authority	$F(1, 90) = .62, p = .43, \eta^2 = .01$	$F(4.02, 361.54) = 8.87, p = .001, \eta^2 = .09$	$F(4.02, 361.54) = 1.80, p = .13, \eta^2 = .02$
PBRs Anxious-Depressed	$F(1, 88) = 9.46, p = .003, \eta^2 = .10$	$F(4.30, 378.62) = 7.63, p < .001, \eta^2 = .08$	$F(4.30, 378.62) = .96, p = .44, \eta^2 = .01$
PBRs Dull-Confused	$F(1, 87) = 27.08, p < .001, \eta^2 = .24$	$F(3.90, 339.21) = 4.28, p = .002, \eta^2 = .05$	$F(3.90, 339.21) = 1.32, p = .26, \eta^2 = .02$
PBRs Total	$F(1, 88) = 7.28, p = .01, \eta^2 = .08$	$F(3.97, 349.18) = 9.84, p < .001, \eta^2 = .10$	$F(3.97, 349.18) = .75, p = .56, \eta^2 = .01$
BPRS Activity	$F(1, 82) = 14.04, p < .001, \eta^2 = .15$	$F(1.55, 127.12) = 2.46, p = .10, \eta^2 = .03$	$F(1.55, 127.12) = .01, p = .99, \eta^2 < .001$
BPRS Anxious-Depressed	$F(1, 82) = 19.34, p < .001, \eta^2 = .19$	$F(2, 163.57) = .91, p = .40, \eta^2 = .01$	$F(2, 163.57) = 2.16, p = .12, \eta^2 = .03$
BPRS Hostility-Suspiciousness	$F(1, 82) = 18.12, p < .001, \eta^2 = .18$	$F(1.56, 128.19) = 7.93, p = .002, \eta^2 = .09$	$F(1.56, 128.19) = 1.71, p = .19, \eta^2 = .02$
BPRS Thought Disorders	$F(1, 82) = 21.05, p < .001, \eta^2 = .20$	$F(1.77, 145.16) = .33, p = .69, \eta^2 = .004$	$F(1.77, 145.16) = .59, p = .54, \eta^2 = .01$
BPRS Withdrawal	$F(1, 82) = 10.15, p = .002, \eta^2 = .11$	$F(1.99, 163.04) = .81, p = .45, \eta^2 = .01$	$F(1.99, 163.04) = .17, p = .84, \eta^2 = .002$
BPRS Total	$F(1, 82) = 36.90, p < .001, \eta^2 = .31$	$F(1.84, 151.31) = 2.82, p = .06, \eta^2 = .03$	$F(1.84, 151.31) = 1.34, p = .26, \eta^2 = .02$

There were significant main effects of time on all variables; however, the results do not support the predicted hypothesis of significant decline in psychological well-being over time. Figure 11 provides the mean change over time for each composite. The only variable showing decreased functioning over time was the withdrawal-alienation composite. Time 4 (6 months incarcerated in CSP) revealed the highest levels of withdrawal-alienation, followed by a significant decline from time 4 to time 5. For the other psychological variables, there was improved functioning over time; however, when comparing sequential time periods, the majority of the variables (i.e., anxiety, depression-hopelessness, hostility-anger control, hypersensitivity, psychosis) only showed statistically significant improvement from the first to the second assessment period. The exception to this basic pattern was for the somatization composite where statistically significant improvement occurred between periods 5 and 6. For the cognitive variables there were also significant time effects. The Trails derived score showed significant improvement from the first assessment to the second. The SLUMS showed significant improvements in cognitive performance between times 2 and 3 and between times 5 and 6.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

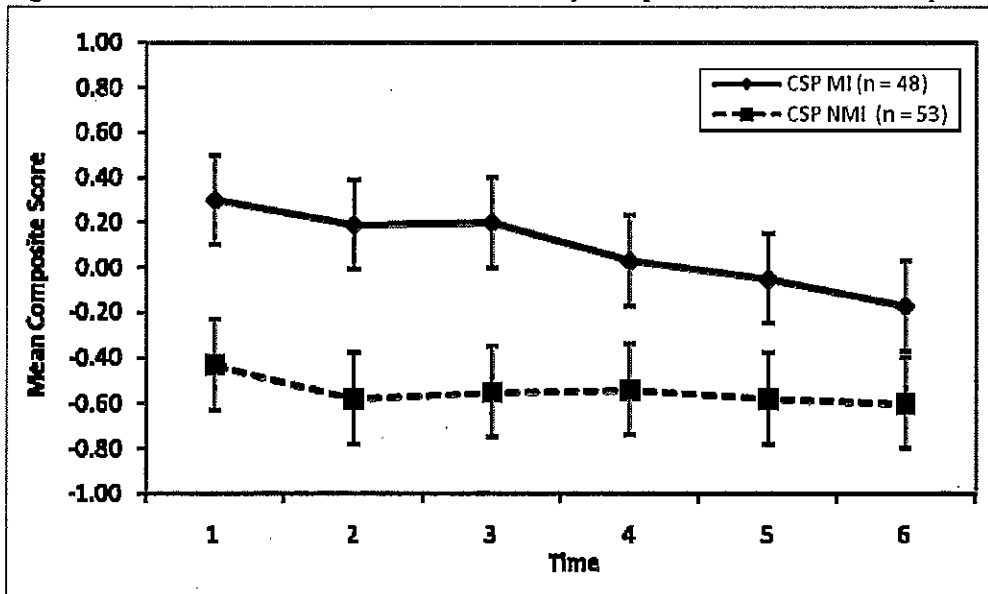
Figure 11. Mean Scores over Time for the 7 Composites Summarized across the CSP Groups



Although there were significant changes across time for all variables, we were particularly interested in the group by time interaction to determine if there was differential change across times based on mental health status. There were statistically significant interactions for the anxiety, hostility-anger, hypersensitivity, and somatization composites. Figures 12 to 15 provide graphical representations of these interactions. To further understand these interactions, simple main effects of time were examined for each group using Bonferroni pairwise comparisons of time periods.

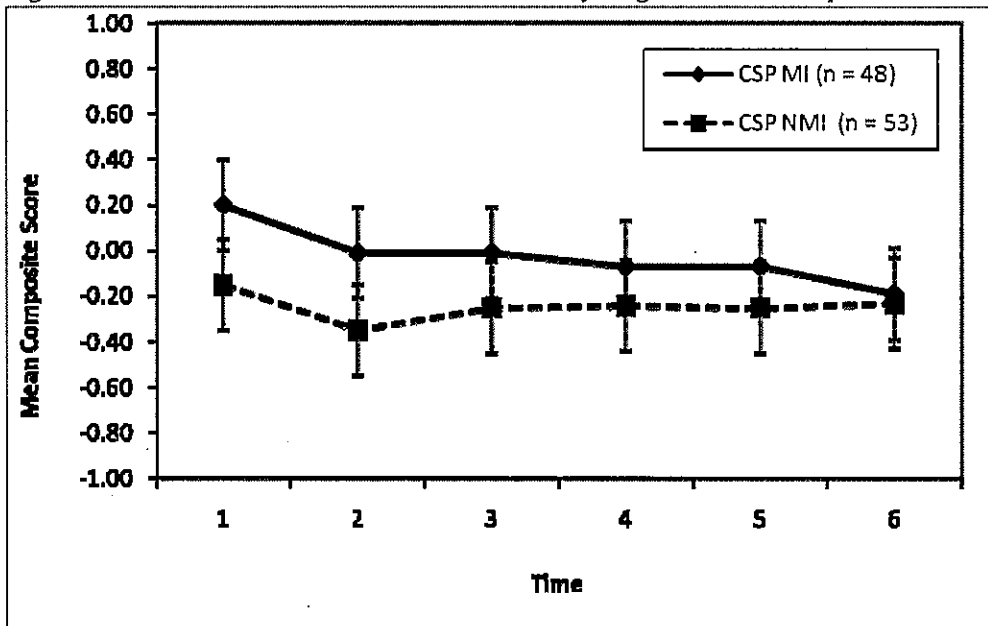
For the anxiety composite, the CSP NMI group showed no significant change over time, but the CSP MI group did. Mean scores at the fourth, fifth and sixth assessments were significantly lower than means at the first three assessment periods, and the sixth assessment mean was significantly lower than the mean at the fourth assessment.

Figure 12. Mean Scores over Time for the Anxiety Composite for each CSP Group



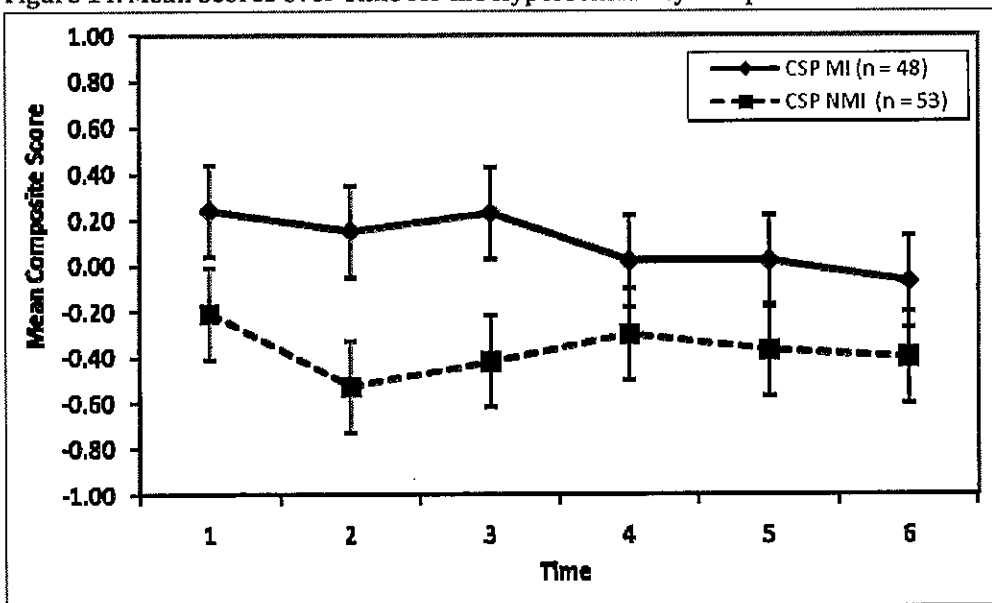
For the hostility-anger control composite, the CSP NMI group showed no significant change over time. In contrast, the CSP MI group showed significant improvement over time with mean hostility-anger control scores significantly elevated at the first assessment compared to all other time periods and the last assessment period significantly lower than the first three assessment periods.

Figure 13. Mean Scores over Time for the Hostility-Anger Control Composite for each CSP Group



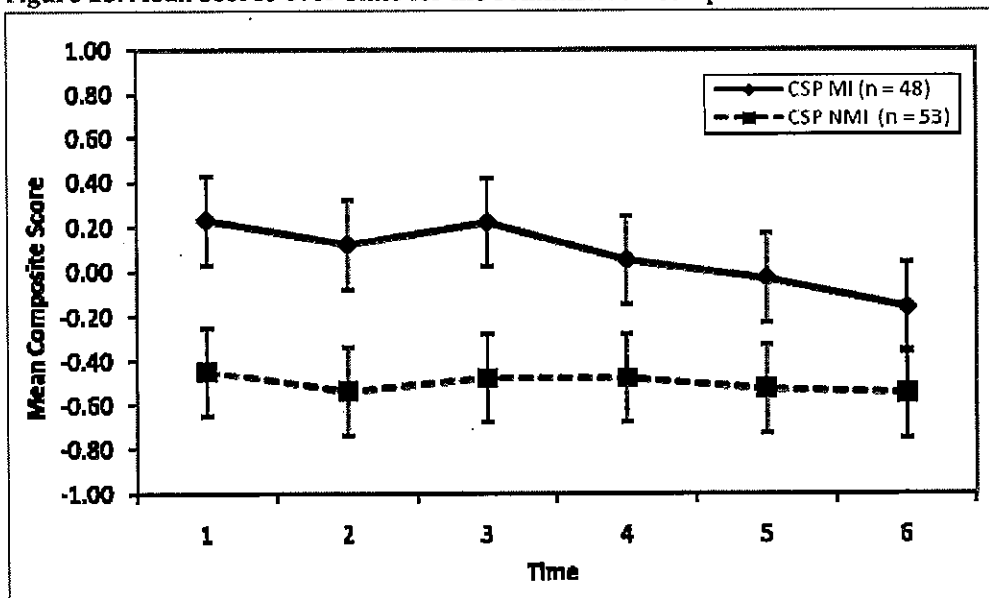
For the hypersensitivity composite, both groups demonstrated significant change over time. The CSP NMI group showed significant improvement from time 1 to time 2 but then scores worsened over time so that scores at the fourth assessment were significantly worse than the scores at the second period. For the CSP MI group, there was a significant decline in scores from the first to the second assessment periods, then an increase in scores with an elevation occurring at time 3 (compared to time 2), and then a significant decline in scores at the fourth assessment period.

Figure 14. Mean Scores over Time for the Hypersensitivity Composite for each CSP Group



For the somatization composite, there was significant change over time for the CSP MI group but not for the CSP NMI group. Significant decreases in scores occurred from the third to the fourth assessment periods and from the fifth to the sixth periods.

Figure 15. Mean Scores over Time for the Somatization Composite for each CSP Group



Comparisons on Staff Report Measures. In addition to self-report measures, data were collected from correctional staff using the PBRs and from clinicians using the BPRS. The PBRs was given at each of the 6 assessment periods whereas the BPRS was given at 6 month intervals so that there were only three assessments. Table 12 provides the means and standard deviations for the PBRs scores, Table 14 provides the summary statistics for the three BPRS assessments, and the inferential results for both variables are in Table 13.

Table 14. Summary Statistics on BPRS Scales across 3 Time Periods for All Study Groups

BPRS Subscale	<i>M</i> ₁	<i>SD</i> ₁	<i>M</i> ₃	<i>SD</i> ₃	<i>M</i> ₅	<i>SD</i> ₅
Activity						
CSP MI (<i>n</i> = 49)	6.39	1.72	6.08	1.74	6.00	1.53
CSP NMI (<i>n</i> = 35)	5.60	1.14	5.26	.74	5.20	.53
GP MI (<i>n</i> = 25)	6.40	1.63	5.88	1.20	6.04	1.59
GP NMI (<i>n</i> = 25)	5.64	1.25	5.36	.86	5.28	.54
SCCF (<i>n</i> = 55)	6.85	2.67	6.45	1.48	6.24	1.98
Anxious-Depressed						
CSP MI (<i>n</i> = 49)	9.51	3.11	9.35	2.93	8.47	3.02
CSP NMI (<i>n</i> = 35)	7.37	2.07	6.74	2.24	7.42	3.14
GP MI (<i>n</i> = 25)	8.68	3.13	7.96	2.47	8.40	2.31
GP NMI (<i>n</i> = 25)	6.68	1.77	6.52	1.83	6.08	1.78
SCCF (<i>n</i> = 55)	10.54	3.28	8.87	2.65	8.85	2.98
Hostility-Suspiciousness						
CSP MI (<i>n</i> = 49)	5.51	2.42	5.35	2.80	4.41	1.94
CSP NMI (<i>n</i> = 35)	4.17	1.99	3.37	.69	3.31	.68
GP MI (<i>n</i> = 25)	4.84	1.84	4.52	2.29	4.36	1.93
GP NMI (<i>n</i> = 25)	3.96	1.97	3.60	1.53	3.72	1.67
SCCF (<i>n</i> = 55)	5.53	3.01	4.51	1.91	4.64	2.12
Thought Disorder						
CSP MI (<i>n</i> = 49)	6.53	2.34	6.71	2.18	6.35	2.24
CSP NMI (<i>n</i> = 35)	5.43	1.04	5.14	.43	5.23	.55
GP MI (<i>n</i> = 25)	5.64	.99	5.40	.91	5.24	.91
GP NMI (<i>n</i> = 25)	5.20	.50	5.04	.20	5.44	1.44
SCCF (<i>n</i> = 55)	8.40	3.55	6.49	1.91	5.24	.83
Withdrawal						
CSP MI (<i>n</i> = 49)	7.59	1.63	7.67	1.98	7.39	1.50
CSP NMI (<i>n</i> = 35)	6.68	.99	7.00	1.37	6.71	1.82
GP MI (<i>n</i> = 25)	7.00	1.55	6.80	1.32	7.16	1.34
GP NMI (<i>n</i> = 25)	6.44	.65	6.20	.50	6.32	.80
SCCF (<i>n</i> = 55)	8.56	2.48	7.69	1.75	7.53	1.49
Total						
CSP MI (<i>n</i> = 49)	35.53	7.19	35.16	8.90	32.61	6.81
CSP NMI (<i>n</i> = 35)	29.26	4.85	27.51	3.71	27.89	4.92
GP MI (<i>n</i> = 25)	32.56	5.86	30.56	4.98	31.20	4.17
GP NMI (<i>n</i> = 25)	27.92	4.81	26.72	3.23	26.84	3.75
SCCF (<i>n</i> = 55)	39.89	9.97	34.02	5.55	33.84	6.80

For the correctional staff ratings, there were statistically significant group differences on the Anxious-Depressed, Dull-Confused, and Total scales with the CSP MI group scoring significantly higher on each subscale compared to the NMI group. There were significant changes across time for both groups with the first assessment showing higher ratings compared to the second assessment period ratings on all PBRs scales. Additionally, there was a statistically significant drop in Anti-Authority scores from the third rating to the

fourth rating. There were no group by time interactions, indicating that change over time was the same for the two CSP groups.

As might be expected, the CSP MI group was elevated on each of the clinical rating scales of the BPRS compared to the CSP NMI group. There was only significant change across time on the Hostility-Suspiciousness subscale with scores at the last time period ($M = 3.86$, $SE = .17$) significantly lower than the first ($M = 4.84$, $SE = .25$) and middle assessment ($M = 4.36$, $SE = .24$) period means. There were no statistically significant group by time interaction effects.

Comparisons between NMI Groups

A significant advantage of this study is the use of comparison groups to determine if the CSP groups change over time differentially compared to similar groups of inmates who are not placed in AS. In the following analyses, participants without mental health issues are compared in order to determine if those in AS change over time on the measures differentially compared to those in the general prison population (CSP NMI vs. GP NMI). (A later section compares the participants who have been identified as mentally ill.) The groups are compared on the five common time assessments. Mixed design analysis of variance was used to compare change across time and between groups.

Comparisons on Self-Report Measures. The summary statistics for the groups are provided in Table 15 and the analysis of variance results and effect sizes are provided in Table 16. For anxiety, depression-hopelessness, hostility-anger control, hypersensitivity, psychosis, and somatization composites, there were statistically significant group differences between the groups with the CSP NMI scoring significantly higher than the GP NMI group. For the withdrawal-alienation composite, the SLUMS cognitive measure, and the Trails derived score, there were no statistically significant differences between groups.

Table 15. Summary Statistics on Measures across 5 Time Periods for the NMI Groups

Variable	M_1	SD_1	M_2	SD_2	M_3	SD_3	M_4	SD_4	M_5	SD_5
Anxiety										
CSP NMI ($n = 55$)	-.44	.53	-.59	.56	-.56	.60	-.55	.59	-.58	.58
GP NMI ($n = 38$)	-.71	.47	-.82	.40	-.85	.40	-.86	.35	-.88	.40
Depression-Hopelessness										
CSP NMI ($n = 55$)	-.39	.52	-.47	.57	-.47	.58	-.50	.55	.50	.57
GP NMI ($n = 38$)	-.73	.35	-.82	.34	-.84	.39	-.81	.29	-.85	.30
Hostility-Anger Control										
CSP NMI ($n = 55$)	-.13	.61	-.34	.52	-.23	.59	-.23	.64	-.24	.64
GP NMI ($n = 38$)	-.34	.49	-.45	.47	-.48	.54	-.54	.50	-.50	.51
Hypersensitivity										
CSP NMI ($n = 55$)	-.25	.65	-.54	.66	-.44	.74	-.31	.75	-.36	.72
GP NMI ($n = 38$)	-.50	.68	-.66	.62	-.73	.68	-.64	.51	-.73	.56
Psychosis										
CSP NMI ($n = 55$)	-.28	.71	-.51	.60	-.47	.63	-.41	.69	-.40	.75
GP NMI ($n = 38$)	-.60	.72	-.81	.53	-.81	.64	-.77	.53	-.86	.64
Somatization										
CSP NMI ($n = 55$)	-.46	.62	-.56	.63	-.50	.63	-.50	.56	-.52	.62
GP NMI ($n = 38$)	-.61	.51	-.77	.41	-.77	.41	-.72	.49	-.76	.49
Withdrawal-Alienation										
CSP NMI ($n = 55$)	-.30	.60	-.12	.85	-.15	.82	.00	.82	-.07	.84
GP NMI ($n = 38$)	-.45	.78	-.32	.68	-.42	.70	-.32	.83	-.32	.80
SLUMS										
CSP NMI ($n = 55$)	21.74	3.46	22.53	3.66	23.98	3.30	24.34	3.05	24.18	3.33
GP NMI ($n = 38$)	23.16	3.97	23.92	3.26	24.47	3.55	24.71	3.69	24.82	3.24
Trails B/A										
CSP NMI ($n = 55$)	3.11	1.52	2.84	.93	3.00	1.22	2.88	1.14	2.71	1.02
GP NMI ($n = 38$)	2.82	.82	3.07	1.64	2.90	1.04	2.65	.76	2.80	.95
PBRs Anti-authority										
CSP NMI ($n = 51$)	7.75	7.75	5.46	6.30	3.92	4.87	4.38	4.79	3.40	4.48
GP NMI ($n = 22$)	5.58	7.23	7.02	5.41	6.69	7.24	7.75	5.91	7.93	6.92
PBRs Anxious-Depressed										
CSP NMI ($n = 51$)	3.21	4.32	1.92	3.02	1.43	2.48	1.55	4.40	1.84	3.27
GP NMI ($n = 20$)	2.70	3.66	2.80	3.49	3.69	3.90	3.86	5.28	3.10	4.05
PBRs Dull-Confused										
CSP NMI ($n = 49$)	1.84	2.40	1.26	2.13	.86	1.19	.98	1.48	1.12	1.60
GP NMI ($n = 20$)	1.71	2.13	1.62	1.77	.09	2.08	2.10	2.65	2.07	2.83
PBRs Total										
CSP NMI ($n = 51$)	12.81	12.44	8.80	10.09	6.31	7.52	6.99	7.51	6.64	7.47
GP NMI ($n = 20$)	10.22	12.32	11.39	7.23	12.82	11.48	13.52	10.98	13.55	11.77

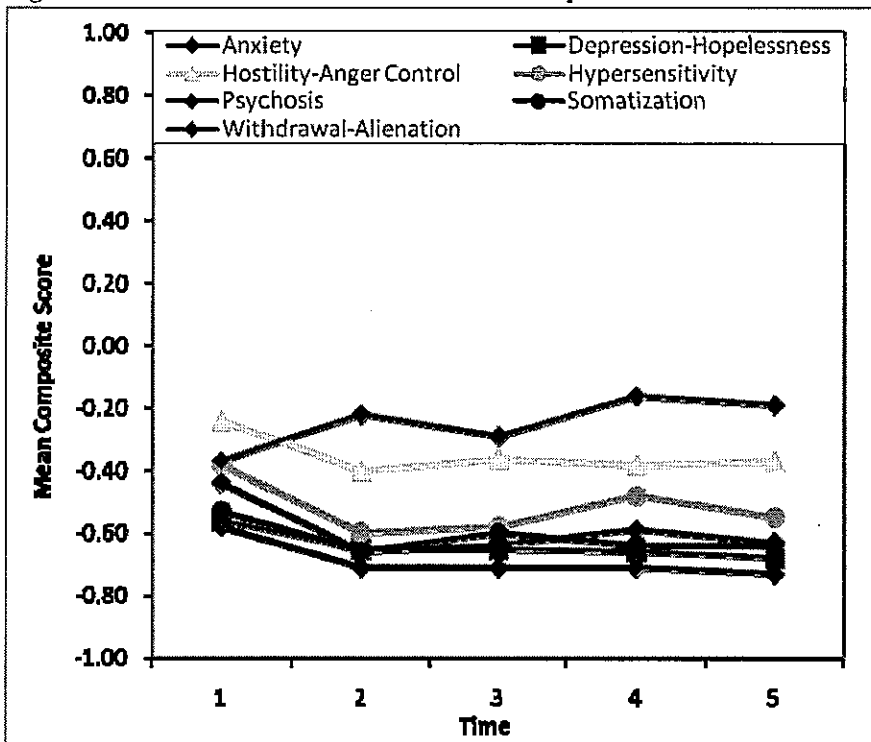
Table 16. F Statistics and Partial η^2 Comparing NMI Groups across 5 Time Periods

Variable	Group Main Effect	Time Main Effect	Interaction Effect
Self-Report			
Anxiety	$F(1, 91) = 8.74, p = .004, \eta^2 = .09$	$F(3.03, 275.73) = 4.52, p = .004, \eta^2 = .05$	$F(3.03, 275.73) = .28, p = .84, \eta^2 = .003$
Depression-Hopelessness	$F(1, 91) = 15.24, p < .001, \eta^2 = .14$	$F(3.86, 351.74) = 4.74, p = .001, \eta^2 = .05$	$F(3.86, 351.74) = 1.45, p = .22, \eta^2 = .02$
Hostility-Anger Control	$F(1, 91) = 4.69, p = .03, \eta^2 = .05$	$F(3.50, 318.75) = 2.77, p = .03, \eta^2 = .03$	$F(3.50, 318.75) = .13, p = .96, \eta^2 = .001$
Hypersensitivity	$F(1, 91) = 5.18, p = .02, \eta^2 = .05$	$F(3.79, 345.07) = 4.70, p = .001, \eta^2 = .05$	$F(3.79, 345.07) = 1.21, p = .31, \eta^2 = .01$
Psychosis	$F(1, 91) = 8.67, p = .004, \eta^2 = .09$	$F(3.09, 281.17) = 5.11, p = .001, \eta^2 = .05$	$F(3.09, 281.17) = .65, p = .59, \eta^2 = .01$
Somatization	$F(1, 91) = 4.75, p = .03, \eta^2 = .05$	$F(3.43, 312.15) = 2.42, p = .06, \eta^2 = .03$	$F(3.43, 312.15) = .69, p = .58, \eta^2 = .01$
Withdrawal-Alienation	$F(1, 91) = 2.79, p = .10, \eta^2 = .03$	$F(3.88, 352.78) = 3.49, p = .01, \eta^2 = .04$	$F(3.88, 352.78) = .52, p = .72, \eta^2 = .01$
SLUMS	$F(1, 91) = 2.09, p = .15, \eta^2 = .02$	$F(3.77, 343.26) = 15.33, p < .001, \eta^2 = .14$	$F(3.77, 343.26) = 1.17, p = .32, \eta^2 = .01$
Trails B/A	$F(1, 86) = .13, p = .72, \eta^2 = .001$	$F(3.53, 303.21) = 1.25, p = .29, \eta^2 = .01$	$F(3.53, 303.21) = 1.29, p = .28, \eta^2 = .02$
Staff Report			
PBRS Anti-Authority	$F(1, 71) = 4.67, p = .03, \eta^2 = .06$	$F(3.43, 243.59) = .60, p = .64, \eta^2 = .01$	$F(3.43, 243.59) = 3.64, p = .01, \eta^2 = .05$
PBRS Anxious-Depressed	$F(1, 69) = 4.63, p = .04, \eta^2 = .06$	$F(3.62, 249.81) = .30, p = .86, \eta^2 = .004$	$F(3.62, 249.81) = 1.92, p = .12, \eta^2 = .03$
PBRS Dull-Confused	$F(1, 67) = 7.61, p = .01, \eta^2 = .10$	$F(3.59, 240.56) = .23, p = .91, \eta^2 = .003$	$F(3.59, 240.56) = 1.45, p = .22, \eta^2 = .02$
PBRS Total	$F(1, 69) = 6.54, p = .01, \eta^2 = .09$	$F(3.37, 232.76) = .40, p = .77, \eta^2 = .01$	$F(3.37, 232.76) = 3.22, p = .02, \eta^2 = .04$
BPRS Activity	$F(1, 58) = .19, p = .67, \eta^2 = .003$	$F(1.43, 83.11) = 4.60, p = .01, \eta^2 = .07$	$F(1.43, 83.11) = .03, p = .93, \eta^2 < .001$
BPRS Anxious-Depressed	$F(1, 58) = 2.66, p = .11, \eta^2 = .04$	$F(1.76, 101.84) = .81, p = .43, \eta^2 = .01$	$F(1.76, 101.84) = 1.59, p = .21, \eta^2 = .03$
BPRS Hostility-Suspiciousness	$F(1, 58) = .22, p = .64, \eta^2 = .004$	$F(1.47, 85.54) = 4.83, p = .02, \eta^2 = .08$	$F(1.47, 85.54) = 1.15, p = .31, \eta^2 = .02$
BPRS Thought Disorder	$F(1, 58) = .07, p = .78, \eta^2 = .001$	$F(1.89, 109.66) = 2.54, p = .09, \eta^2 = .04$	$F(1.89, 109.66) = 1.79, p = .17, \eta^2 = .03$
BPRS Withdrawal	$F(1, 58) = 5.95, p = .02, \eta^2 = .09$	$F(1.73, 100.26) = .08, p = .90, \eta^2 = .001$	$F(1.73, 100.26) = .97, p = .37, \eta^2 = .02$
BPRS Total	$F(1, 58) = 1.35, p = .25, \eta^2 = .02$	$F(1.83, 106.28) = 3.69, p = .03, \eta^2 = .06$	$F(1.83, 106.28) = .11, p = .88, \eta^2 = .002$

For all variables except the somatization composite and Trails derived score, there were statistically significant changes across time; however, there were not any significant group by time interactions, indicating that the two groups changed similarly across time. Figure 16 gives the mean change over time on the composites (summarized across the two NMI groups). For all composite variables except withdrawal-alienation, the pattern of change was the same when examining differences between sequential time periods. There were statistically significant improvements in reported psychological well-being from the first to the second assessment but no other significant differences between time periods. The withdrawal-alienation composite was the only variable that showed significantly higher scores over time, with statistically significant change on average from the first to second assessments and from the third to the fourth assessments. The SLUMS also showed change across time with significant improvement from the first ($M = 22.45, SE = .39$) to second ($M = 23.22, SE = .27$) assessment and from the second to third ($M = 24.23, SE = .36$) assessment.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Figure 16. Mean Scores over Time for the 7 Composites Summarized across the NMI Groups



Comparisons on Staff Report Measures. The summary statistics for the correctional officer ratings on the PBRS are given in Table 15, the summary statistics for the clinician ratings are given in Table 14, and the analysis of variance results for the staff report comparisons are given in Table 16.

For the correctional officer ratings, there are significant group differences on all four PBRS scales. When averaged over time, the CSP NMI group scored significantly lower than the GP NMI group on each of the four scales. There were no significant main effects of time on any of the scales and no significant interaction effects for the Anxious-Depressed or Dull-Confused scales; however, there were significant interaction effects for Anti-Authority and Total scales. These interactions are displayed in Figures 17 and 18. For both scales, the same basic pattern occurs with the CSP NMI scores tending to decrease across time with significant drops from the first to the second assessment and with the GP NMI scores showing no significant change across time (although scores tend to increase).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Figure 17. Mean Scores over Time for the PBRs Anti-Authority Subscale for each NMI Group

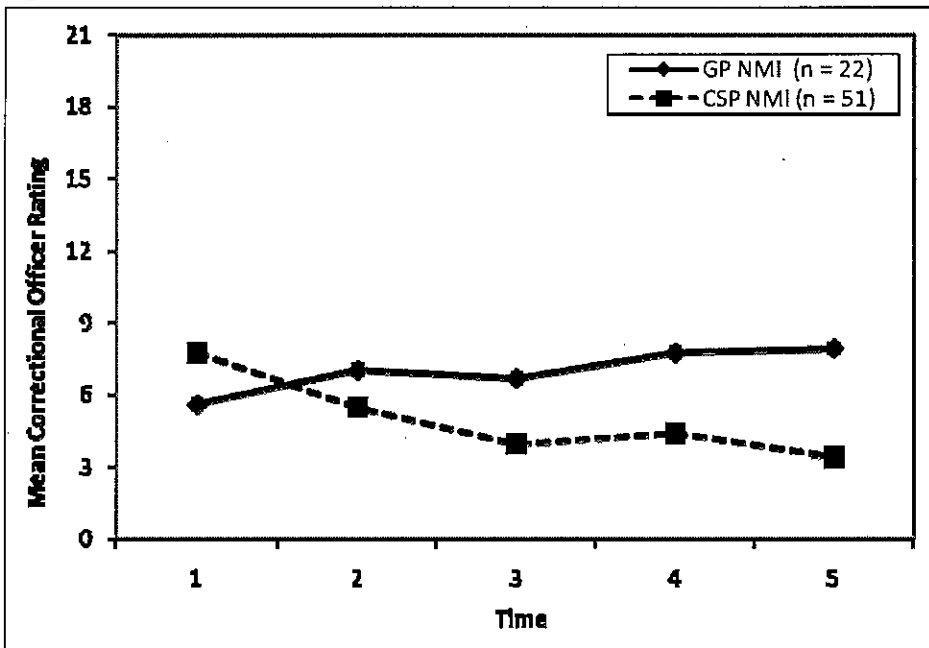
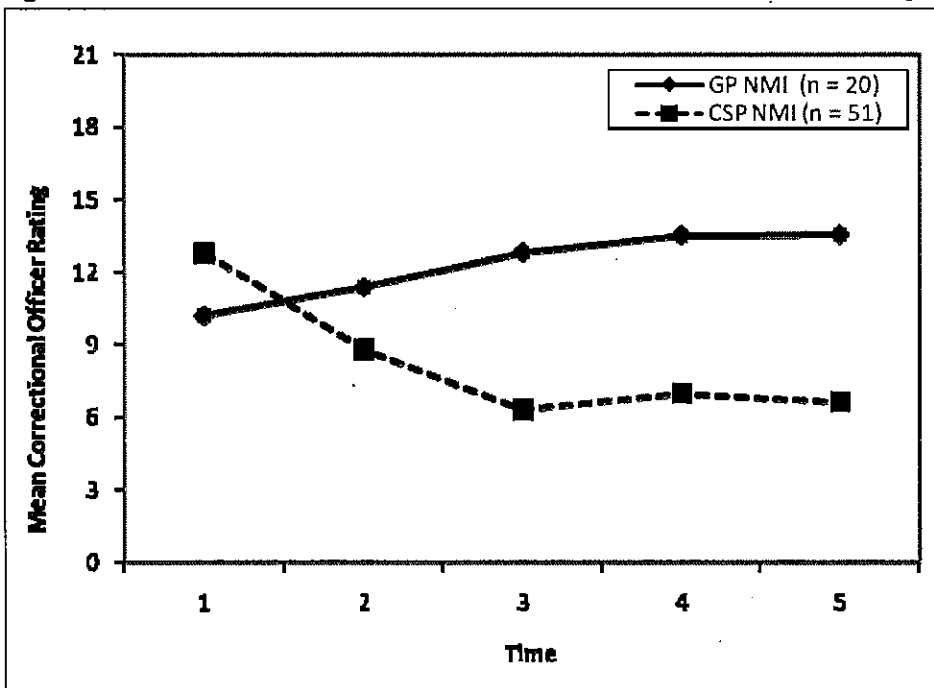


Figure 18. Mean Scores over Time for the PBRs Total Scale for each NMI Group



For the clinician ratings, there was a significant group difference on the Withdrawal subscale of the BPRS with the CSP NMI group ($M = 6.80, SE = .13$) rated significantly higher compared to the GP NMI group ($M = 6.32, SE = .15$). No other BPRS subscales had statistically significant group differences. There were significant time effects on Activity, Hostility-Suspiciousness, and Total scores but no significant interaction effects for any of the BPRS subscales. For the Activity subscale, ratings at the first assessment ($M = 5.62, SE = .16$) were significantly higher than ratings at the third assessment ($M = 5.24, SE = .10$) but not different from ratings at

the second assessment ($M = 5.24, SE = .07$). For the Hostility-Suspiciousness subscale, the ratings at the first assessment ($M = 7.03, SE = .26$) were significantly greater than ratings at the second ($M = 6.63, SE = .27$) and the third ($M = 6.75, SE = .35$) assessment periods. The total score showed this same pattern with first assessment ($M = 28.59, SE = .63$) ratings significantly higher than second ($M = 27.12, SE = .47$) and third ($M = 27.36, SE = .59$) periods.

Comparisons between MI Groups

In the following analyses, the three groups with participants identified as mentally ill are compared. Like the comparisons between the NMI groups, there is a CSP MI group and a GP MI group plus a third group of inmates who have been placed in a psychiatric treatment facility (SCCF). Analyses were completed on the five common time periods using mixed design analysis of variance techniques.

Comparisons on Self-Report Measures. The summary statistics for the groups are provided in Table 17 and the analysis of variance results and effect sizes are provided in Table 18. Significant group differences were found on the anxiety, depression-hopelessness, psychosis, somatization, and withdrawal-alienation composites. Using Bonferroni corrected pairwise comparisons, the SCCF group was always significantly higher than the GP MI group on these composites. Additionally, the SCCF group was significantly higher than the CSP MI group for the depression-hopelessness, psychosis, and withdrawal-alienation composites but not significantly different for anxiety and somatization composites. The GP MI and CSP MI groups did not show any statistically significant mean differences although the CSP MI group always had a higher mean. There were not significant group differences on the hostility-anger control and hypersensitivity composites or on the cognitive measures (SLUMS and Trails derived).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 17. Summary Statistics on Measures across 5 Time Periods for the MI Groups

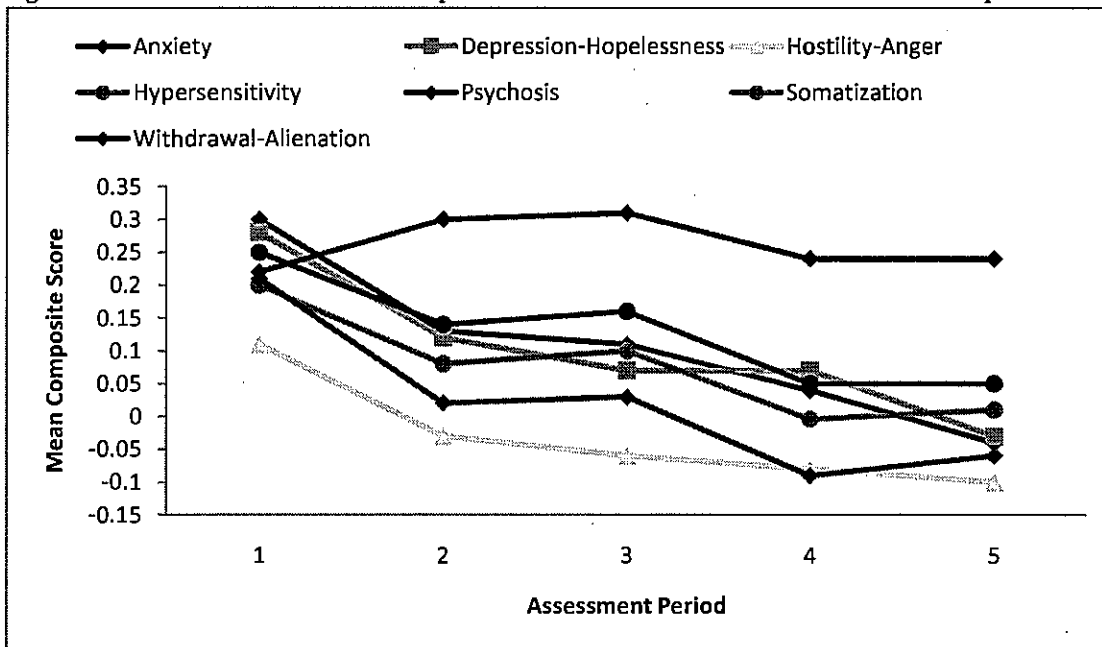
Variable	<i>M</i> ₁	<i>SD</i> ₁	<i>M</i> ₂	<i>SD</i> ₂	<i>M</i> ₃	<i>SD</i> ₃	<i>M</i> ₄	<i>SD</i> ₄	<i>M</i> ₅	<i>SD</i> ₅
Anxiety										
CSP MI (<i>n</i> = 55)	.26	.84	.14	.82	.12	.86	-.01	.86	-.10	.86
GP MI (<i>n</i> = 28)	.11	.77	-.09	.65	-.06	.59	-.20	.56	-.22	.61
SCCF (<i>n</i> = 55)	.51	.71	.35	.81	.27	.88	.32	.86	.21	.76
Depression-Hopelessness										
CSP MI (<i>n</i> = 55)	.19	.78	.07	.74	.04	.80	.03	.84	-.11	.81
GP MI (<i>n</i> = 28)	.00	.76	-.19	.62	-.18	.53	-.22	.59	-.30	.54
SCCF (<i>n</i> = 55)	.64	.89	.47	.93	.35	.98	.39	.95	.33	.86
Hostility-Anger Control										
CSP MI (<i>n</i> = 55)	.20	.67	.00	.67	-.03	.69	-.09	.74	-.12	.72
GP MI (<i>n</i> = 28)	.11	.68	.01	.65	-.04	.58	-.08	.58	-.08	.55
SCCF (<i>n</i> = 55)	.02	.69	-.08	.73	-.09	.65	-.05	.79	-.10	.71
Hypersensitivity										
CSP MI (<i>n</i> = 55)	.16	.85	.08	.87	.11	.94	-.08	.90	-.05	.81
GP MI (<i>n</i> = 28)	.13	.92	-.02	.71	.60	.65	-.14	.66	-.11	.69
SCCF (<i>n</i> = 55)	.32	.78	.17	.78	.14	.91	.20	.77	.20	.75
Psychosis										
CSP MI (<i>n</i> = 55)	.16	.80	.04	.78	.09	.95	-.14	.87	-.02	.94
GP MI (<i>n</i> = 28)	.02	.80	-.29	.73	-.31	.65	-.46	.52	-.39	.64
SCCF (<i>n</i> = 55)	.46	.84	.31	.94	.31	.88	.33	.95	.22	.86
Somatization										
CSP MI (<i>n</i> = 55)	.23	.82	.14	.83	.21	.79	.09	.81	-.01	.85
GP MI (<i>n</i> = 28)	.07	.71	-.10	.65	-.03	.83	-.24	.53	-.14	.59
SCCF (<i>n</i> = 55)	.46	.67	.40	.68	.29	.81	.30	.83	.31	.76
Withdrawal-Alienation										
CSP MI (<i>n</i> = 55)	.15	.83	.34	.88	.25	.83	.30	.81	.18	.80
GP MI (<i>n</i> = 28)	.08	.88	.03	.78	.12	.68	-.06	.71	-.02	.62
SCCF (<i>n</i> = 55)	.43	.83	.53	.75	.56	.75	.50	.75	.55	.73
SLUMS										
CSP MI (<i>n</i> = 55)	20.80	5.44	21.20	4.86	22.27	4.68	22.84	4.38	23.51	4.03
GP MI (<i>n</i> = 28)	21.36	4.18	23.11	3.66	23.71	2.99	24.32	3.73	24.96	3.29
SCCF (<i>n</i> = 55)	20.96	3.55	21.81	4.24	23.11	4.37	23.52	4.07	23.35	4.04
PBRS Trails B/A										
CSP MI (<i>n</i> = 55)	2.99	1.04	2.74	1.15	2.67	.89	2.60	1.05	2.67	.76
GP MI (<i>n</i> = 28)	3.23	1.30	2.81	.68	2.75	.85	2.76	.87	2.62	.92
SCCF (<i>n</i> = 55)	2.97	1.16	2.95	1.13	2.90	1.16	2.80	.98	2.74	1.29
PBRS Anti-Authority										
CSP MI (<i>n</i> = 50)	7.04	6.96	6.66	6.34	6.75	7.60	3.56	5.40	4.64	6.35
GP MI (<i>n</i> = 16)	5.31	4.61	6.48	5.05	8.58	10.14	7.00	6.75	7.56	9.77
SCCF (<i>n</i> = 41)	2.85	5.60	4.00	4.73	3.51	4.18	5.67	6.64	5.70	6.94
PBRS Anxious-Depressed										
CSP MI (<i>n</i> = 49)	5.96	6.58	3.86	5.23	3.45	4.31	2.94	3.44	2.88	3.90
GP MI (<i>n</i> = 16)	2.31	3.53	4.32	6.30	5.00	4.63	2.89	3.83	3.62	5.20
SCCF (<i>n</i> = 41)	5.15	4.77	5.43	5.05	5.06	4.46	6.84	6.35	6.82	6.01
PBRS Dull-Confused										
CSP MI (<i>n</i> = 49)	3.94	4.32	2.71	3.57	2.63	3.04	2.82	3.64	2.26	3.12
GP MI (<i>n</i> = 16)	1.69	2.15	2.50	3.14	2.62	3.28	1.81	2.95	1.25	1.95
SCCF (<i>n</i> = 41)	3.50	3.92	3.54	2.85	3.16	2.94	4.20	3.99	4.30	4.24
PBRS Total										
CSP MI (<i>n</i> = 49)	17.30	15.89	13.63	13.63	13.22	13.25	9.20	9.89	9.93	10.98
GP MI (<i>n</i> = 16)	9.38	9.62	12.35	12.68	16.28	15.04	11.72	11.77	12.44	15.05
Both Groups	11.70	13.46	13.26	10.59	11.97	9.31	16.99	15.90	17.22	15.58

Table 18. F Statistics and Partial η^2 Comparing MI Groups across 5 Time Periods

Variable	Group Main Effect	Time Main Effect	Interaction Effect
Self-Report			
Anxiety	$F(2, 135) = 3.97, p = .02, \eta^2 = .06$	$F(3.67, 495.07) = 9.11, p < .001, \eta^2 = .06$	$F(7.34, 495.07) = .42, p = .90, \eta^2 = .01$
Depression-Hopelessness	$F(2, 135) = 8.44, p < .001, \eta^2 = .11$	$F(3.52, 475.64) = 6.76, p < .001, \eta^2 = .05$	$F(7.04, 475.64) = .24, p = .98, \eta^2 = .004$
Hostility-Anger Control	$F(2, 135) = .13, p = .88, \eta^2 = .002$	$F(3.45, 465.81) = 5.49, p = .001, \eta^2 = .04$	$F(6.90, 465.81) = .70, p = .67, \eta^2 = .01$
Hypersensitivity	$F(2, 135) = 1.41, p = .25, \eta^2 = .02$	$F(3.92, 529.43) = 2.80, p = .02, \eta^2 = .02$	$F(7.84, 529.43) = .62, p = .76, \eta^2 = .01$
Psychosis	$F(2, 135) = 7.28, p = .001, \eta^2 = .10$	$F(3.82, 148.17) = 6.55, p < .001, \eta^2 = .05$	$F(7.66, 148.17) = .97, p = .46, \eta^2 = .01$
Somatization	$F(2, 135) = 4.26, p = .02, \eta^2 = .06$	$F(3.87, 522.39) = 4.83, p = .001, \eta^2 = .04$	$F(7.74, 522.39) = .89, p = .52, \eta^2 = .01$
Withdrawal-Alienation	$F(2, 135) = 5.51, p = .01, \eta^2 = .08$	$F(4.00, 539.26) = .79, p = .53, \eta^2 = .01$	$F(7.99, 539.26) = .77, p = .63, \eta^2 = .01$
SLUMS	$F(2, 134) = 1.32, p = .27, \eta^2 = .02$	$F(3.69, 494.42) = 28.64, p < .001, \eta^2 = .18$	$F(7.38, 494.42) = .76, p = .63, \eta^2 = .01$
Trails B/A	$F(2, 128) = .47, p = .63, \eta^2 = .01$	$F(4, 512) = 4.50, p = .001, \eta^2 = .03$	$F(8, 512) = .62, p = .76, \eta^2 = .01$
Staff Report			
PBRS Anti-Authority	$F(2, 104) = 2.56, p = .08, \eta^2 = .05$	$F(3.58, 372.72) = .63, p = .62, \eta^2 = .01$	$F(7.17, 372.72) = 3.30, p = .002, \eta^2 = .06$
PBRS Anxious-Depression	$F(2, 103) = 5.92, p = .004, \eta^2 = .10$	$F(3.62, 373.16) = .06, p = .99, \eta^2 = .001$	$F(7.25, 373.16) = 2.97, p = .004, \eta^2 = .05$
PBRS Dull-Confused	$F(2, 103) = 5.03, p = .01, \eta^2 = .09$	$F(3.46, 355.98) = .23, p = .90, \eta^2 = .002$	$F(6.91, 355.98) = 1.38, p = .21, \eta^2 = .03$
PBRS Total	$F(2, 103) = .40, p = .67, \eta^2 = .01$	$F(3.43, 352.95) = .13, p = .96, \eta^2 = .001$	$F(6.85, 352.95) = 3.46, p = .001, \eta^2 = .06$
BPRS Activity	$F(2, 126) = 1.36, p = .26, \eta^2 = .02$	$F(1.82, 228.85) = 2.78, p = .07, \eta^2 = .02$	$F(3.63, 228.85) = .16, p = .95, \eta^2 = .003$
BPRS Anxious-Depressed	$F(2, 126) = 2.32, p = .10, \eta^2 = .04$	$F(2, 252) = 5.15, p = .01, \eta^2 = .04$	$F(4, 252) = 1.66, p = .16, \eta^2 = .03$
BPRS Hostility-Suspiciousness	$F(2, 126) = .73, p = .48, \eta^2 = .01$	$F(1.92, 242.10) = 5.40, p = .01, \eta^2 = .04$	$F(3.84, 242.10) = 1.31, p = .27, \eta^2 = .02$
BPRS Thought Disorder	$F(2, 126) = 9.91, p < .001, \eta^2 = .14$	$F(1.61, 203.27) = 5.50, p = .01, \eta^2 = .04$	$F(3.23, 203.27) = 4.54, p = .003, \eta^2 = .07$
BPRS Withdrawal	$F(2, 126) = 5.46, p = .005, \eta^2 = .08$	$F(1.99, 250.81) = 1.76, p = .17, \eta^2 = .01$	$F(3.98, 250.81) = 2.02, p = .09, \eta^2 = .03$
BPRS Total	$F(2, 126) = 7.10, p = .001, \eta^2 = .10$	$F(1.85, 232.85) = 8.94, p < .001, \eta^2 = .07$	$F(3.70, 232.85) = 2.95, p = .02, \eta^2 = .04$

There were significant changes across times for all composites except the withdrawal-alienation composite. The hostility-anger control composite also showed a significant interaction indicating differential change across time between groups. Figure 19 provides the mean plot demonstrating change across time for the five composites that had a significant time effect but no interaction effect. For the anxiety and depression-hopelessness composites, there were significant decreases in mean scores from the first to second assessment periods. For the psychosis and somatization composites, there were significant decreases in mean ratings from the first to second and from the third to fourth assessment periods. The hypersensitivity composite had a significant time effect, but the comparison of sequential time periods showed no significant change (pairwise comparisons indicated that the first assessment mean was significantly higher than mean scores at the fourth and fifth periods).

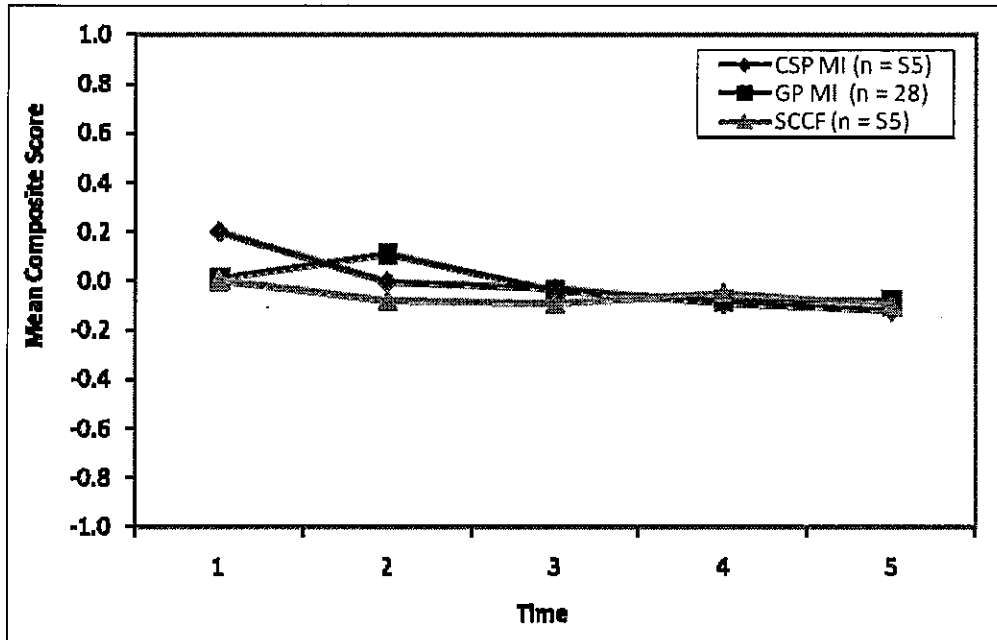
Figure 19. Mean Scores for the 7 Composites over Time Summarized across MI Groups



For the SLUMS cognitive assessment, there were significant increases in performance from the first ($M = 21.04, SE = .40$) to second ($M = 22.04, SE = .39$) and from the second to third ($M = 23.03, SE = .38$) assessment periods. For the Trails derived score there were significant improvements in performance from the first ($M = 3.08, SE = .10$) to second ($M = 2.83, SE = .10$) assessment periods (indicated by a decrease in mean scores).

The significant interaction for the hostility-anger control composite is graphed in Figure 20. There was significant change over time for the CSP MI group with time 1 scores significantly higher than all other assessment periods. There were not significant changes over time for the SCCF and GP MI groups.

Figure 20. Mean Change over Time on the Hostility-Anger Control Composite for each MI Group



Comparisons on Staff Report Measures. The summary statistics for the correctional staff ratings on the PBRs are given in Table 17, the summary statistics for the clinician ratings are given in Table 14, and the analysis of variance results for the staff report comparisons are given in Table 18.

For the correctional staff ratings, there were significant group differences for PBRs Anxious-Depressed and Dull-Confused subscale scores. The SCCF group ($M_{ad} = 5.86, SE = .48; M_{dc} = 3.74, SE = .31$) scored significantly higher than the both the CSP MI ($M_{ad} = 3.82, SE = .44; M_{dc} = 2.87, SE = .28$) and the GP MI ($M_{ad} = 3.63, SE = .76; M_{dc} = 1.98, SE = .50$) groups on both subscales. The Anti-Authority rating scale did not show statistically significant group differences ($p = .08$) but there was a small to moderate effect size ($\eta^2 = .05$); the only significant difference was between the GP MI and the SCCF groups ($p = .04$). There were no statistically significant time effects for any of the subscales of the correctional officer ratings; however, there were significant group by time interactions for the Anti-Authority, Anxious-Depressed, and Total scores. Figures 21 to 23 demonstrate the interaction for these three variables.

Further analyses of the interaction effects showed the same pattern of significance – there were significant changes over time for the CSP MI group but not for the other two groups. Specifically, for the Anti-Authority subscale, the fourth assessment had lower scores than the first three assessment periods; for the Anxious-Depressed subscale, the first assessment was higher than all other assessment periods; and for the Total PBRs scale, the first assessment was significantly higher than the fourth and fifth assessments. There were not significant changes over time for the other two groups; however it is noteworthy that these scores tended to increase over time.

Figure 21. Mean Scores over Time for the PBRs Anti-Authority Subscale for each MI Group

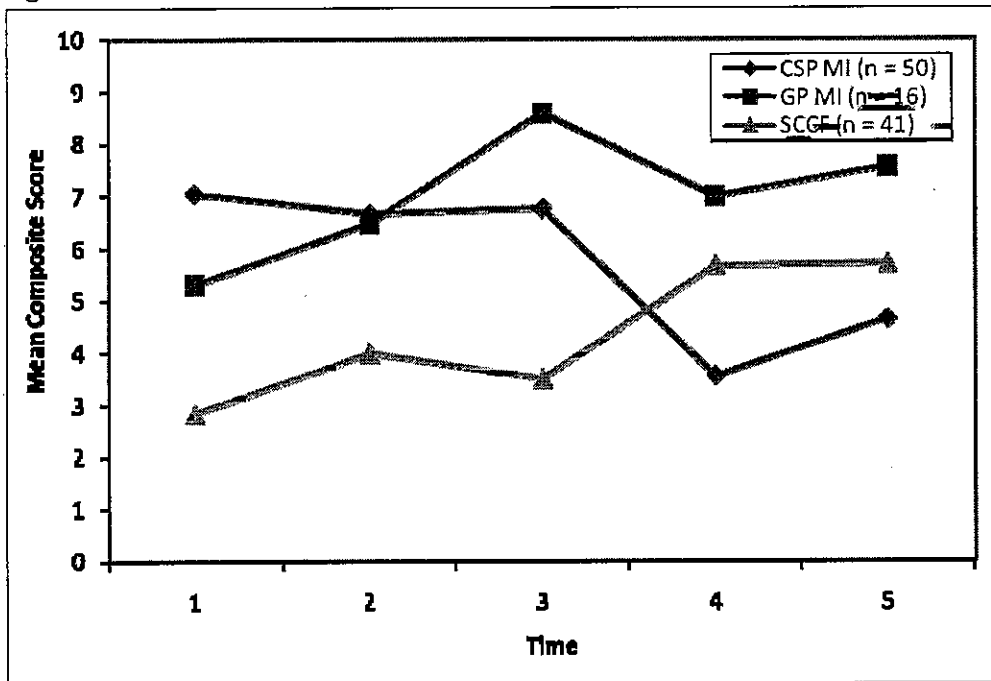


Figure 22. Mean Scores over Time for the PBRs Anxious-Depressed Subscale for each MI Group

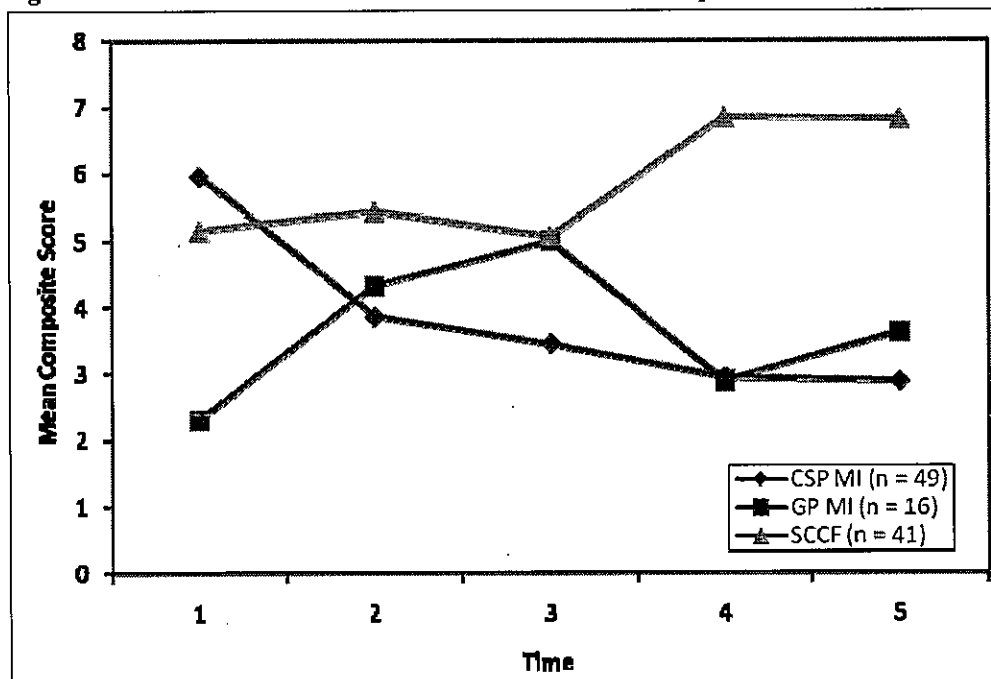
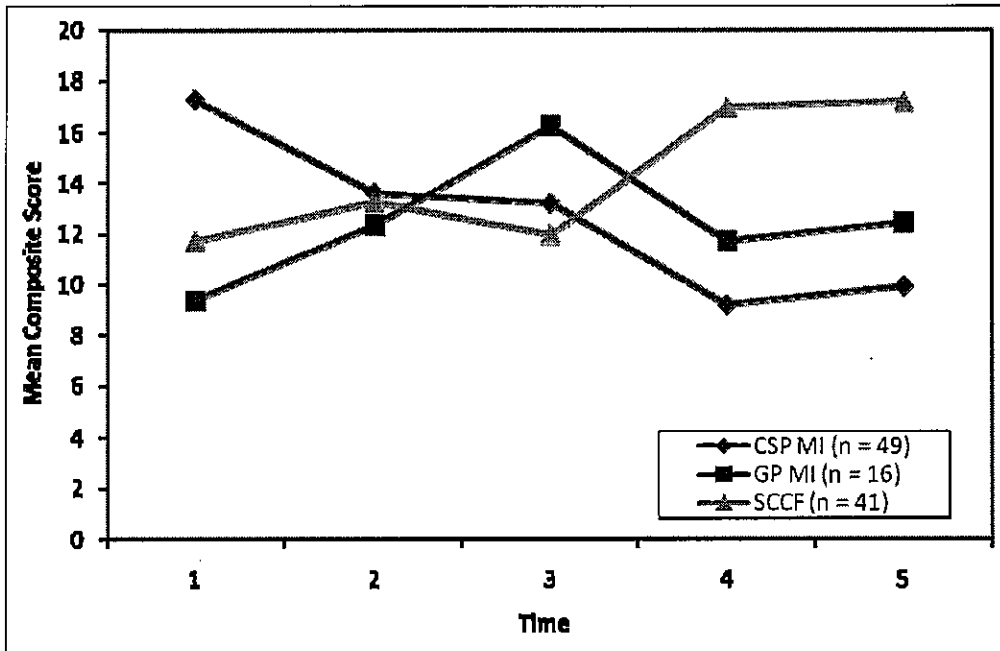


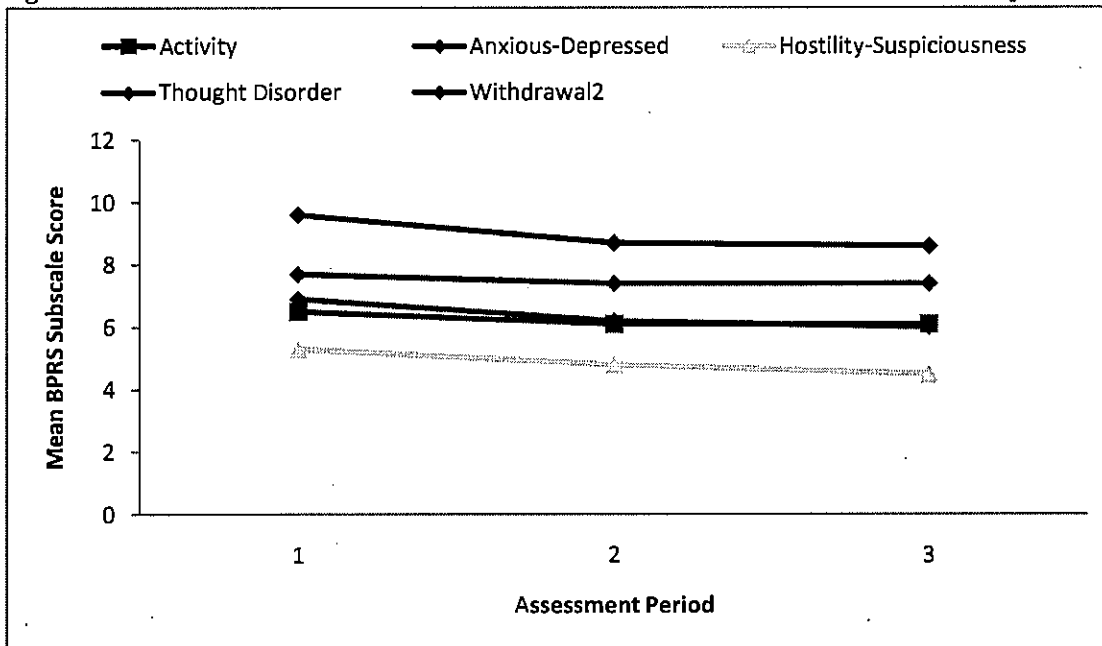
Figure 23. Mean Scores over Time for the PBRs Total Scale for each MI Group



For clinician ratings on the BPRS, there were significant mean group differences for the Total scale, the Thought Disorder subscale, and the Withdrawal subscale. The GP MI group ($M = 31.44, SE = .99$) had significantly lower means on the total scale compared to both the CSP MI ($M = 34.44, SE = .70$) and the SCCF ($M = 35.92, SE = .66$) groups, but there was not a significant difference between the CSP MI and SCCF groups. All three groups were significantly different from each other on the Thought Disorder subscale with the GP MI ($M = 5.43, SE = .32$) having the lowest scores followed by CSP MI group ($M = 6.53, SE = .23$) and then the SCCF group ($M = 7.16, SE = .22$). For the Withdrawal subscale, the SCCF group had significantly higher means ($M = 7.93, SE = .16$) compared to both the CSP MI ($M = 7.55, SE = .17$) and the GP MI ($M = 6.99, SE = .24$) groups, but there was not a significant difference between the CSP MI and GP MI groups.

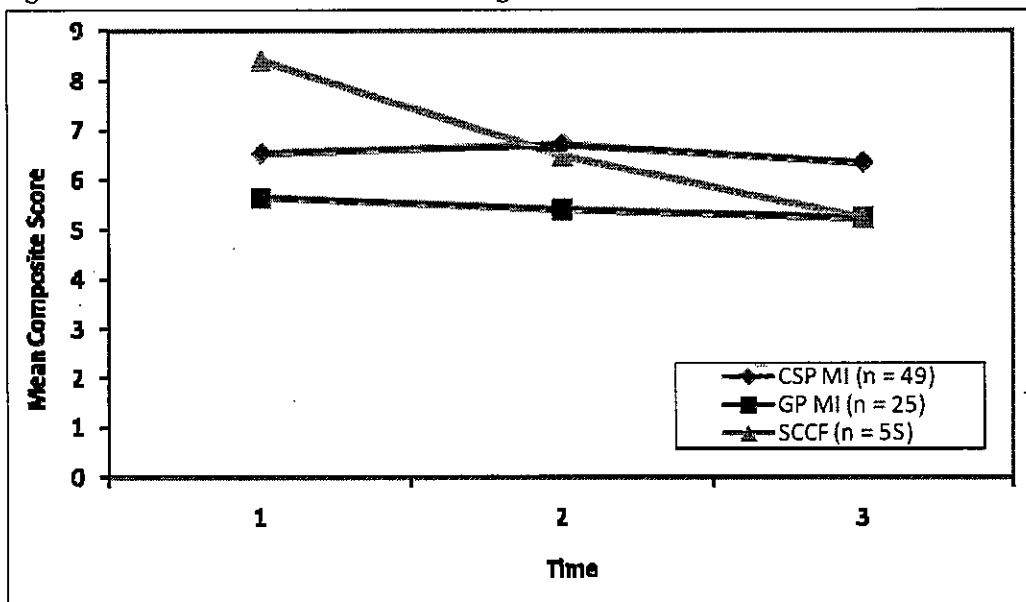
Time effects were statistically significant for all BPRS scales except Activity and Withdrawal subscales; however, there were also significant interactions for Thought Disorder and Total scales. Figure 24 provides the means for change over time for the Activity, Anxious-Depressed, Hostility-Suspiciousness, Thought Disorder, and Withdrawal subscales of the BPRS. For the Anxious-Depressed and Hostility-Suspiciousness scales, mean ratings at the first assessment were greater than mean ratings at the third assessment for all three subscales. Additionally, for the Anxious-Depressed subscale, the mean rating at the first assessment was also significantly greater than the mean at the second assessment.

Figure 24. Mean Scores for the BPRS Subscales over Time Summarized across MI Groups



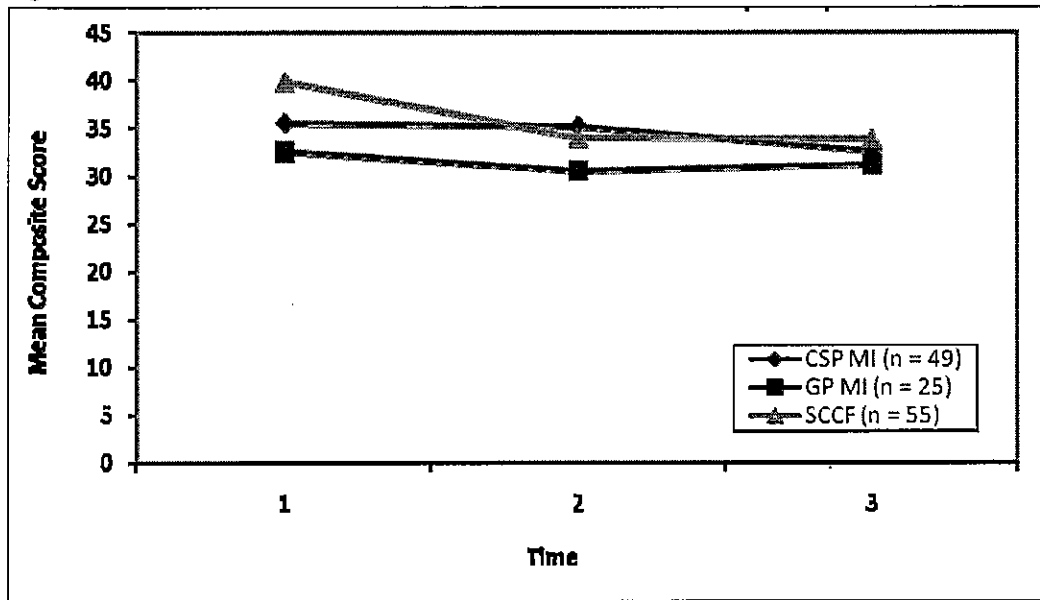
Figures 25 and 26 are mean plots to demonstrate the interaction effects for BPRS Thought Disorder subscale and BPRS Total scores. Simple main effects for each group on the Thought Disorder subscale indicate that there are significant changes over time for the SCCF group but not for the other two groups. The SCCF group had significantly higher scores at the first assessment compared to the other two assessment periods.

Figure 25. Mean Scores for the BPRS Thought Disorder Subscale over Time for each MI Group



For BPRS Total scores, the GP group does not change significantly over time but both of the other groups have significant time effects. In particular, the last assessment scores for the CSP MI group were significantly lower than the first two assessment periods, and the first assessment scores for the SCCF group were significantly higher than each of the other assessment periods.

Figure 26. Mean Scores for the BPRS Total Scale over Time for each MI Group



Slopes Analysis

To compare change over time in another way, slopes analyses were conducted in addition to the means analyses. For these analyses, a slope and intercept were computed for each person on each composite using the available time periods for anyone who completed two or more assessments. These slopes and intercepts were then compared across groups. If AS was impacting change across time, we would expect slopes to be different across study groups. We also computed an intercept value for each person on each self-report variable; these intercepts were computed so that they represented an estimated value at initial assessment. Thus differences in groups would indicate different starting points. As a reminder, for all dependent variables except the SLUMS, lower scores indicate better performance. Thus a positive slope would indicate a worsening of psychological well-being over time and a negative slope would indicate an improvement over time. Similarly, a positive or larger intercept value indicates higher psychological distress (or lower cognitive functioning) at the outset compared to lower (or negative) values for all measures except the SLUMS.

Table 19 gives the means and standard deviations for the slopes and intercepts for each group on each composite and Table 20 provides the statistical results from a one-way analysis of variance on each variable comparing if there are group differences in mean slopes and intercepts.

Table 19. Summary Statistics on Slopes and Intercepts for each Self-Report Variable for the Study Groups

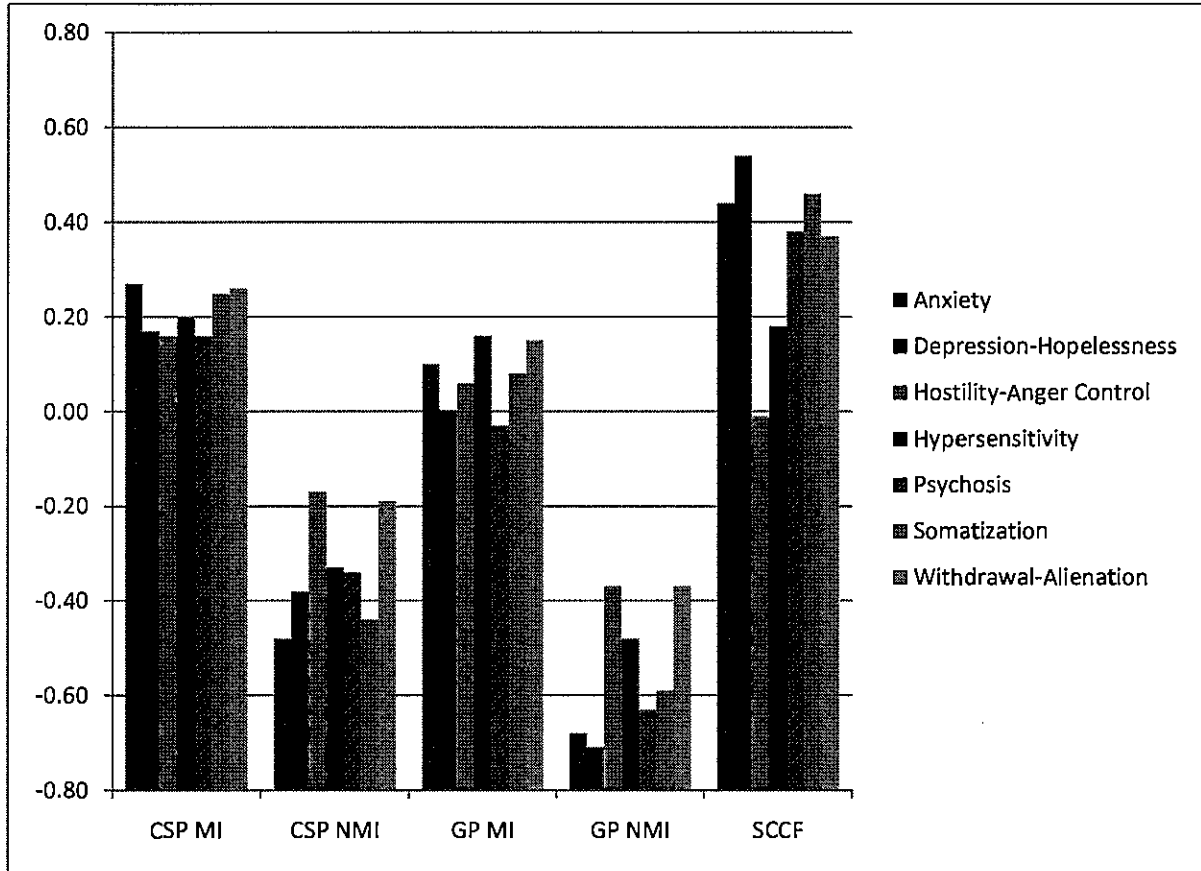
Variable	CSP MI		CSP NMI		GP MI		GP NMI		SCCF		All	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Anxiety												
Intercept	.27	.82	-.48	.56	.10	.73	-.68	.48	.44	.69	-.02	.80
Slope	-.11	.17	-.05	.20	-.06	.20	.03	.10	-.07	.19	-.06	.18
Depression-Hopelessness												
Intercept	.17	.70	-.38	.55	.003	.65	-.71	.40	.54	.84	-.02	.80
Slope	-.07	.13	-.04	.15	-.05	.22	-.01	.10	-.21	.21	-.05	.17
Hostility-Anger Control												
Intercept	.16	.64	-.17	.57	.06	.64	-.37	.50	-.01	.65	-.05	.63
Slope	-.07	.14	-.02	.14	-.04	.15	-.03	.09	-.04	.18	-.04	.14
Hypersensitivity												
Intercept	.20	.78	-.33	.67	.16	.81	-.48	.72	.18	.66	-.04	.77
Slope	-.09	.18	-.05	.26	-.04	.21	-.05	.15	.03	.26	-.04	.23
Psychosis												
Intercept	.16	.72	-.34	.65	-.03	.79	-.63	.63	.38	.80	-.05	.80
Slope	-.08	.19	-.05	.23	-.08	.29	-.04	.10	-.06	.28	-.06	.23
Somatization												
Intercept	.25	.78	-.44	.68	.08	.68	-.59	.52	.46	.66	-.01	.79
Slope	-.08	.17	-.03	.17	-.05	.16	-.04	.10	-.06	.27	-.05	.19
Withdrawal-Alienation												
Intercept	.26	.76	-.19	.71	.15	.78	-.37	.73	.37	.81	.07	.81
Slope	-.02	.18	.05	.18	-.01	.23	.02	.17	.05	.34	.02	.24
SLUMS												
Intercept	20.74	4.71	22.01	3.29	21.93	3.62	23.53	3.16	20.72	3.83	21.61	3.93
Slope	.65	.97	.78	.87	.81	.82	.38	.70	.83	1.42	.70	1.04
Trails B/A												
Intercept	2.94	.94	3.11	1.15	3.14	.87	3.08	1.09	3.05	.99	3.05	1.02
Slope	-.10	.22	-.06	.30	-.15	.36	-.07	.28	-.98	.40	-.09	.32

Table 20. F Statistics and Partial η^2 Comparing Study Groups on Slopes and Intercepts

Self-Report Measures	Intercept Comparisons	Slope Comparisons
Anxiety	$F(4, 257) = 27.14, p < .001, \eta^2 = .30$	$F(4, 257) = 1.39, p = .24, \eta^2 = .02$
Depression-Hopelessness	$F(4, 257) = 28.62, p < .001, \eta^2 = .31$	$F(4, 257) = 1.30, p = .27, \eta^2 = .02$
Hostility-Anger Control	$F(4, 257) = 5.86, p < .001, \eta^2 = .08$	$F(4, 257) = .88, p = .48, \eta^2 = .01$
Hypersensitivity	$F(4, 257) = 10.23, p < .001, \eta^2 = .14$	$F(4, 257) = 2.84, p = .02, \eta^2 = .04$
Psychosis	$F(4, 257) = 16.34, p < .001, \eta^2 = .20$	$F(4, 257) = .24, p = .92, \eta^2 = .004$
Somatization	$F(4, 257) = 24.10, p < .001, \eta^2 = .27$	$F(4, 257) = .58, p = .68, \eta^2 = .01$
Withdrawal-Alienation	$F(4, 257) = 9.00, p < .001, \eta^2 = .12$	$F(4, 257) = 1.18, p = .32, \eta^2 = .02$
SLUMS	$F(4, 257) = 4.50, p = .002, \eta^2 = .07$	$F(4, 257) = 1.46, p = .21, \eta^2 = .02$
Trails B/A	$F(4, 257) = .30, p = .88, \eta^2 = .005$	$F(4, 257) = .54, p = .71, \eta^2 = .008$

As might be expected, there were significant group differences on intercepts for each self-report variable except for the Trails B/A derived task. Figure 27 provides the mean intercept values for each self-report composite for each group. In general, the MI groups had worse performance on these psychological variables compared to the NMI groups. The exceptions to this general finding is that the CSP NMI was not significantly different from SCCF and GP MI on hostility composite and was not different from GP MI on psychosis and withdrawal-alienation composites. Similarly, the GP NMI group was not significantly different from the GP MI on SLUMS. Another general finding for intercept differences is that groups with the same mental health status (MI, NMI) tended to be similar to one another. The exceptions to this general finding were that for the MI groups, the SCCF group was significantly higher than the GP MI group on depression-hopelessness, psychosis, and withdrawal-alienation composites.

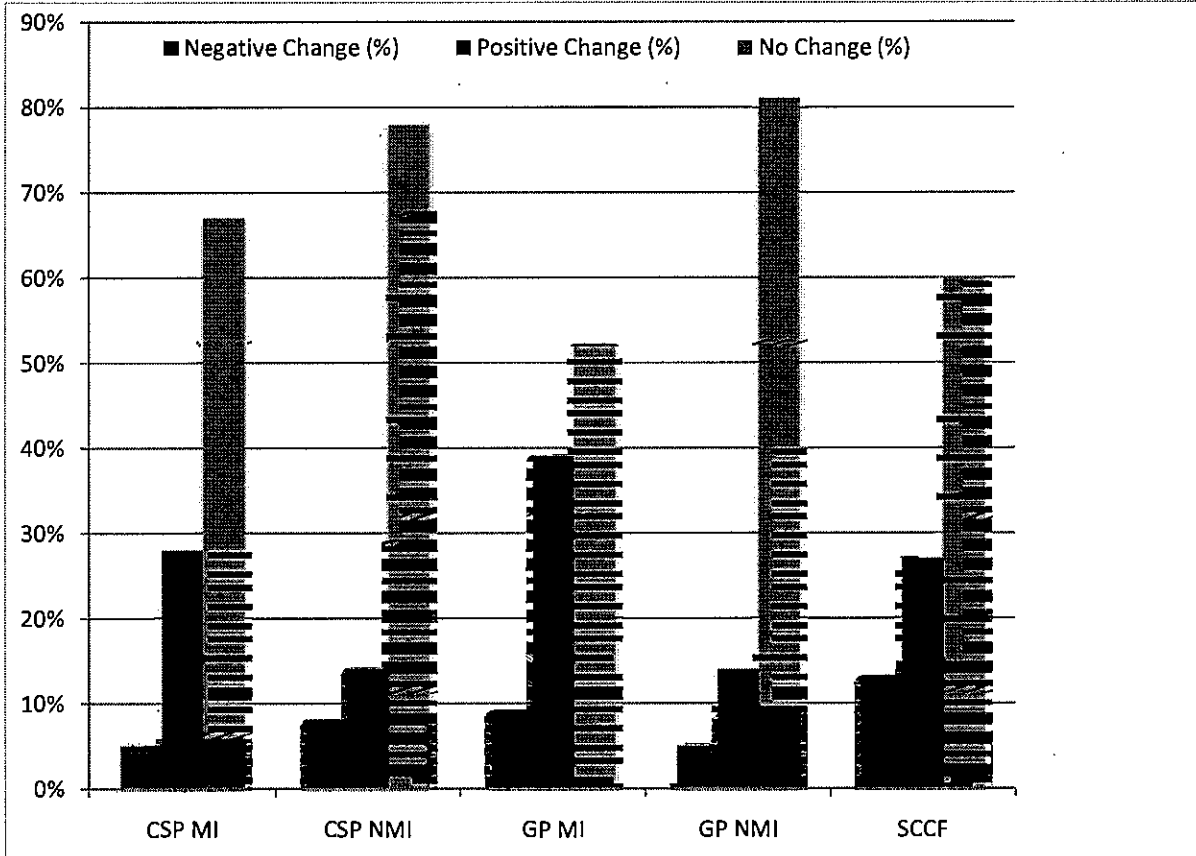
Figure 27. Mean Intercept Values for each Composite by Study Group



In contrast to intercept analyses that showed many group differences, only hypersensitivity had significant differences in slopes between groups (see Table 20). The only significant differences were between the CSP MI group and the SCCF group. The CSP MI group had a negative slope indicating improvement over time and the SCCF group had a positive slope indicating a worsening trend over time.

To better understand how change is occurring across groups, we identified participants as having positive, negative, or no change over time. Participants were classified as positive changers if they had strong positive change on at least one variable (i.e., slope was more than 2 standard deviations from mean and in the direction of positive change) or had smaller positive change on three or more variables (i.e., slopes on at least three variables were more than 1 standard deviation from mean and in the direction of positive change). Likewise, participants were classified as negative changers if slopes were in negative direction. The remaining participants who did not meet the rules for either positive or negative change were classified as not changing. Figure 28 provides the percentage of change types for each study group.

Figure 28. Percentage of Change Types by Study Group



Groups were significantly different in the percentage of change types, $\chi^2(8, N = 270) = 16.26, p = .04$. Using standardized residuals, the following conditions were found to be different from expectations: for the CSP MI group, there was a lower than expected percentage of persons changing negatively (5% vs. 8%); for the CSP NMI group, there was a lower percentage of people changing in positive direction than expected (14% vs. 24%); for the GP MI, there was a higher percentage of people changing in the positive direction (39% vs. 24%) and fewer than expected stable patterns (52% vs. 68%); for the GP NMI group, there was a lower percentage of people changing positively (14% vs. 24%) and more stable patterns than expected (81% vs. 68%); and for the SCCF group, there was a higher percentage of persons changing negatively (13% vs. 8%).

PREDICTOR ANALYSES

The purpose of these analyses was to explore if there were predictors of the rate of change across time on each composite. Using regression analyses to predict individual slopes as the dependent variable, we examined if the variables listed in Table 23 could explain rate of change. These variables were identified by the literature or the study advisory board as potential predictors. All study participants are used in these analyses, and MI status and AS status are used as two of the predictors.

Table 23. Variables Used to Predict Change over Time

Predictor	Notes
MI status (study group membership)	0: NMI group; 1: MI group
AS status (study group membership)	0: Not AS; 1: AS
Demographics	
Age (at start of study)	DCIS variable
Education	DCIS; 0: Less than HS; 1: HS diploma or GED
Minority status	DCIS; 0: Not a minority; 1: Minority
Criminal History	
Offense degree	DCIS
Previous AS confinement	DCIS; 0: No; 1: Yes
Prior incarcerations	DCIS
Gang membership	DCIS; 0: No; 1: Known gang member
Psychological History	
Anger needs	DCIS
Anti-social personality disorder	CCI
Anxiety (Axis I)	CCI
Avoidant personality disorder	CCI
Borderline personality disorder	CCI
Dependent personality disorder	CCI
Depression (Axis I)	CCI
Depressive personality disorder	CCI
History of deliberate self harm	DSHI (life time incidence)
Histrionic personality disorder	CCI
Impulsivity	CCI
Narcissistic personality disorder	CCI
Obsessive-compulsive personality disorder	CCI
Paranoid personality disorder	CCI
Passive-aggressive personality disorder	CCI
Psychotic thinking (Axis I)	CCI
Post traumatic stress disorder (Axis I)	CCI
PSI Attitudes towards AS	PSI Time 1
PSI Fear Level	PSI Time 1
PSI Safety	PSI Time 1
Sadistic personality disorder	CCI
Schizoid personality disorder	CCI
Schizophrenia (Axis I)	CCI
Schizotypal personality disorder	CCI
Self-defeating personality disorder	CCI
Self-destruction needs	DCIS
Sex offender needs	DCIS
Social Phobia (Axis I)	CCI
Trauma symptoms	TSI
Withdrawal (Axis I)	CCI

To determine which variables are potential predictors of self-report outcomes, a forward statistical regression was used. The information in Table 24 presents the regression analysis results providing the adjusted R² (proportion of variance explained in the slope variable by the predictors) and lists which variables were found to be significant predictors, along with standardized regression coefficients. The sign of the regression coefficient provides information about the direction of the relationship between the dependent variable and the predictor. Recall that the dependent variable is rate of change (slope) with positive scores indicating worsening of performance over time for all variables except on the SLUMS and negative scores indicating improving performance over time for all variables except on the SLUMS. Thus, a negative relationship of a

predictor with a slope implies that high scores on a variable covary with lower slope scores and thus more improvement for all variables (except SLUMS where a negative relationship implies higher scores on predictor goes with more decline in SLUMS performance over time).

Table 24. Regression Results: Significant Predictors of Rate of Change over Time in Composite Variables

Construct	Adj. R ²	Significant Predictors
Anxiety	.11***	Schizophrenia (Axis I; $\beta = -0.32$) PSI Safety ($\beta = -0.20$) TSI Total ($\beta = 0.21$)
Depression-Hopelessness	.17***	Paranoid PD ($\beta = -0.23$) Sadistic PD ($\beta = 0.19$) TSI Total ($\beta = 0.30$) Schizophrenia (Axis I; $\beta = -0.38$)
Hostility-Anger Control	.12***	Passive-Aggressive PD ($\beta = -0.29$) TSI Total ($\beta = 0.16$) Withdrawal (Axis I; $\beta = -0.21$) PSI Safety ($\beta = 0.19$) Anger Needs ($\beta = 0.13$)
Hypersensitivity	.04***	DSHI ($\beta = -0.22$)
Psychosis	.03**	Narcissistic PD ($\beta = -0.20$)
SLUMS	.08***	Obsessive-Compulsive PD ($\beta = 0.30$) PSI Fear Level ($\beta = -0.24$)
Somatization	.09***	PSI Safety ($\beta = -0.26$) Narcissistic PD ($\beta = -0.14$)
Withdrawal-Alienation	.03**	DSHI ($\beta = -0.18$)
Trails B/A	.08***	PSI AS Attitude ($\beta = -0.25$) Antisocial PD ($\beta = 0.19$) Narcissistic PD ($\beta = -0.22$) Depressive PD ($\beta = 0.20$)

* $p < .05$; ** $p < .01$; *** $p < .001$

The study variables—MI status and AS confinement—were never significant predictors of outcomes. There were 15 different significant predictors on at least one outcome. For a predictor to have practical meaning in an applied setting, it would be important for predictors to be related to multiple outcome variables. There were no predictors that were significantly related to a majority of outcomes; however there were predictors that were significantly related to change over time on two or three multiple constructs. These were trauma history (positive relationship with change), PSI Safety (both positive and negative relationships with change), narcissistic personality disorder (negative relationship with change), schizophrenia scores (negative relationship with change), and history of self harm (negative relationship with change). To provide an interpretive example, the positive relationships between trauma history and slopes for the anxiety composite indicate that higher scores on trauma co-vary with higher anxiety slopes. This implies that more trauma leads to worsening over time on the anxiety composite. Thus, generalizing to significant predictors, higher scores on trauma and lower scores on narcissistic personality disorder, schizophrenia (axis I), and self harm, lead to more negative change over time. The PSI Safety subscale had both positive and negative relationships with outcome variables. Higher scores on the PSI Safety subscale (i.e., feeling safer in AS) was related to improvements in anxiety and somatization but also to more hostility over time.

DISCUSSION

The results of this study were largely inconsistent with our hypotheses and the bulk of literature that indicates AS is extremely detrimental to inmates with and without mental illness. We hypothesized that inmates in segregation would experience greater psychological deterioration over time than comparison inmates, who were comprised of similar offenders confined in non-segregation prisons. Similar to other research, our study found that segregated offenders were elevated on multiple psychological and cognitive measures when compared to normative adult samples (Andersen et al, 2000; Haney, 2003; Suedfeld et al., 1982; Zinger et al., 2001). However, there were elevations among the comparison groups too, suggesting that high degrees of psychological disturbances are not unique to the AS environment. The GP NMI group was the only one that was similar to the normative group on a number of scales.

In examining change over time patterns, there was initial improvement in psychological well-being across all study groups, with the bulk of the improvements occurring between the first and second testing periods, followed by relative stability for the remainder of the study. On only one measure – withdrawal – did offenders worsen over time, but this finding was only true for the two NMI groups, so it is not attributable to AS. Even given the improvements that occurred within the study timeframe, the elevations in psychological and cognitive functioning that were evident at the start of the study remained present at the end of the study.

Another hypothesis was that offenders with mental illness would deteriorate over time in AS at a rate more rapid and more extreme than for those without mental illness. Patterns indicated that the MI groups (CSP MI, GP MI, SCCF) tended to look similar to one another but were significantly elevated compared to the NMI groups (CSP NMI, GP NMI), regardless of their setting. For the AS offenders, the MI group scored worse than the NMI group on all self-report measures except the Trails test and all staff measures except the PBRs Anti-Authority scale. In addition to the changes over time described above, PBRs scores decreased significantly for segregated inmates regardless of their mental health status, which would be an indicator that staff may be perceiving improvements, but the significant differences were from the first to the second assessment periods when the majority of participants changed facilities, which suggests this is perhaps a measurement error rather than a true improvement. As hypothesized there was a differential time effect for the MI and NMI groups on several composite measures (i.e., anxiety, hostility-anger control, hypersensitivity, somatization), but the interactions were in the opposite direction of our hypothesis; on average, the CSP NMI group did not change while the CSP MI group improved.

We stated that offenders in segregation would develop an array of psychological symptoms consistent with the SHU syndrome. As already discussed, all of the study groups, with the exception of the GP NMI group, showed symptoms that were associated with the SHU syndrome. These elevations were present from the start and were more serious for the mentally ill than non-mentally ill. In classifying people as improving, declining, or staying the same over time, the majority remained the same. There was a small percentage (7%) who worsened and a larger proportion (20%) who improved. Therefore, this study cannot attribute the presence of SHU symptoms to confinement in AS. The features of the SHU syndrome appear to describe the most disturbed offenders in prison, regardless of where they are housed. In fact, the group of offenders who were placed in a psychiatric care facility (SCCF) had the greatest degree of psychological disturbances and the greatest amount of negative change.

Finally, in this study, we conducted some exploratory predictive analyses to determine if there were individual characteristics that could identify who may be at greater risk of psychological harm from segregation. There were no individual predictors that showed strong effects for predicting change. This could indicate that we did not have the correct predictors or that patterns of decompensation are individualized (i.e., not predictable), but it is more likely that the relative stability over time makes it difficult to predict change.

A review of the findings warrants a discussion of plausible alternative explanations that might account for our results. The use of a repeated measures design enabled us to determine that change was occurring and in which direction. Even given the debate about whether or not harmful effects resulted from AS, it was never suggested that inmates might improve as this study found. The presence of comparison groups avoids an attribution error; the changes, improvements in this case (i.e., 20%), are not due to segregation. These conclusions replicate those drawn by Zinger and colleagues (2001) where there was a similar lack of evidence of harm. These studies suffered criticism for high refusal rates, high attrition rates, small sample sizes, and short durations – limitations that were corrected in the present study (note, however, that no generalizations should be made beyond the 1 year follow-up period in this study). Furthermore, the use of reliable and valid standardized measures enabled the present research study to assess psychological functioning in an objective manner. Although the majority of these tests were not normed for prisoner populations, the current reliability and validity findings increased our confidence in these measures.

The most difficult finding to interpret is the improvement that occurred between the first and second testing sessions, which was significant for all groups except the GP NMI group. This effect may be due to reactivity—the participants know they are in a study and respond in a particular way. Perhaps they have a need to respond in a way that puts them in the most favorable light (e.g., ability to handle demands of prison confinement). (Sometimes improvement in performance due to being observed is called the Hawthorne effect; however this effect seems to be misunderstood and it was not merely the fact of being studied that led to those original finding of improvement [Gottfredson, 1996]). It is also possible that there are demand characteristics introduced by the field researcher that cues the participants on how to respond; this seems unlikely as the participants would be expected to respond in the hypothesized direction. Although a testing or practice effect might explain the improvements on cognitive measures, we were unable to find support in the literature or from the study advisory board that psychological measures should be influenced by testing effects. Because the changes occurred in the AS and comparison groups, it is not possible to attribute the improvements to the confinement conditions; however it may be that participating in the study produces some unknown expectation. Although study demands may lead to positive ratings, it seems unlikely that these response biases would overshadow the negative impacts of AS if they really existed. However, there is not enough information in the data collected to understand the reason for the positive change. The most likely explanation is that study participants were included in our study when they were in the midst of a crisis and, with time, the crisis dissipated.

LIMITATIONS

This study was able to incorporate several design features that improved upon the capability of previous research to draw conclusions about the effects of AS. On the other hand, this study has several limitations that affect its generalizability to other settings. First, this study included literate adult male offenders and should therefore not be generalized to female offenders, illiterate offenders, or juveniles. Second, this study can only be generalized to other prison systems to the extent that their conditions of AS confinement are

similar to Colorado's. The same findings might not be found in other AS units that have different offender populations or criteria for placement, more restrictive confinement, or fewer mental health services.

The duration of this study was limited to one year. We believed this time period to be adequate to detect harmful effects because it was postulated in earlier research that the effects of segregation would be quickly evident (Grassian & Friedman, 1986; Haney, 2003; Kupers, 2008; Lovell et al., 2007; Rhodes, 2004; Toch, 1992). Kupers (2008, p. 1006) stated "that for just about all prisoners, being held in isolated confinement for longer than 3 months causes lasting emotional damage if not full-blown psychosis and functional disability." Therefore, we expected that deleterious effects would become evident within a year, but it is possible they do not appear until after longer periods of segregation.

This study used a moderate sample size because we anticipated moderate to large effects based on the literature (e.g., Grassian, 1983; Haney, 2003; Pizarro & Stenius, 2004). It is possible that the true magnitude of the negative effect is small and, therefore, larger sample sizes would be required to detect negative changes and predict the types of offenders who might be harmed by segregation. In support of this postulate, the present study found small to moderate effect sizes for change over time, however they were in the opposite (positive) direction.

This study examined group averages. It was not designed to identify if certain individuals might be worsened by the conditions of AS; rather the purpose was to examine whether offenders on the whole, both mentally ill and non-mentally ill, are harmed by long-term segregation. Also, in the design of this study, we assumed a general linear trend in the data and were not able to capture nonlinear changes over time that might have occurred. It is possible that a person in segregation could have had one or more brief episodes, possibly even severe episodes, of psychopathology that were not reflected in our data because testing occurred at three month intervals and that would not have been reflected in trend analyses of their psychological functioning. This study was not designed to assess brief changes in psychological functioning, however serious.

This study attempted to triangulate data between inmate self-report, staff observations, and official records. In the research study, we had the largest degree of success in gathering self-report data. Some may question whether inmates' self-report is reliable because they may have reason to exaggerate their symptoms, but our testing of the measures' psychometric properties indicated that the participants responded in remarkably reliable and valid ways. The official record data, which was intended to help us understand the varying degree of social isolation to which study participants were exposed, was inconsistent and incomplete. Because our findings did not show negative change over time, the official record data would not have been as useful as originally intended; however, it would have still been beneficial in describing the conditions of confinement. Additionally, the data from the clinical staff suggested that there were issues with the BPRS data, where clinicians were able to rank order groups, but they did not estimate elevations to similar heights as those reported by inmates. These data also raise the possibility that clinical staff are aware of less distress than inmates validly report.

FUTURE RESEARCH

The definition of AS varies greatly from state to state, so much so that it is difficult to define or count the number of inmates held nationally in AS; therefore, replication is needed in other prison systems to determine whether these findings still hold true when the conditions of confinement are varied. The present research was unable to determine which elements of CSP were essential to prevent harm from occurring, and

it did not assess the interventions used at CSP to monitor and treat inmates. Ongoing research is needed to better understand how the different components of segregation may impact offenders differentially. For example, the type and intensity of psychiatric services provided to AS inmates may have a particularly strong effect on whether they decompensate during extended periods of segregation. Research that incorporates qualitative data, such as mental health records and historical patient records, may also help to understand how individuals are impacted by their confinement conditions. It is important to study other high-security settings that permit more out of cell time or increased interpersonal contact (i.e., group treatment).

Similar research is needed with female offenders in AS. Although they represent a small percentage of the AS population, there is a stunning lack of information about the pathways that lead women to segregation or how they adapt to this environment. Women offenders have high rates of mental illness (James & Glaze, 2006; O'Keefe & Schnell, 2008). Trauma appears to be a major determinant of mental illness in female offender populations (Green, Miranda, Daroowalla, & Siddique, 2005; Zlotnik & Pearlstein, 1997); incarcerated women report much higher rates of abuse than incarcerated men (McClellan, Farabee, & Crouch, 1997). In examining coping mechanisms, the most unique gender differences exhibited by women are their strong need for social interaction and their propensity to cope with the prison environment predominantly through relationship formation (Severance, 2005). Given their higher rates of mental illness, trauma history, and needs for social interaction, women may be particularly vulnerable to potentially harmful effects of segregation.

One untapped topic in the area of segregation research is the role of staff, both how they affect the setting and the effects of the setting on them (Haney, 2008; Mears, 2008; Pizarro & Narag, 2008). Recognizing that correctional officers have the greatest amount of contact with offenders, Dvoskin and Spiers (2004) have suggested they need a larger role in the treatment and psychiatric care of offenders, including, but not limited to, psychotherapeutic techniques to diffuse crisis situations, consultation with mental health professionals, and monitoring inmates' compliance and adjustment to psychotropic medications. Evaluation research would benefit the field in understanding the effectiveness of staff training and intervention programs implemented in AS. Finally, assessing inmate perceptions of staff may have value in understanding the impact of long-term segregation on inmates because how they are treated may have a significant impact on their adjustment to AS.

There were some findings in this study that were difficult to interpret or that did not fit into the same general patterns described above. The Trails test did not differentiate between groups as did the other measures. In contrast, the BPRS tended to differentiate between offenders with and without mental illness even better than did the self-report measures. The hypersensitivity construct showed more variability and more differential changes over time than the other constructs; for offenders in AS, those with mental illness improved in hypersensitivity between times 1 and 2 but then showed a worsening trend while those without mental illness significantly improved between times 1 and 2, but then worsened between times 2 and 3, and then improved at the fourth interval. Further research may be needed to explore the reason for these different patterns or to determine if these were spurious findings.

POLICY IMPLICATIONS

Does this study legitimize the use of segregation with offenders, including those with serious and persistent mental illness? Because this study may not generalize to other prison systems, especially those that have conditions of confinement dissimilar to CSP, it is not possible to conclude that AS is not detrimental for all

offenders. Systems that are more restrictive and have fewer treatment and programming resources should not generalize these findings to their prisons. Replication is needed to understand how increased services, privileges, and out of cell time ameliorate the unintended consequences of AS, and research needs to inform prison systems about the standards and practices necessary to protect inmates in segregation from harmful psychological effects.

It is also important to note that there may be other negative consequences of AS that we did not study. For example, Lovell et al. (2007) found that inmates released directly from segregation to the streets had dramatically higher rates and severity of detected recidivism than AS inmates who first released to GP (but see Mears & Bales, 2009). We also did not study the degree to which AS met its purported goal of changing inmate behavior for the better over time. The only questions addressed by this study were related to psychological changes over time in segregation. Thus, we make no empirical or value judgments about whether and to what degree the use of AS balances the benefits (e.g., a safer prison system) with costs (e.g., significant reductions in freedom).

It is impossible to ignore the extremely disproportionate rate at which inmates with serious mental illness are assigned to AS (Lovell, 2008; Metzner & Fellner, 2010; O'Keefe, 2008a), which has to some degree "shocked the conscience" of the courts (see *Jones 'El v. Berge*, 2001; *Madrid v. Gomez*, 1995; *Ruiz v. Johnson*, 1999). In an era when prisons are expected to implement evidence-based practices and to rehabilitate offenders who will be releasing back to the community, is it enough to avoid harm? Must we ask ourselves another question: what are the conditions required to *improve* inmates' mental well-being while in segregation? Prison systems are held to a standard of treatment that is at least equivalent to community standards. It is likely that this most difficult segment of society has failed at all levels of community treatment and earlier criminal justice interventions, but the quest to treat and improve services for the most needy is an important reality facing corrections agencies.

Regarding their psychological functioning and levels of distress, these data suggest, although the differences were small, that inmates with serious mental illness are less likely to improve in segregation *and* are less likely to get worse compared to mentally ill inmates in GP. We do not assume that the reasons for these apparently contradictory findings are the same. For example, it is possible that fewer inmates with mental illness get worse because segregation is a safer and more structured environment. On the other hand, hypotheses regarding their unlikeliness to improve include the significant limitations that segregation places on various types of therapeutic activities and services such as group therapy. Further, the data do not tell us which aspects of AS prevent psychological improvement and deterioration, respectively, among inmates with mental illness. However, since prisons have a constitutional duty to respond to serious medical (including psychiatric) needs, the possibility that segregation may prevent improvement is cause for concern and further study.

There remain significant implications for mental health staff who work in prison systems that permit the placement of mentally ill offenders in long-term segregation. It is critical for mental health staff to screen and assess offenders prior to AS placement to determine their vulnerability to harm that might occur as a result of their segregation. While in segregation, it is important that the mental status of all offenders be assessed on a frequent, regular basis through rounds and individual sessions. Prison systems need to have a range of confinement options, such that offenders who are at risk of or are showing signs of decompensa-

tion can be removed from segregation and placed in an alternative high security environment that permits greater out of cell time and interaction with others.

Other systems have rejected confinement models that isolated offenders and held them in extremely restrictive spaces. Even if the segregation models of the early 1900's and the state psychiatric hospitals of the mid-19th century are viewed as "primitive" compared to modern-day AS facilities, it is important to examine and understand why these models failed and were ultimately dismantled. Although there are a number of researchers who predict that there is no end in sight to the supermax model (King, 1999; Mears, 2008; Pizarro & Narag, 2008; Pizarro & Stenius, 2004), they have also raised empirical questions regarding their efficacy. Questions about the efficacy of AS will be asked until more is known about whether the use of AS in prison systems improves conditions for the rest of the system, whether and how they improve inmate behavior within and beyond the prison walls, whether they are cost-effective, whether they increase risks to public safety, and whether there are settings or individuals that are prone to psychological deterioration.

REFERENCES

- American Correctional Association (2009). *2009 directory: Adult correctional departments, institutions, agencies, and probation and parole authorities*. Alexandria, VA: Author.
- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). Washington, DC: Author.
- American Psychological Association. (2009). *Publication Manual of the American Psychological Association* (6th ed.). Washington, DC: Author.
- Andersen, H. S., Sestoft, D., Lillebaek, T., Gabrielsen, G., & Hemmingsen, R. (2003). A longitudinal study of prisoners on remand: Repeated measures of psychopathology in the initial phase of solitary versus nonsolitary confinement. *International Journal of Law and Psychiatry*, 26, 165-177. Retrieved from <http://sciencedirect.com>
- Andersen, H. S., Sestoft, D., Lillebaek, T., Gabrielsen, G., Hemmingsen, R., & Kramp, P. (2000). A longitudinal study of prisoners on remand: Psychiatric prevalence, incidence, and psychopathology in solitary vs. non-solitary confinement. *Acta Psychiatrica Scandinavica*, 102, 19-25.
- Andrews, D. A., & Bonta, J. L. (1995). *Level of Service Inventory—Revised*. Toronto, ON: Multi-Health Systems.
- Andrews, D. A., & Bonta, J. L. (2003). *Level of Service Inventory—Revised (LSI-R): U.S. norms manual supplement*. North Tonawanda, NY: Multi-Health Systems, Inc.
- Arrigo, B. A., & Bullock, J. L. (2008). The psychological effects of solitary confinement on prisoners in super-max units: Reviewing what we know and recommending what should change. *International Journal of Offender Therapy and Comparative Criminology*, 52, 622-641. doi: 10.1177/0306624X07309720
- Atherton, E. (2001). Incapacitation with a purpose. *Corrections Today*, 63, 100-103. Retrieved from <http://vnwebnhwwilsonweb.com>
- Austin, J., Alexander, J., Anuskiewicz, S., & Chin, L. (1995). *Evaluation of the Colorado Objective Prison Classification System*. San Francisco, CA: National Council on Crime and Delinquency.
- Beck, A. T., & Steer, R. A. (1993). *Beck Hopelessness Scale: Manual*. San Antonio, TX: The Psychological Corporation—Harcourt Brace & Company.
- Benjamin, T. B., & Lux, K. (1975). Constitutional and psychological implications of the use of solitary confinement: Experience at the Maine state prison. *Clearinghouse Review*, 9, 83-84.
- Briere, J. (1995). *Trauma Symptom Inventory: Professional manual*. Lutz, FL: Psychological Assessment Resources, Inc.

- Briggs, C. S., Sundt, J. L., & Castellano, T.C. (2003). The effects of supermaximum security prisons on aggregate levels of institutional violence. *Criminology*, 41, 1341-1376. Retrieved from <http://www3.interscience.wiley.com/cgi-bin/fulltext/118866214/PDFSTART>
- Brodsky, S., & Scogin, F. (1988). Inmates in protective custody: First data on emotional effects. *Forensic Reports*, 1, 267-280.
- Burger, K. G., Calsyn, R. J., Morse, G. A., Klinkenberg, W. D., & Trusty, M. L. (1997). Factor structure of the Expanded Brief Psychiatric Rating Scale. *Journal of Clinical Psychology*, 53, 451-454. doi: 10.1002/(SICI)1097-4679(199708)53:5<451::AID-JCLP5>3.0.CO;2-Q
- Cohen, F. (2008). Penal Isolation: Beyond the Seriously Mentally Ill. *Criminal Justice and Behavior*, 35, 1017-1047.
- Collins, W. C. (2004). *Correctional law for the correctional officer* (4th ed.). Lanham, MA: American Correctional Association.
- Cooke, D. J. (1998). The development of the Prison Behavior Rating Scale. *Criminal Justice and Behavior*, 25, 482-506. doi: 10.1177/0093854898025004005
- Coolidge, F. L. (2004). *An intraduction to the Coolidge Correctional Inventory (CCI)*. Retrieved from <http://www.uccs.edu/~faculty/fcoolidge/>
- CTB/McGraw-Hill. (1994). *Tests of Adult Basic Education (TABE): Versian 7-8*. Monterey, CA: CTB/McGraw-Hill.
- CTB/McGraw-Hill. (2004). *Tests of Adult Basic Education (TABE): TABE Form 7 to the 2002 GED Linking Report*. Retrieved from http://www.ctb.com/media/articles/pdfs/AdultEducation/TABE_Form7_2002_GED_Linking_Report.pdf?FOLDER<>folder_id=59997&bmUID=1089131925485
- Derogatis, L. (1993). *Brief Symptom Inventory: Administration, scoring, and procedures manual*. Minneapolis, MN: NCS Pearson, Inc.
- Ditton, P. M. (1999, July). *Mental health and treatment of inmates and probationers*. Washington, DC: U.S. Department of Justice, Bureau of Justice Statistics.
- Dvoskin, J. A., & Spiers, E. M. (2004). On the role of correctional officers in prison mental health. *The Psychiatric Quarterly*, 75, 41-59.
- Ecclestone, C. E. J., Gendreau, P., & Knox, C. (1974). Solitary confinement of prisoners: An assessment of its effects on prisoners' personal constructs and andrenocortical activity. *Canadian Journal of Behavioral Science*, 6, 178 – 191.
- Gawande, A. (2009, March 30). Hellhole: The United States holds tens of thousands of inmates in long-term solitary confinement. Is this torture? *The New Yorker*, 36-45. Retrieved from http://www.newyorker.com/reporting/2009/03/30/090330fa_fact_gawande

- Gendreau, P., & Bonta, J. (1984). Solitary confinement is not cruel and unusual punishment: People sometimes are! *Canadian Journal of Criminology*, 26, 467-478.
- Gendreau, P. E., Freedman, N., Wilde, G. J. S., & Scott, G. D. (1968). Stimulation seeking after seven days of perceptual deprivation. *Perceptual and Motor Skills*, 26, 547-550.
- Gendreau, P. E., Freedman, N., Wilde, G. J. S., & Scott, G. D. (1972). Changes in EEG alpha frequency and evoked response latency during solitary confinement. *Abnormal Psychology*, 79, 54-59.
- Gendreau, P., McLean, R., Parsons, T., Drake, R., and Ecclestone, J. (1970). Effect of two days' monotonous confinement on conditioned eyelid frequency and topography. *Perceptual and Motor Skills*, 31, 291-293.
- Gottfredson, G. D. (1996). The Hawthorne misunderstanding (and how to get the Hawthorne effect in action research). *Journal of Research in Crime and Delinquency*, 33, 28-48. doi: 10.1177/0022427896033001003
- Grassian, S. (1983). Psychopathological effects of solitary confinement. *American Journal of Psychiatry*, 140, 1450-1454.
- Grassian, S., & Friedman, N. (1986). Effects of sensory deprivation in psychiatric seclusion and solitary confinement. *International Journal of Law and Psychiatry*, 8, 49-65.
- Gratz, K. L. (2001). Measurement of deliberate self-harm: Preliminary data on the deliberate self-harm inventory. *Journal of Psychopathology and Behavioral Assessment*, 23, 253-263.
- Green, B. L., Miranda, J., Daroowalla, A., & Siddique, J. (2005). Trauma exposure, mental health functioning and program needs of women in jail. *Crime & Delinquency*, 51, 133-151. doi: [10.1177/0011128704267477](https://doi.org/10.1177/0011128704267477)
- Haney, C. (1993). Infamous punishment: The psychological effects of isolation. *National Prison Project Journal*, 8, 3-21.
- Haney, C. (2003). Mental health issues in long-term solitary and "supermax" confinement. *Crime and Delinquency*, 49, 124-156.
- Haney, C. (2008). A culture of harm: Taming the dynamics of cruelty in supermax prisons. *Criminal Justice and Behavior*, 35, 956-984.
- Hedlund, J. L., & Vieweg, B. W. (1980). The Brief Psychiatric Rating Scale (BPRS): A comprehensive review. *Journal of Operational Psychiatry*, 11, 48-64.
- Hewitt v. Helms, 459 U.S. 460 (1983).
- Hodgins, S., & Côté, G. (1991). The mental health of penitentiary inmates in isolation. *Canadian Journal of Criminology*, 33, 175-182.
- Human Rights Watch. (1997). *Cold storage: Super-maximum security confinement in Indiana*. New York, NY: Human Rights Watch.

- Human Rights Watch. (1999). Red Onion State Prison: Super-maximum security confinement in Virginia. *Human Rights Watch*, 11, 1-24.
- Human Rights Watch. (2000). Out of sight: Super-maximum security confinement in the United States. *Human Rights Watch*, 12, 1-9.
- James, D., & Glaze, L. (2006). *Mental health problems of prisons and jail inmates*. Washington, DC: Bureau of Justice Statistics.
- Jones v. Baker, 115 F.3d. 810 (6th Cir. 1998).
- Jones 'El v. Berge, 164 F. Supp. 2d 1096 (W.D. Wis. 2001).
- King, R. D. (1999). The rise and rise of supermax: An American solution in search of a problem? *Punishment and Society*, 1, 163-186. doi: 10.1177/14624749922227766
- Kupers, T. (2008). What to do with the survivors? Coping with the long-term effects of isolated confinement. *Criminal Justice and Behavior*, 35, 1005-1016.
- Kurki, L., & Morris, N. (2001). The purposes, practices, and problems of supermax prisons. *Crime and Justice*, 28, 1-21.
- Lovell, D. (2008). Patterns of disturbed behavior in a supermax population. *Criminal Justice and Behavior*, 35, 985-1004.
- Lovell, D., Johnson, L. C., & Cain, K. C. (2007). Recidivism of supermax prisoners in Washington state. *Crime and Delinquency*, 53, 633-656. doi: 10.1177/0011128706296466
- Lukoff, D., Nuechterlein, K., & Ventura, A. (1986). Manual for the expanded brief psychiatric rating scale. *Schizophrenia Bulletin*, 13, 261-276.
- Madrid v. Gomez, 899 F. Supp. 1146 (N.D. Cal. 1995).
- McClary v. Kelly, 4 F. Supp. 2d 195 (W.D.N.Y. 1998).
- McClellan, D. S., Farabee, D., & Crouch, B. M. (1997). Early victimization, drug use, and criminality: A comparison of male and female prisoners. *Criminal Justice and Behavior*, 24, 455-476.
- McNair, D. M., Lorr, M., & Droppleman, L. F. (1992). *Manual: Profile of Mood States, Revised*. San Diego, CA: Educational and Industrial Testing Service.
- Mears, D. P. (2008). An assessment of supermax prisons using an evaluation research framework. *The Prison Journal*, 88, 43-68.
- Mears, D. P., & Bales, W. D. (2009). Supermax incarceration and recidivism. *Criminology*, 47, 1131-1166. doi: 10.1111/j.1745-9125.2009.00171.x
- Mears, D. P., & Watson, J. (2006). Towards a fair and balanced assessment of supermax prisons. *Justice Quarterly*, 23, 232-270. doi: 10.1080/07418820600688867

- Metzner, J. L., & Dvoskin, J. A. (2006). An overview of correctional psychiatry. *Psychiatric Clinics of North America*, 29, 761-772. Retrieved from [http://psych.theclinics.com/issues/contents?issue_key=S0193-953X\(06\)X0019-9](http://psych.theclinics.com/issues/contents?issue_key=S0193-953X(06)X0019-9)
- Metzner, J. L., & Fellner, J. (2010). Solitary confinement and mental illness in U.S. prisons: A challenge for medical ethics. *Journal of the American Academy of Psychiatry and the Law*, 38, 104-108. Retrieved from <http://www.jaapl.org/cgi/reprint/38/1/104.pdf>
- Morey, L. C. (1997). *Personality Assessment Screener: Professional manual*. Lutz, FL: Psychological Assessment Resources, Inc.
- Motiuk, L. L., & Blanchette, K. (2001). Characteristics of administratively segregated offenders in federal corrections. *Canadian Journal of Criminology*, 43, 131-144.
- Motiuk, M. S., Motiuk, L. L., & Bonta, J. (1992). A comparison between self-report and interview-based inventories in offender classification. *Criminal Justice and Behavior*, 19, 143-159. doi: 10.1177/0093854892019002003
- Naday, A., Freilich, J. D., & Mellow, J. (2008). The elusive data on supermax confinement. *The Prison Journal*, 88, 69-93.
- National Institute of Corrections. (1997). *Supermax housing: A survey of current practice, special issues in corrections*. Longmont, CO: U.S. Department of Justice, National Institute of Corrections.
- National Institute of Corrections. (1999). *Supermax prisons: Overview and general considerations*. Longmont, CO: US Department of Justice, National Institute of Corrections.
- National Institute of Mental Health. (2010). The Numbers Count: Mental Disorders in America. Retrieved from <http://www.nimh.nih.gov/health/publications/the-numbers-count-mental-disorders-in-america/index.shtml>
- O'Keefe, M. L. (2005). *Analysis of Colorado's administrative segregation*. [Technical Report]. Colorado Springs, CO: Department of Corrections.
- O'Keefe, M. L. (2008a). Administrative segregation for mentally ill inmates. *Journal of Offender Rehabilitation*, 45, 149-165.
- O'Keefe, M. L. (2008b). Administrative segregation from within: A corrections perspective. *The Prison Journal*, 88, 123-143.
- O'Keefe, M. L., Klebe, K., & Hromas, S. (1998). *Validation of the Level of Supervision Inventory (LSI) for community based offenders in Colorado: Phase II*. [Technical Report]. Colorado Springs, CO: Colorado Department of Corrections.
- O'Keefe, M. L., & Schnell, M. (2008). Offenders with Mental Illness in the Correctional System. *Journal of Offender Rehabilitation*, 45, 81-104.
- Overall, J. E. & Gorham, D. R. (1962). The brief psychiatric rating scale. *Psychological Reports*, 10, 799-812.

- Pizarro, J. M., & Narag, R. E. (2008). Supermax prisons: What we know, what we do not know, and where we are going. *The Prison Journal, 88*, 23-42.
- Pizzaro, J. & Stenius, V. M. K. (2004). Supermax prisons: Their rise, current practices and effect on inmates. *The Prison Journal, 84*, 248-264.
- Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills, 8*, 271-276.
- Rhodes, L. (2004). *Total confinement: Madness and reason in maximum security*. Berkeley, CA: University of California Press.
- Ruiz v. Johnson, 37 F. Supp. 855 (S.D. Texas 1999).
- Sandin v. Conner, 515 U.S. 473 (1995).
- Severance, T. A. (2005). You know who you can go to: Cooperation and exchange between incarcerated women. *The Prison Journal, 85*, 343-367.
- Smith, P. S. (2008). "Degenerate criminals:" Mental health and psychiatric studies of Danish prisoners in solitary confinement, 1870-1920. *Criminal Justice and Behavior, 35*, 1048-1064. doi: 10.1177/0093854808318782
- Spielberger, C., Gorsuch, R., & Lushene, R. (1970). *The State-Trait Anxiety Inventory: Test manual for Form X*. Palo Alto, CA: Consulting Psychologists Press.
- Suedfeld, P., Ramirez, C., Deaton, J., & Baker-Brown, G. (1982). Reactions and attributes of prisoners in solitary confinement. *Criminal Justice and Behavior, 9*, 303-340. doi: 10.1177/0093854882009003004
- Sundt, J. L., Castellano, T. D., & Briggs, C. S. (2008). The sociopolitical context of prison violence and its control: A case study of supermax and its effect in Illinois. *The Prison Journal, 88*, 94-122.
- Tariq, S. H., Tumosa, N., Chibnall, J. T., Perry, M. H., & Morley, J. E. (2006). Comparison of the Saint Louis University Mental Status Examination and the Mini-Mental State Examination for detecting dementia and mild neurocognitive disorder—a pilot study. *American Journal of Geriatric Psychiatry, 14*, 900-910.
- Toch, H. (1992). *Living in prison: The ecology of survival*. Washington, DC: American Psychological Association.
- Toch, H. (2001). The future of supermax confinement. *The Prison Journal, 81*, 376-388.
- Turner v. Safley, 482 U.S. 78 (1987).
- Ventura, J., Lukoff, D., Nuechterlein, K. H., Liberman, R. P., Green, M. F., & Shaner, A. (1993). Brief Psychiatric Rating Scale (BPRS) expanded version (4.0): Scales, anchor points, and administration manual. *International Journal of Methods in Psychiatric Research, 3*, 227-244.
- Ward, D. A., & Werlich, T. G. (2003). Alcatraz and Marion: Evaluating super-maximum custody. *Punishment and Society, 5*, 53-75. doi: 10.1177/1462474503005001295

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Widows, M. R., & Smith, G. P. (2005). *Structured Inventory of Malingered Symptomatology: Professional manual*. Lutz, FL: Psychological Assessment Resources, Inc.

Wilkinson v. Austin, 545 U.S. 209 (2005).

Zinger, I., Wichman, C., & Andrews, D. A. (2001). The psychological effects of 60 days in administrative segregation. *Canadian Journal of Criminology*, 43, 47-88.

Zlotnik, C., & Pearlstein, T. (1997). Posttraumatic stress disorder (PTSD) comorbidity, and childhood abuse among incarcerated women. *Journal of Nervous and Mental Disease*, 185, 761-763.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

APPENDIX A

PSI

Please rate how often the following items have applied to you in the past week.

	Never True	Rarely True	Sometimes True	Often True	Usually True	Always True	
1. I find that even quiet noises are loud and disturbing.	0	1	2	3	4	5	
2. My heart races faster than normal at times.	0	1	2	3	4	5	
3. I am afraid for no reason.	0	1	2	3	4	5	
4. I might go a day without brushing my teeth.	0	1	2	3	4	5	
5. I do cardiovascular activity (jogging, running, speed walking, etc.).	0	1	2	3	4	5	
6. I have difficulty catching my breath even when I am not exercising.	0	1	2	3	4	5	
7. There are smells here that make me queasy.	0	1	2	3	4	5	
8. I have pounding headaches that make it hard to concentrate.	0	1	2	3	4	5	
9. I comb or brush my hair daily.	0	1	2	3	4	5	
10. I struggle to get air.	0	1	2	3	4	5	
11. I have a lot of energy.	0	1	2	3	4	5	
12. My fear prevents me from doing things that I'd like to do.	0	1	2	3	4	5	
13. I feel dizzy at times.	0	1	2	3	4	5	
14. I look forward to getting back to the general population.	0	1	2	3	4	5	N/A
15. I do strength training (weight lifting, pull-ups, push-ups, etc.).	0	1	2	3	4	5	
16. I feel as though I am choking.	0	1	2	3	4	5	
17. I feel lightheaded or like I am going to faint.	0	1	2	3	4	5	
18. I avoid shaving or grooming my facial hair.	0	1	2	3	4	5	
19. I sleep most of the day.	0	1	2	3	4	5	
20. I find my whole body trembling for no apparent reason.	0	1	2	3	4	5	

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

	Never True	Rarely True	Sometimes True	Often True	Usually True	Always True	
21. I have episodes where I am certain I will die soon.	0	1	2	3	4	5	
22. I do not speak to anyone even when they talk to me.	0	1	2	3	4	5	
23. I shower every day that I am allowed.	0	1	2	3	4	5	
24. I am troubled by physical pain or aches.	0	1	2	3	4	5	
25. I cannot stop myself from shaking.	0	1	2	3	4	5	
26. I am bored to death.	0	1	2	3	4	5	
27. Exercise is not important to me.	0	1	2	3	4	5	
28. I sleep soundly at night.	0	1	2	3	4	5	
29. It is important to me to keep good hygiene.	0	1	2	3	4	5	
30. I break out in a sweat when I am not doing anything.	0	1	2	3	4	5	
31. I find the quiet to be peaceful.	0	1	2	3	4	5	
32. I start conversations with other people.	0	1	2	3	4	5	
33. It is unsafe for me in the general population.	0	1	2	3	4	5	
34. My cell temperature is comfortable.	0	1	2	3	4	5	
35. I feel calm and relaxed.	0	1	2	3	4	5	
36. I need a single cell for my own protection.	0	1	2	3	4	5	
37. This place makes me feel miserable.	0	1	2	3	4	5	
38. I am not bothered by thoughts of dying.	0	1	2	3	4	5	
39. I prefer administrative segregation to the general population.	0	1	2	3	4	5	N/A

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions of points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

PRISON BEHAVIOR RATING SCALE

Instructions: Based on your interactions and records such as Chronological Notes, disciplinary infractions or other incidents, please rate the inmate’s behavior by circling your answer.

	Never/ Rarely	Sometimes	Often	Most of the Time
1. Tried but failed to follow instructions	0	1	2	3
2. Appeared tense and unable to relax	0	1	2	3
3. Appeared close to tears	0	1	2	3
4. Caused trouble during his free time	0	1	2	3
5. Cursed and swore (in an abusive manner)	0	1	2	3
6. Appeared easily upset	0	1	2	3
7. Appeared sluggish and drowsy	0	1	2	3
8. Been held out of normal circulation (e.g. dry cell, mental health watch, special controls, RFP, punitive segregation etc.)	0	1	2	3
9. Had trouble sleeping at night	0	1	2	3
10. Appeared lacking in energy	0	1	2	3
11. Sought reassurance	0	1	2	3
12. Appeared to be brooding on something	0	1	2	3
13. Victimized weaker inmates	0	1	2	3
14. Appeared dull and unintelligent	0	1	2	3
15. Fidgeted and been unable to sit still	0	1	2	3
16. Tried to con staff	0	1	2	3
17. Appeared frightened of other inmates	0	1	2	3
18. Complained about staff	0	1	2	3
19. Not been aware of what is going on around him	0	1	2	3
20. Been aggressive towards staff	0	1	2	3

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

	Never/ Rarely	Sometimes	Often	Most of the Time
21. Had a quick temper	0	1	2	3
22. Been on report (e.g., got a negative CHRON, written up for COPD violation)	0	1	2	3
23. Appeared preoccupied/dreamy	0	1	2	3
24. Tried to play staff against each other	0	1	2	3
25. Openly defied rules	0	1	2	3
26. Appeared sad and depressed	0	1	2	3
27. Stirred up trouble among other inmates	0	1	2	3
28. Aided or abetted others to break the rules	0	1	2	3
29. Been out of touch with what is happening around him	0	1	2	3
30. Been victimized by other inmates	0	1	2	3
31. Not understood orders	0	1	2	3
32. Appeared to be scared	0	1	2	3
33. Has few if any friends	0	1	2	3
34. Avoided other inmates	0	1	2	3
35. Given the impression of ignorance/inability	0	1	2	3
36. Appeared depressed, gloomy, or sulky	0	1	2	3
37. Had poor hygiene	0	1	2	3

Completed by _____

Please Print Name

APPENDIX B

SUMMARY OF STUDY MEASURES

Most of the measures included in this study were self-report pencil-and-paper tests; however, we also collected data from clinicians who completed the Brief Psychiatric Rating Scale (BPRS) and correctional staff who completed the Prison Behavior Rating Scale (PBRs). Two additional measures assessing cognitive functioning (i.e., St. Louis University Memory Scale, Trail Making Test) were administered by a researcher. Instruments used in this study were selected to assess a broad range of symptoms believed to be associated with long-term segregation. We assessed eight constructs by means of 10 different measures (and/or appropriate subscales). The constructs of interest in this study were anxiety, cognitive impairment, depression-hopelessness, hostility-anger control, hypersensitivity, psychosis, somatization, and withdrawal-alienation. In addition to these key variables, we measured other variables that may be predictors of outcomes, including trauma, personality disorders, malingering, and history of self-harm.

Measures were selected for ease of administration and strength of psychometric properties. In this appendix, we describe the measures used in this study, provide results concerning the psychometric properties of the measures, and describe the composites used for analyses in the report. Data are reported for the entire sample at each time period. Table B1 provides a quick reference guide to the tests used in this study as well as the constructs assessed by each of them.

Table B1. Study Measures

Measure	Construct	Administration	Times Assessed
Outcome Variables			
Beck Hopelessness Scale (BHS)	Depression-Hopelessness	Self-Report	B, every 3 months
Brief Psychiatric Rating Scale (BPRS)		Clinicians	B, every 6 months
Activity			
Anxious-Depressed	Anxiety, Depression-Hopelessness		
Hostility-Suspiciousness	Hostility-Anger Control		
Thought Disorder	Psychosis		
Withdrawal	Withdrawal-Alienation		
Brief Symptom Inventory (BSI)		Self-Report	B, every 3 months
Anxiety	Anxiety		
Depression	Depression-Hopelessness		
Hostility	Hostility-Anger Control		
Interpersonal Sensitivity	Hypersensitivity		
Obsessive-Compulsive	Anxiety		
Paranoid Ideation	Psychosis		
Phobic Anxiety	Anxiety		
Psychoticism	Psychosis		
Somatization	Somatization		
Personality Assessment Screener (PAS)		Self-Report	B, every 3 months
Acting Out	Hostility-Anger Control		
Alienation	Withdrawal-Alienation		
Anger Control	Hostility-Anger Control		
Health Problems	Somatization		
Hostile Control	Hostility-Anger Control		
Negative Affect	Anxiety, Depression-Hopelessness		
Psychotic Features	Psychosis		
Social Withdrawal	Withdrawal-Alienation		
Suicidal Thinking	Depression-Hopelessness		

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Measure	Construct	Administration	Times Assessed
Prison Behavior Rating Scale (PBRs)		Officers	B, every 3 months
Anti-Authority	Hostility-Anger Control		
Anxious-Depressed	Anxiety, Depression-Hopelessness		
Dull-Confused	Cognitive Impairment		
Prison Symptom Inventory (PSI)		Self-Report	B, every 3 months
Panic Disorder	Anxiety		
Hypersensitivity-External Stimuli	Hypersensitivity		
Physical Symptoms	Somatization		
Profile of Mood States (POMS)		Self-Report	B, every 3 months
Anger-Hostility	Hostility-Anger Control		
Depression-Dejection	Depression-Hopelessness		
Fatigue-Inertia	Somatization		
Tension-Anxiety	Anxiety		
Saint Louis University Memory Scale (SLUMS)	Cognitive Impairment	Researcher	B, every 3 months
State-Trait Anxiety Inventory (STAI)		Self-Report	B, every 3 months
State Anxiety	Anxiety		B, every 3 months
Trait Anxiety	Anxiety		B, every 3 months
Trail Making Test (TMT)		Researcher	B, every 3 months
Time to Complete A Task	Cognitive Impairment		
Time to Complete B Task	Cognitive Impairment		
B - A Time	Cognitive Impairment		
B/A Time	Cognitive Impairment		
Predictor Variables			
Coolidge Correctional Inventory (CCI)	Personality Disorders	Yes	B
Deliberate Self-harm Inventory (DSHI)	Self-Harm	Yes	B
Structured Inventory of Malingered Symptoms (SIMS)	Malingering	Yes	B, every 3 months
Trauma Symptom Inventory (TSI)	Trauma		2 nd

Note. Times assessed include the first time the test was administered as well as the interval at which it was given (unless it was only conducted at specific testing periods). B stands for baseline test.

DESCRIPTION OF INDIVIDUAL MEASURES

In this section, descriptions of the measures and summary statistics about reliability and validity are provided. Summary statistics include those published in the literature as well as those conducted with our study population. Cronbach's alpha is used to estimate internal consistency reliability at each time period for the entire sample. Correlations between consecutive time periods are used to estimate test-retest reliability. Convergent validity is estimated by correlations of each measure with other measures of the same construct for the entire sample at each time period. Tables for descriptive statistics on the measures are given with the description of the measure; reliability and validity statistics are presented in the description of the measure and/or with the description of the composites.

Beck Hopelessness Scale (BHS)

Designed to measure an individual's degree of despair/depression, the BHS (Beck & Steer, 1993; Beck, Weissman, Lester, & Trexler, 1974) is a 20-item self-report measure on which scores can range from 0 to 20, with higher scores indicating a greater degree of despair about the future (Clum & Yang, 1995). Since this measure does not have any subscales, one total score is derived. Respondents answer *true* or *false* to

statements about their attitudes over the past week (Beck & Steer, 1993). It takes approximately 5 to 10 minutes to complete this measure (Beck & Steer, 1993).

The psychometric properties of the BHS are solid, as it has demonstrated internal consistency estimates of .65 to .89 for nonclinical samples of college students (e.g., Beck et al., 1974; Durham, 1982; Steed, 2001), .86 to .93 for clinical in- or out-patient samples (e.g., Beck et al., 1974; Durham, 1982; Dyce, 1996), and .83 in a forensic sample (Durham, 1982). Three-week test-retest reliability in a university sample was found to be acceptable ($r = .85$) for the entire sample and was slightly higher for males ($r = .94$; Holden & Fekken, 1988). In clinical samples, test-retest correlations ranged from .66 (six-week test-retest correlation) to .69 (one-week test-retest correlation; Beck & Steer, 1993).

BHS self-report ratings have also been correlated to clinician ratings of hopelessness, with correlations ranging from .78 to .98 (Beck et al., 1974), which suggests that this measure possesses acceptable convergent validity. Additionally, the BHS is considered to be a predictor of suicide risk in clinical populations (e.g., Beck, 1986; Beck, Brown, Berchick, Stewart, & Steer, 1990; Beck et al., 1974; Brown, Beck, Steer, & Grisham, 2000), with scores of 9 and above being predictive of suicidal ideation (Beck, Steer, Kovacs, & Garrison, 1985). Additionally, the correlation between the BHS and the Modified Scale for Suicide Ideation was found to be moderate at .46 (Clum & Yang, 1995) in a nonclinical sample of college students. Further evidence for the BHS's convergent validity comes from a study on a clinical inpatient sample; the BHS was found to be significantly correlated with the Beck Depression Inventory (BDI; $r = .68$) and the Current Suicidal Intent ($r = .68$; Kovacs, Beck, & Weissman, 1975).

Summary statistics are available for several different groups, including criminal psychiatric inpatients (Durham, 1982). Durham (1982) assessed college students, general psychiatric patients, and forensic psychiatric patients on the BHS. He found that the mean for the nonclinical, college student sample was 2.32 ($SD = 2.25$, $n = 197$), 6.04 ($SD = 4.67$, $n = 118$) for the clinical, general psychiatric sample, and 6.62 ($SD = 4.88$, $n = 99$) for the clinical, forensic sample. In another study, including 2,067 psychiatric outpatients, the mean total BHS score was 9.06 ($SD = 5.61$; Bieling, Beck, & Brown, 2000). Palmer and Connelly (2005) assessed BHS scores for prisoners with ($n = 24$) and without ($n = 24$) a history of self-harming behavior. They found that the mean BHS score for prisoners with a history of self-harming behavior was 10.13 ($SD = 4.81$); for offenders without a history of self-harming behavior the BHS mean score was found to be significantly lower at 6.29 ($SD = 4.49$).

Summary statistics on the BHS for the current study are given in Table B2 for each group. Internal consistency estimates indicated excellent consistency across items at each time (mean Cronbach's alpha = .93; range = .92 to .94). Test-retest correlation coefficients ranged between .66 and .79 ($M = .71$). Examination of the validity coefficients, given in the composite section of this appendix, indicated that the BHS is correlated with other self-report measures of depression (mean $r = .58$, range = .43 to .77); however, the correlations with relevant staff reports (BPRS Anxious-Depressed and PPRS Anxious-Depressed) were lower (mean $r = .18$, range = -.02 to .33).

Table B2. Summary Statistics (M, SD, n) on BHS by Group and Time

Time	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
1	7.16 (5.63) n = 63	5.14 (4.52) n = 63	7.09 (5.38) n = 33	2.26 (3.16) n = 43	9.84 (6.28) n = 67	6.56 (5.75) n = 269
2	7.59 (6.62) n = 62	4.66 (4.96) n = 59	5.56 (4.68) n = 32	2.35 (3.06) n = 41	9.20 (6.89) n = 64	6.24 (6.11) n = 258
3	7.75 (6.37) n = 60	4.79 (4.88) n = 57	5.38 (4.58) n = 32	2.56 (3.92) n = 41	8.57 (6.85) n = 61	6.12 (5.99) n = 251
4	8.12 (6.90) n = 60	3.87 (4.11) n = 56	5.62 (4.91) n = 29	2.60 (2.89) n = 39	9.20 (7.05) n = 59	6.22 (6.16) n = 243
5	6.32 (6.31) n = 56	4.71 (5.03) n = 56	4.07 (4.22) n = 29	1.58 (2.06) n = 38	8.93 (6.88) n = 57	5.55 (5.92) n = 236
6	5.86 (5.57) n = 51	3.51 (4.10) n = 54	NA	NA	NA	4.61 (4.98) n = 106

Brief Psychiatric Rating Scale (BPRS)

The expanded Brief Psychiatric Rating Scale (BPRS-E; Ventura et al., 1993) is a 24-item measure administered by clinicians to assess patients with psychiatric disorders. It is designed to allow for the rapid review of psychological symptoms over time (e.g., Ventura et al., 1993). Ratings are made after a semi-structured clinical interview with a client. Clinicians rate the different items on the BPRS-E by means of a 7-point severity scale (1- *not present* to 7- *extremely severe*); higher scores on this measure generally indicate greater severity of psychopathology (Segal & Silverman, 2002; Thomas, Donnell, & Young, 2004). However, since clinicians also have an option of using a not assessed (N/A) rating on any given item, scores may not accurately reflect the degree of psychopathology (Ventura et al., 1993). The clinical interview takes approximately 10 to 40 minutes, depending on familiarity with the client as well as presenting symptoms at the time of the assessment (Thomas et al., 2004). Research has indicated that there are five factors to which the individual items of the BPRS-E are associated: thought disorder (directly reflecting psychosis), withdrawal, anxious-depressed, hostility-suspiciousness, and activity (Burger, Calsyn, Morse, Klinkenberg, & Trusty, 1997).

Internal consistency reliability for the total BPRS-E was found to be between .74 and .79 for clinical populations (Perlick, Rosenheck, Clarkin, Sirey, & Raue, 1999; Segal & Silverman, 2002; Thomas et al., 2004). Furthermore, when considering the internal consistency reliability for the 5-factor structure, the coefficients for four out of the five scales ranged from .73 (i.e., anxiety-depression) to .81 (i.e., activity); the Cronbach's alpha for the hostility-suspiciousness factor was found to be lower at .49 (Burger et al., 1997).

Mean total scores for clinical populations were found to be between 37.9 ($SD = 11.1$) and 61.6 ($SD = 12.9$; Biancosino et al., 2004; Brown, Chhina, & Dye, 2008; Segal & Silverman, 2002), while the mean total for the BPRS-E among inmates with psychiatric problems in the prison population was found to be 49.29 ($SD = 14.78$; Gray, Bressington, Lathlean, & Mills, 2008). When individuals were tested over time in a clinical setting, mean scores significantly decreased at each testing interval (Biancosino et al., 2004; Brown et al., 2008), which indicates that the test may be sensitive to change over time following an intervention.

The BPRS-E has been shown to be correlated with the Brief Symptom Inventory (BSI), a self-report measure of psychological symptoms (Morlan & Tan, 1998), indicating convergent validity. Furthermore, a study on female inmates showed that mental health referrals are more often done for those inmates who have received higher BPRS-E scores, suggesting that this assessment tool is useful in detecting psychopathology in inmates (Nicholls, Lee, Corrado, & Ogloff, 2004).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table B3 provides the summary statistics for the study groups on the BPRS scales at each time period. The BPRS scores tend to be lower than normative data found with other clinical populations, indicating a potential floor effect and potential rater bias.

Table B3. Summary Statistics (*M*, *SD*, *n*) on BPRS Scales by Group at each 6 month Time Period

Assessment	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Activity						
Time 1	6.30 (1.74) <i>n</i> = 59	5.56 (1.05) <i>n</i> = 57	6.21 (1.54) <i>n</i> = 33	5.82 (1.34) <i>n</i> = 39	6.82 (2.58) <i>n</i> = 62	6.18 (1.84) <i>n</i> = 250
Time 3	5.92 (1.49) <i>n</i> = 54	5.26 (.67) <i>n</i> = 47	5.90 (1.27) <i>n</i> = 30	5.42 (1.18) <i>n</i> = 38	6.25 (1.44) <i>n</i> = 56	5.78 (1.31) <i>n</i> = 225
Time 5	6.17 (2.36) <i>n</i> = 48	5.18 (.49) <i>n</i> = 45	5.96 (1.56) <i>n</i> = 27	5.28 (.53) <i>n</i> = 29	6.27 (2.05) <i>n</i> = 51	5.82 (1.73) <i>n</i> = 200
Anxious-Depressed						
Time 1	9.51 (3.03) <i>n</i> = 59	7.37 (2.22) <i>n</i> = 57	9.03 (2.90) <i>n</i> = 33	6.67 (1.81) <i>n</i> = 39	10.40 (3.34) <i>n</i> = 62	8.74 (3.08) <i>n</i> = 250
Time 3	9.07 (2.95) <i>n</i> = 54	6.62 (2.15) <i>n</i> = 47	8.20 (2.44) <i>n</i> = 30	6.50 (1.84) <i>n</i> = 38	8.91 (2.64) <i>n</i> = 56	7.97 (2.71) <i>n</i> = 225
Time 5	8.44 (2.85) <i>n</i> = 48	7.04 (2.95) <i>n</i> = 45	8.22 (2.31) <i>n</i> = 27	5.97 (1.68) <i>n</i> = 29	9.06 (3.13) <i>n</i> = 51	7.90 (2.92) <i>n</i> = 200
Hostility-Suspiciousness						
Time 1	5.52 (2.46) <i>n</i> = 59	4.04 (1.76) <i>n</i> = 57	4.70 (1.72) <i>n</i> = 33	3.90 (1.83) <i>n</i> = 39	5.47 (3.01) <i>n</i> = 62	4.81 (2.39) <i>n</i> = 250
Time 3	5.17 (2.45) <i>n</i> = 54	3.34 (.64) <i>n</i> = 47	4.43 (2.16) <i>n</i> = 30	3.42 (1.26) <i>n</i> = 38	4.38 (1.54) <i>n</i> = 56	4.20 (1.86) <i>n</i> = 225
Time 5	4.54 (2.16) <i>n</i> = 48	3.36 (.71) <i>n</i> = 45	4.37 (1.86) <i>n</i> = 27	3.76 (1.62) <i>n</i> = 29	4.72 (2.17) <i>n</i> = 51	4.19 (1.88) <i>n</i> = 200
Thought Disorder						
Time 1	6.59 (2.35) <i>n</i> = 59	5.32 (.87) <i>n</i> = 57	5.64 (.99) <i>n</i> = 33	5.18 (.51) <i>n</i> = 39	8.29 (3.33) <i>n</i> = 62	6.38 (2.41) <i>n</i> = 250
Time 3	6.50 (1.87) <i>n</i> = 54	5.17 (.48) <i>n</i> = 47	5.33 (.84) <i>n</i> = 30	5.10 (.39) <i>n</i> = 38	6.59 (1.94) <i>n</i> = 56	5.85 (1.54) <i>n</i> = 225
Time 5	6.35 (2.45) <i>n</i> = 48	5.18 (.49) <i>n</i> = 45	5.44 (1.22) <i>n</i> = 27	5.38 (1.35) <i>n</i> = 29	6.61 (2.11) <i>n</i> = 51	5.89 (1.84) <i>n</i> = 200
Withdrawal						
Time 1	7.73 (1.76) <i>n</i> = 59	6.79 (1.18) <i>n</i> = 57	7.00 (1.41) <i>n</i> = 33	6.38 (.63) <i>n</i> = 39	8.61 (2.60) <i>n</i> = 62	7.43 (1.92) <i>n</i> = 250
Time 3	7.83 (1.96) <i>n</i> = 54	7.06 (1.40) <i>n</i> = 47	6.83 (1.26) <i>n</i> = 30	6.21 (.53) <i>n</i> = 38	7.59 (1.56) <i>n</i> = 56	7.20 (1.58) <i>n</i> = 225
Time 5	7.50 (1.62) <i>n</i> = 48	6.71 (1.74) <i>n</i> = 45	7.22 (1.42) <i>n</i> = 27	6.31 (.76) <i>n</i> = 29	7.59 (1.55) <i>n</i> = 51	7.14 (1.57) <i>n</i> = 200
Total						
Time 1	35.66 (7.60) <i>n</i> = 59	29.07 (4.71) <i>n</i> = 57	32.58 (5.38) <i>n</i> = 33	27.95 (4.88) <i>n</i> = 39	39.60 (9.69) <i>n</i> = 62	33.52 (8.28) <i>n</i> = 250
Time 3	34.50 (7.64) <i>n</i> = 54	27.45 (3.51) <i>n</i> = 47	30.70 (4.76) <i>n</i> = 30	26.66 (3.77) <i>n</i> = 38	33.71 (4.67) <i>n</i> = 56	31.00 (6.13) <i>n</i> = 225
Time 5	33.00 (8.56) <i>n</i> = 48	27.47 (4.59) <i>n</i> = 45	31.22 (4.20) <i>n</i> = 27	26.69 (3.57) <i>n</i> = 29	34.25 (7.12) <i>n</i> = 51	30.92 (6.93) <i>n</i> = 200

Table B4 provides the Cronbach's alpha estimates for the subscales at each assessment period. The internal consistency estimates for the BPRS subscales (*M* = .55, range = .40 to .66) were lower than those found in normative samples but similar across time periods; however, the internal consistency estimates for the total score is similar to that found in normative samples. Correlations between sequential time periods (6 months

apart) are provided in Table B5 and show low stability across time. It is possible that participants changed facilities from one testing session to the next, causing a switch in their assigned clinicians. This change in clinicians could lower correlations between BPRS scores across time and present a picture of inmates' psychopathological instability when, in fact, inter-rater disparity might be causing the change in BPRS scores over time. Correlations between the BPRS scales and relevant self-report scales of the same construct ranged between .15 and .35 ($M = .28$). Correlations between the BPRS scales and the relevant correctional officer ratings (PBRS scales) ranged between .08 and .29 ($M = .19$). These convergent validity estimates are lower than expected and are likely impacted by restriction of range (i.e., scores on the BPRS are averaging at the low end of the possible scores and standard deviations are small).

Table B4. Internal Consistency Estimates (Cronbach's alpha) for BPRS Scales at each Time Period

BPRS Scale	Time 1	Time 3	Time 5
Activity	.58	.53	.64
Anxious-Depressed	.55	.60	.66
Hostility-Suspiciousness	.57	.61	.51
Thought Disorder	.64	.52	.57
Withdrawal	.47	.49	.40
Total Scale	.81	.80	.79

Table B5. Correlations between Consecutive Time Periods for BPRS Scales

BPRS Scale	T1-T3	T3-T5
Activity	.36	.40
Anxious-Depressed	.45	.43
Hostility-Suspiciousness	.36	.48
Thought Disorder	.33	.58
Withdrawal	.30	.23
Total Scale	.41	.51

Brief Symptom Inventory (BSI)

The BSI (Derogatis, 1993) is a 53-item self-report measure that is widely employed to assess a broad range of psychological symptoms. It measures clinical symptoms across nine subscales (i.e., Somatization, Obsessive-Compulsive, Interpersonal Sensitivity, Depression, Anxiety, Hostility, Phobic Anxiety, Paranoid Ideation, and Psychoticism) and three global scales (i.e., General Severity Index [GSI]; Positive Symptom Total; Positive Symptom Distress Index; Boulet & Boss, 1991). Respondents are asked to rate the degree of distress experienced over the last week, using a 5-point rating scale (0 – *not at all* to 4 – *extremely*). Higher scores on the BSI indicate a greater degree of psychopathology. Despite having different subscales, the BSI seems to be better at providing information on the general degree of psychopathology instead of the nature of it (Boulet & Boss, 1991). A minimum of 6th grade reading ability is required to complete this measure, and it generally takes 10 minutes to complete (Boulet & Boss, 1991).

Internal consistency reliabilities across subscales are acceptable for clinical populations (range = .57 to .89; Boulet & Boss, 1991; Broday & Mason, 1991; Hayes, 1997; Kellett, Beail, Newman, & Frankish, 2003). Additionally, internal consistency reliabilities for nonclinical, community samples for the different subscales ranged from .60 to .81 (Kellett et al., 2003), whereas they ranged from .52 to .86 for forensic populations (Kellett et al., 2003; Zinger, Wichmann, & Andrews, 2001). Item-total correlations for the scales ranged from

.57 to .79, with a median correlation of .69, for clinical populations (Hayes, 1997) and ranged from .73 to .91 for forensic populations (Boulet & Boss, 1991). Two-week test-retest reliability is acceptable for the subscales (range = .68 [Somatization] to .91 [i.e., Phobic Anxiety]) and the GSI ($r = .90$; Cundick, 1975; Derogatis, 1993; Kellett et al., 2003; Piersma, Reaume, Boes, 1994) across nonclinical, clinical, and forensic samples. The BSI has been shown to be valid for studying change over time (Long, Harring, Brekke, Test, & Greenberg, 2007).

Normative data are widely available for psychiatric in- and out-patients and the general population but not for a prison population (Derogatis, 1993). Normative means for the different subscales ranged from .67 ($SD = .71$) to 1.65 ($SD = 1.11$) in psychiatric outpatients, from .71 ($SD = .97$) to 1.26 ($SD = 1.15$) in psychiatric in-patients, and from .11 ($SD = .25$) to .37 ($SD = .41$) in nonclinical populations (Derogatis, 1993). Cochran and Hale (1985) conducted a normative study on male and female college students at a 4-year college. They found that mean scores ranged from .29 ($SD = .27$) to 1.17 ($SD = .77$) for males ($n = 143$) and from .32 ($SD = .45$) to 1.12 ($SD = .66$) for females ($n = 204$). Furthermore, normative data are available on the global scales of the BSI; the normative mean for the BSI GSI was 1.20 ($SD = .70$) for psychiatric outpatients, .25 ($SD = .24$) for nonclinical populations, and .97 ($SD = .78$) for psychiatric inpatients (Derogatis, 1993).

Convergent validity has been assessed by means of comparing dimensions of the BSI to clusters on the Minnesota Multiphasic Personality Inventory (MMPI; Boulet & Boss, 1991; Cundick, 1975). Correlations of these comparisons were between .30 and .72 in Cundick's (1975) study. In Boulet and Boss's (1991) study, correlations between the most relevant MMPI and BSI subscales were found to be moderate, ranging from .50 (MMPI Depression and BSI Depression) to .53 (MMPI Hypochondriasis and BSI Somatization). In a clinical sample, some of the BSI subscales were significantly correlated with the associated subscales on the BPRS: the depression scale on the BSI was significantly correlated to the depressive mood scale on the BPRS ($r = .69$), the anxiety scales on the BSI and BPRS correlated as well ($r = .49$), and the two hostility scales of both measures were also significantly correlated with one another ($r = .49$; Morlan & Tan, 1998). Overall, moderate to high correlations with other measures seem to indicate that the BSI does, indeed, have adequate convergent validity.

Table B6 provides the summary statistics for the present study on the BSI scales at each time period. Internal consistency estimates for the BSI subscales were strong with Cronbach's alphas ranging between .71 and .91 ($M = .85$). Test-retest reliability estimates ranged between .53 and .79 ($M = .72$) indicating good stability within three month testing intervals. The BSI subscales showed reasonable convergent validity as correlations with other self-report measures of the same constructs ranged between .15 and .89 ($M = .56$) but there were lower validity estimates with staff reports with correlations ranging between -.01 and .43 ($M = .23$).

Table B6. Summary Statistics (M, SD, n) on BSI Scales by Group and Time

Assessment	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Anxiety						
1	7.87 (6.52) $n = 64$	3.43 (4.00) $n = 63$	7.30 (5.56) $n = 33$	2.30 (3.15) $n = 43$	9.69 (5.75) $n = 67$	6.33 (5.92) $n = 270$
2	7.21 (6.01) $n = 62$	2.24 (3.71) $n = 58$	5.56 (5.58) $n = 32$	1.46 (2.65) $n = 41$	7.91 (6.29) $n = 64$	5.14 (5.77) $n = 257$
3	6.69 (6.50) $n = 59$	2.68 (3.65) $n = 57$	6.25 (5.75) $n = 32$	1.85 (4.19) $n = 41$	7.92 (6.11) $n = 61$	5.23 (5.88) $n = 250$
4	5.92 (6.39) $n = 60$	2.71 (3.36) $n = 56$	5.17 (5.52) $n = 29$	1.49 (2.21) $n = 39$	8.11 (6.27) $n = 59$	4.91 (5.66) $n = 243$

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Assessment	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
5	5.61 (6.25) n = 56	2.43 (3.20) n = 56	4.79 (5.12) n = 29	1.08 (1.85) n = 38	7.47 (5.77) n = 57	4.47 (5.34) n = 236
6	4.84 (5.88) n = 51	2.78 (3.56) n = 54	NA	NA	NA	3.74 (4.91) n = 106
Depression						
1	9.32 (6.43) n = 63	5.94 (5.24) n = 63	7.97 (6.34) n = 33	3.24 (3.65) n = 43	11.91 (6.22) n = 67	8.03 (6.42) n = 269
2	7.92 (5.92) n = 62	4.72 (5.05) n = 58	6.44 (5.86) n = 32	2.22 (3.49) n = 41	10.46 (7.02) n = 64	6.74 (6.34) n = 257
3	8.07 (6.94) n = 59	4.64 (5.10) n = 57	7.38 (6.01) n = 32	2.49 (4.18) n = 41	9.87 (7.24) n = 61	6.73 (6.62) n = 250
4	7.18 (6.48) n = 60	4.27 (4.63) n = 56	6.62 (5.25) n = 29	1.95 (2.62) n = 39	10.20 (7.28) n = 59	6.34 (6.32) n = 243
5	7.30 (6.41) n = 56	4.05 (4.76) n = 56	5.76 (4.98) n = 29	1.68 (2.58) n = 38	10.07 (7.16) n = 57	6.10 (6.28) n = 236
6	6.29 (5.98) n = 51	3.96 (4.70) n = 54	NA	NA	NA	5.05 (5.46) n = 106
Hostility						
1	6.74 (5.20) n = 64	4.02 (3.95) n = 63	5.58 (5.01) n = 33	1.93 (2.77) n = 43	6.50 (5.47) n = 67	5.14 (4.95) n = 270
2	5.66 (4.56) n = 62	2.53 (3.72) n = 58	4.34 (3.92) n = 32	1.71 (1.98) n = 41	5.76 (5.22) n = 64	4.19 (4.47) n = 257
3	5.78 (5.23) n = 59	3.10 (4.04) n = 57	5.12 (4.16) n = 32	1.71 (2.78) n = 41	6.30 (5.12) n = 61	4.54 (4.76) n = 250
4	5.30 (5.50) n = 60	3.66 (4.54) n = 56	5.31 (5.25) n = 29	1.91 (2.64) n = 39	6.64 (6.23) n = 59	4.70 (5.31) n = 243
5	5.61 (5.15) n = 56	3.45 (4.36) n = 56	5.17 (4.23) n = 29	1.21 (1.92) n = 38	5.79 (4.96) n = 57	4.38 (4.69) n = 236
6	4.32 (4.77) n = 51	3.56 (4.64) n = 54	NA	NA	NA	3.89 (4.69) n = 106
Interpersonal Sensitivity						
1	5.39 (4.67) n = 63	3.43 (3.22) n = 63	5.15 (3.99) n = 33	2.40 (3.43) n = 43	6.91 (3.96) n = 67	4.80 (4.20) n = 269
2	5.14 (4.02) n = 62	2.36 (2.81) n = 58	3.72 (3.60) n = 32	1.68 (2.36) n = 41	5.98 (4.04) n = 64	4.00 (3.85) n = 257
3	4.62 (4.44) n = 60	2.44 (3.21) n = 57	4.44 (3.99) n = 32	1.27 (2.42) n = 41	6.28 (4.34) n = 61	3.96 (4.19) n = 251
4	4.00 (4.07) n = 60	2.64 (3.11) n = 56	3.90 (4.04) n = 29	1.23 (2.04) n = 39	6.39 (4.57) n = 59	3.81 (4.10) n = 243
5	4.42 (4.21) n = 56	2.52 (2.79) n = 56	3.41 (3.63) n = 29	0.89 (1.72) n = 38	6.56 (4.33) n = 57	3.80 (4.02) n = 236
6	4.13 (4.43) n = 51	2.24 (2.72) n = 54	NA	NA	NA	3.13 (3.75) n = 106
Obsessive-Compulsive						
1	9.76 (6.52) n = 64	5.41 (4.19) n = 63	10.02 (6.54) n = 33	3.02 (3.29) n = 43	11.03 (5.36) n = 67	8.02 (6.07) n = 270
2	8.92 (6.48) n = 62	4.28 (4.09) n = 58	8.31 (5.99) n = 32	2.78 (3.35) n = 41	10.22 (6.44) n = 64	7.14 (6.18) n = 257
3	8.60 (7.16) n = 59	4.46 (4.14) n = 57	9.03 (5.86) n = 32	2.66 (3.76) n = 41	9.95 (6.66) n = 61	7.06 (6.39) n = 250
4	7.87 (6.54) n = 60	4.75 (96) n = 56	8.14 (5.60) n = 29	2.26 (3.19) n = 39	10.54 (6.62) n = 59	6.93 (6.32) n = 243
5	8.07 (6.51) n = 56	4.27 (4.33) n = 56	7.59 (5.21) n = 29	2.23 (3.19) n = 38	9.18 (6.36) n = 57	6.44 (5.95) n = 236

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Assessment	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
6	6.96 (6.09) n = 51	4.23 (4.62) n = 54	NA	NA	NA	5.50 (5.53) n = 106
Paranoid Ideation						
1	8.66 (4.78) n = 64	6.68 (4.52) n = 62	7.94 (5.12) n = 33	4.49 (4.71) n = 43	9.29 (4.90) n = 67	7.60 (5.03) n = 269
2	7.71 (4.41) n = 62	4.72 (3.95) n = 58	5.88 (4.93) n = 32	3.56 (3.79) n = 41	8.78 (5.32) n = 64	6.41 (4.90) n = 257
3	7.51 (5.35) n = 59	4.95 (3.98) n = 57	6.75 (4.98) n = 32	2.70 (3.63) n = 41	9.26 (4.84) n = 61	6.47 (5.11) n = 250
4	6.43 (5.06) n = 60	5.24 (4.45) n = 56	5.31 (3.87) n = 29	3.02 (3.22) n = 39	9.78 (5.39) n = 59	6.29 (5.11) n = 243
5	7.10 (5.23) n = 56	5.18 (4.73) n = 56	5.45 (4.73) n = 29	2.47 (3.83) n = 38	8.56 (4.71) n = 57	6.05 (5.10) n = 236
6	6.78 (5.30) n = 51	5.07 (4.48) n = 54	NA	NA	NA	5.85 (4.96) n = 106
Phobic Anxiety						
1	5.06 (5.18) n = 64	1.89 (3.28) n = 63	3.82 (4.81) n = 33	1.28 (2.37) n = 43	6.15 (5.15) n = 67	3.84 (4.75) n = 270
2	4.60 (5.09) n = 62	1.43 (2.70) n = 58	2.72 (3.87) n = 32	0.83 (1.73) n = 41	4.86 (5.21) n = 64	3.11 (4.42) n = 257
3	4.56 (5.31) n = 59	1.22 (2.48) n = 57	3.59 (4.65) n = 32	0.86 (1.67) n = 41	5.41 (5.22) n = 61	3.28 (4.59) n = 250
4	3.98 (5.50) n = 59	1.05 (2.57) n = 56	2.41 (3.64) n = 29	0.69 (1.30) n = 39	5.47 (4.84) n = 59	2.95 (4.44) n = 242
5	3.84 (4.53) n = 56	1.27 (2.34) n = 56	2.3 (3.67) n = 29	0.50 (1.11) n = 38	4.74 (4.71) n = 57	2.80 (3.97) n = 236
6	3.20 (4.74) n = 51	1.13 (2.84) n = 54	NA	NA	NA	2.11 (3.99) n = 106
Psychoticism						
1	7.30 (5.16) n = 63	4.83 (4.26) n = 63	6.27 (5.20) n = 33	3.23 (3.51) n = 43	9.05 (4.99) n = 67	6.38 (5.08) n = 269
2	6.84 (4.72) n = 62	3.28 (3.31) n = 58	5.28 (4.62) n = 32	2.07 (3.09) n = 41	7.94 (4.93) n = 64	5.36 (4.76) n = 257
3	6.20 (5.16) n = 59	3.32 (3.41) n = 57	5.44 (4.58) n = 32	2.07 (3.75) n = 41	7.90 (5.18) n = 61	5.18 (4.96) n = 250
4	5.06 (4.81) n = 60	3.52 (3.34) n = 56	4.66 (3.70) n = 29	1.90 (2.73) n = 39	8.04 (5.19) n = 59	4.87 (4.66) n = 243
5	5.48 (5.04) n = 56	3.54 (3.77) n = 56	4.76 (3.38) n = 29	1.50 (3.16) n = 38	7.60 (4.81) n = 57	4.80 (4.67) n = 236
6	5.27 (4.64) n = 51	3.14 (3.72) n = 54	NA	NA	NA	4.14 (4.31) n = 106
Somatization						
1	7.64 (7.35) n = 64	3.89 (5.22) n = 63	5.21 (5.21) n = 33	2.46 (3.31) n = 43	8.09 (5.76) n = 67	5.76 (6.07) n = 270
2	5.76 (5.63) n = 62	2.59 (4.27) n = 58	4.34 (4.45) n = 32	1.61 (2.99) n = 41	7.07 (5.62) n = 64	4.53 (5.23) n = 257
3	6.12 (6.59) n = 59	3.37 (3.82) n = 57	5.59 (6.23) n = 32	1.66 (2.97) n = 41	6.67 (6.52) n = 61	4.83 (5.77) n = 250
4	5.55 (6.65) n = 60	2.93 (4.38) n = 56	3.58 (3.87) n = 29	1.20 (2.23) n = 39	7.02 (6.50) n = 59	4.37 (5.63) n = 243
5	5.09 (5.18) n = 56	3.05 (4.35) n = 56	4.52 (5.12) n = 29	1.47 (3.34) n = 38	6.30 (6.33) n = 57	4.24 (5.53) n = 236
6	3.74 (5.50) n = 51	2.71 (3.82) n = 54	NA	NA	NA	3.18 (4.70) n = 106

Coolidge Correctional Inventory (CCI)

The Coolidge Correctional Inventory (CCI; Coolidge, 2004) is an adaptation of the Coolidge Axis II Inventory (CATI; Coolidge, n.d. a; Coolidge, Segal, Klebe, Cahill, & Whitcomb, 2009) designed for use by CDOC with new prison admissions to identify personality disorders and neuropsychological problems among inmates. The assessment follows a self-report format with a 4-point scale (1—*strongly false* to 4—*strongly true*) across 250 items. Scores are obtained for a total of 33 different scales (Coolidge et al., 2009) based on the American Psychiatric Association's (2000) diagnostic criteria (DSM-IV-TR). The CCI can be used to assess 14 personality disorders – 10 from the DSM-IV-TR Axis II, 2 from the DSM-IV-TR (American Psychiatric Association, 2000) appendix, and 2 from the DSM-III Axis II (American Psychiatric Association, 1980; Coolidge et al., 2009). The personality disorders assessed by the CCI are as follows: Antisocial, Avoidant, Borderline, Dependent, Depressive, Histrionic, Narcissistic, Obsessive-Compulsive, Paranoid, Passive-Aggressive, Sadistic, Schizoid, Schizotypal, and Self-Defeating. Furthermore, the CCI is used to assess other psychological and neuropsychological problems and syndromes (i.e., Introversive-Extroversion, Maladjustment, Executive Functions, Decision Difficulty, Planning Problems, Neuropsychological Dysfunction, Language, Memory, Neurosomatic Issues, Hostility-Anger, Hostility-Danger, Hostility-Impulsivity, Hypersensitivity, Drug and Alcohol Problems) as well as five selected Axis I scales and associated subscales (i.e., ADHD, Post-Traumatic Stress Disorder, Psychotic Thinking, Schizophrenia, Social Phobia, Withdrawal, Anxiety, and Depression; Coolidge et al., 2009). The CCI also has response validity scales available. For this study, the CCI personality disorders and Axis I scales were used as potential predictors of outcomes.

The Cronbach's alphas for the CCI's subscales were found to be acceptable within prison populations, with a median Cronbach's alpha of .78 for the personality subscales (range = .65 - .86) in a sample of 3,962 inmates (Coolidge et al., 2009) and a median Cronbach's alpha of .75 (range = .47 - .84) in a sample of 3,090 inmates (Whitcomb, 2006). Mean scores on the personality disorders subscales of the CCI ranged from 42.76 ($SD = 8.59$) to 54.27 ($SD = 10.88$), with a median of 47.67 (Coolidge et al., 2009). In another study, Whitcomb (2006) found mean scores ranging from 41.25 ($SD = 9.47$) to 58.80 ($SD = 9.97$) for violent offenders and from 41.38 ($SD = 9.50$) to 58.71 ($SD = 9.29$) for nonviolent offenders.

Because the CCI is an adaptation of the CATI and little research has been done on the CCI, test-retest reliability as well as convergent validity can, at the very least, be evaluated for the CATI as it measures many of the same components as the CCI but was not designed for correctional populations (Coolidge, n.d. a, Coolidge, n.d. b). One-week test-retest reliabilities were found to be strong, with an average correlation of .90 for the personality disorders (Coolidge, n.d. b). Scores on personality disorder scales of the CATI were correlated with scores on the respective Brief Millon Clinical Multiaxial Inventory II (MCMI-II) scales (range = .10 to .87; $Mdn = .58$; Coolidge, n.d. b).

Cronbach's alphas for the single assessment period of the CCI were varied with values ranging between .46 and .88 ($M = .74$). The majority of the internal consistency estimates were greater than .70 with lower estimates for Histrionic (.66), Self-defeating (.64), Schizoid (.55), and Impulsivity (.46) scales.

Deliberate Self-Harm Inventory (DSHI)

In order to assess self-harming behavior in inmates who participated in this study, we used the Deliberate Self-Harm Inventory (DSHI; Gratz, 2001; Gratz & Chapman, 2007). The DSHI is a 17-item measure that questions respondents about various self-harming behaviors. Engagement in as well as frequency of engagement

in different self-harming behaviors are assessed. At the first testing session in this study, participants were asked about their lifetime history of deliberate self-harm. Specifically, they were asked to indicate whether they have ever engaged in the various self-harming behaviors and, if so, how old they were when they first engaged in the activity, how many times they engaged in the activity, when they most recently engaged in the behavior, how many years they engaged in the behavior, and whether engaging in the activity ever led to required medical treatment and/or hospitalization. Completion of this assessment takes most people less than 5 minutes (Fliege et al., 2006). For this study, self-harming behavior was coded as a dichotomous variable; a self-harm total score was computed by summing the *yes/no* responses across the 17 self-harming behaviors. Additionally, the DSHI was administered at the last testing period for some of the participants; however, the number of participants was small and those assessments were not used in this report.

Internal consistency reliability for the DSHI in clinical populations was found to be .81, with a split-half correlation of .78 (Fliege et al., 2006). Item-total correlations in Fliege et al.'s (2006) study were between .23 and .55. Item-total correlations in a nonclinical sample ranged from .00 to .65, with a median item-total correlation of .45 (Gratz, 2001). Two- to four-week test-retest reliability was acceptable at .91 in a clinical population (Fliege et al., 2006) and at .68 in a nonclinical population (Gratz, 2001). Among a nonclinical population, the internal consistency coefficient was .82 for the DSHI (Gratz, 2001). Gratz also assessed the convergent validity of the DSHI, finding significant moderate correlations with other self-harm measures (range = .35 to .49), such as the mental health history self-harm item, Diagnostic Interview for Borderlines-Revised self-harm item, and Suicide Behaviors Questionnaire self-harm item.

For the current sample, the internal consistency estimate for the DSHI with dichotomous responses on the 17 items was acceptable at .84. Table B7 provides the proportion of people who responded yes to each item and summary statistics for the total score.

Table B7. Summary Statistics for the DSHI Items and Total Score

Item	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Cutting	47%	17%	33%	5%	61%	35%
Burn with cigarette	44%	19%	24%	12%	25%	26%
Burn with match/lighter	33%	14%	24%	9%	22%	21%
Carved words into skin	16%	14%	21%	5%	25%	17%
Carved pictures into skin	14%	14%	15%	2%	18%	13%
Purposefully scratched	14%	5%	6%	0%	24%	11%
Bitten self (broke skin)	9%	2%	3%	0%	18%	7%
Rubbed sandpaper on body	3%	3%	0%	0%	9%	4%
Dripped acid on skin	3%	0%	0%	2%	3%	2%
Used bleach, comet, oven cleaner to scrub skin	6%	2%	0%	0%	4%	3%
Stuck sharp objects into skin	30%	11%	9%	2%	28%	18%
Rubbed glass into skin	5%	0%	0%	0%	6%	3%
Broken own bones	5%	2%	0%	0%	6%	3%
Banged head	20%	5%	18%	2%	27%	15%
Punched self	11%	3%	9%	2%	13%	8%
Prevented wounds from healing	8%	0%	3%	2%	16%	7%
Any other self-harm	34%	8%	18%	0%	42%	23%
Total <i>M (SD)</i>	3.05 (3.43)	1.19 (1.82)	1.85 (2.28)	0.46 (1.44)	3.50 (3.48)	2.17 (2.96)

Note. Item statistics are the percentage of persons who indicated a history of the self-harm behavior over their lifetime prior to the study.

Personality Assessment Screener (PAS)

The PAS is a quick and effective screening tool that gauges the social functioning of an individual globally and across 10 subscales (viz., Negative Affect, Acting Out, Health Problems, Psychotic Features, Social Withdrawal, Hostile Control, Suicidal Thinking, Alienation, Alcohol Problem, Anger Control; Harrison & Rogers, 2007). The alcohol problem subscale was not included in this study because it was not a construct of interest. The PAS is a 22-item screening measure that was originally derived from the larger Personality Assessment Inventory (Morey, 1991). Respondents rate each statement on a 4-point scale (F—*false*, ST—*slightly true*, MT—*mostly true*, VT—*very true*); higher scores on this measure indicate greater severity of clinical problems (Morey, 1997) or problems with impression management, as Holden, Book, Edwards, Wasylkiw, and Starzyk (2003) termed it. In order to complete this test, a 4th grade reading level is required; it should take no longer than 5 minutes to complete this assessment (Morey, 1997). For this study, raw scores rather than P-scores were evaluated.

Levels of internal consistency reliability are acceptable for the PAS total score as well as for subscale scores. In a sample of county jail inmates ($N = 100$), the Cronbach's alpha was .74 (Harrison & Rogers, 2007) for total scores but lower for subscales, which is likely due to the small number of items (2) on each subscale. Despite the low number of items on each subscale, 6 of the 10 subscales exhibited alpha coefficients of .60 or greater (i.e., Negative Affect, Health Problems, Psychotic Features, Social Withdrawal, Suicidal Thinking, and Alienation; Harrison & Rogers, 2007); alpha coefficients on the remaining subscales were not provided in Harrison and Rogers' (2007) study. Alpha coefficients for the PAS total score have also been assessed in clinical and nonclinical samples. In a nonclinical sample, internal consistency was found to be .75 for the total score and ranged between .34 (i.e., Alcohol Problem) and .68 (i.e., Suicidal Thinking) for subscales (Morey, 1997). In a clinical sample, internal consistency was found to be .79 for the total score and ranged between .48 and .84 for subscales (Morey, 1997).

Both the total score and subscale scores were assessed to have good test-retest reliability. For a nonclinical sample, 1-month test-retest reliability was .89 for the total PAS score and ranged between .66 and .92 for the subscales (Morey, 1997), with a median test-retest reliability of .77 across subscales. For a clinical sample, 1-month test-retest reliability was .85 for the total PAS score and ranged between .47 and .81 for the subscales (Morey, 1997), with a median test-retest reliability of .66 across subscales.

Normative data on the PAS are available for both clinical and nonclinical populations. The mean raw PAS total score for a nonclinical, community sample was found to be 16.66 ($SD = 7.40$; Morey, 1997). Mean raw scores on subscales ranged from .37 ($SD = .94$) to 4.05 ($SD = .54$; Holden et al., 2003; Morey, 1997). The mean raw PAS total score for a clinical sample was found to be 25.83 ($SD = 9.99$; Morey, 1997). Mean raw scores on subscales ranged from 1.19 ($SD = 1.53$) to 4.99 ($SD = 2.48$; Morey, 1997). Additionally, the PAS has shown good convergent validity (Morey, 1997). The total score on the PAS has been positively correlated with scores on the PAI and MMPI (Gondolf, 2008; Morey, 1997). Furthermore, adequate convergent validity has also been shown for the different subscales of the PAS (Gondolf, 2008; Morey, 1997).

Table B8 provides the summary statistics for the study groups on the PAS scales at each time period. Cronbach's alphas were computed to assess internal consistency reliability with coefficients ranging between .27 and .95 ($M = .64$). These estimates were somewhat lower than those found in the literature. The lowest reliability estimates were for the Acting Out and Hostile Control subscales. Test-retest correlation coefficients ranged between .54 and .84 ($M = .69$). Correlations of the PAS subscales with other self-report measures of

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

the same construct ranged between .34 and .67 ($M = .50$) and with correctional officer and clinician ratings the correlations ranged between .08 and .34 ($M = .21$).

Table B8. Summary Statistics (M , SD , n) on PAS Scales by Group and Time

Assessment	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Acting Out						
Time 1	6.38 (1.94) $n = 64$	6.40 (2.06) $n = 63$	7.24 (1.85) $n = 33$	6.40 (1.90) $n = 43$	6.79 (1.74) $n = 67$	6.59 (1.91) $n = 270$
Time 2	6.21 (2.07) $n = 62$	6.01 (1.99) $n = 59$	6.66 (1.94) $n = 32$	6.29 (1.93) $n = 41$	6.26 (1.88) $n = 64$	6.24 (1.96) $n = 258$
Time 3	5.98 (2.04) $n = 60$	5.79 (2.15) $n = 57$	6.81 (1.49) $n = 32$	5.95 (2.12) $n = 41$	6.03 (1.97) $n = 61$	6.05 (2.01) $n = 251$
Time 4	5.80 (2.07) $n = 60$	5.84 (2.21) $n = 56$	6.48 (1.57) $n = 29$	5.51 (2.39) $n = 39$	6.23 (2.02) $n = 59$	5.95 (2.10) $n = 243$
Time 5	5.52 (2.25) $n = 56$	6.07 (2.21) $n = 56$	6.66 (1.26) $n = 29$	5.78 (1.89) $n = 38$	6.32 (2.15) $n = 57$	6.02 (2.08) $n = 236$
Time 6	5.81 (2.42) $n = 51$	6.15 (2.06) $n = 54$	NA	NA	NA	5.96 (2.25) $n = 106$
Alienation						
Time 1	3.44 (1.96) $n = 63$	3.08 (1.71) $n = 63$	3.88 (1.71) $n = 33$	2.63 (1.83) $n = 43$	4.06 (1.94) $n = 67$	3.43 (1.90) $n = 269$
Time 2	4.12 (1.73) $n = 60$	3.41 (1.92) $n = 59$	3.84 (1.48) $n = 32$	2.93 (1.65) $n = 41$	4.42 (1.73) $n = 64$	3.80 (1.80) $n = 256$
Time 3	3.92 (1.82) $n = 60$	3.35 (1.88) $n = 57$	4.06 (1.41) $n = 32$	2.83 (1.67) $n = 41$	4.44 (1.68) $n = 61$	3.76 (1.80) $n = 251$
Time 4	3.86 (1.81) $n = 59$	3.55 (1.90) $n = 56$	3.32 (1.47) $n = 28$	2.85 (1.71) $n = 39$	4.58 (1.45) $n = 59$	3.74 (1.78) $n = 241$
Time 5	3.86 (1.66) $n = 56$	3.45 (1.84) $n = 56$	3.48 (1.30) $n = 29$	2.84 (1.79) $n = 37$	4.46 (1.42) $n = 57$	3.70 (1.70) $n = 235$
Time 6	3.78 (1.79) $n = 51$	3.52 (1.79) $n = 54$	NA	NA	NA	3.64 (1.78) $n = 106$
Anger Control						
Time 1	3.52 (1.82) $n = 64$	2.79 (1.63) $n = 63$	3.06 (1.69) $n = 33$	2.19 (1.30) $n = 43$	2.92 (1.78) $n = 67$	2.93 (1.72) $n = 270$
Time 2	3.39 (1.83) $n = 62$	2.47 (1.56) $n = 58$	2.91 (1.78) $n = 32$	2.10 (1.18) $n = 41$	3.03 (1.94) $n = 64$	2.82 (1.75) $n = 257$
Time 3	3.38 (1.76) $n = 60$	2.70 (1.65) $n = 56$	2.81 (1.47) $n = 32$	2.12 (1.42) $n = 41$	3.02 (1.78) $n = 61$	2.86 (1.69) $n = 250$
Time 4	3.24 (1.72) $n = 59$	2.61 (1.67) $n = 56$	2.66 (1.54) $n = 29$	2.08 (1.44) $n = 39$	2.90 (1.87) $n = 59$	2.75 (1.72) $n = 242$
Time 5	3.09 (1.69) $n = 56$	2.55 (1.49) $n = 56$	2.96 (1.73) $n = 28$	2.18 (1.61) $n = 38$	2.89 (1.75) $n = 57$	2.75 (1.67) $n = 235$
Time 6	2.98 (1.61) $n = 51$	2.52 (1.66) $n = 54$	NA	NA	NA	2.74 (1.63) $n = 106$
Health Problems						
Time 1	2.00 (1.62) $n = 64$	1.24 (1.64) $n = 63$	2.30 (1.78) $n = 33$	1.28 (1.79) $n = 43$	2.94 (1.70) $n = 67$	1.98 (1.81) $n = 270$
Time 2	2.05 (2.02) $n = 61$	1.12 (1.66) $n = 58$	1.84 (1.72) $n = 32$	1.17 (1.32) $n = 41$	2.84 (1.64) $n = 64$	1.87 (1.83) $n = 256$
Time 3	2.43 (1.97) $n = 60$	1.18 (1.69) $n = 56$	1.91 (1.75) $n = 32$	1.05 (1.24) $n = 41$	2.54 (1.53) $n = 61$	1.88 (1.77) $n = 250$
Time 4	1.95 (1.66) $n = 59$	1.07 (1.26) $n = 56$	1.62 (1.29) $n = 29$	0.95 (1.10) $n = 39$	2.60 (1.67) $n = 58$	1.70 (1.58) $n = 241$

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Assessment	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Time 5	2.02 (1.93) n = 54	0.94 (1.13) n = 55	2.00 (1.58) n = 29	0.97 (1.28) n = 38	2.53 (1.59) n = 57	1.72 (1.66) n = 233
Time 6	1.96 (1.65) n = 50	1.17 (1.38) n = 54	NA	NA	NA	1.53 (1.56) n = 105
Hostile Control						
Time 1	3.35 (1.76) n = 62	3.44 (1.62) n = 63	2.76 (1.58) n = 33	3.28 (1.45) n = 42	2.61 (1.75) n = 67	3.10 (1.68) n = 267
Time 2	3.00 (1.67) n = 61	3.26 (1.50) n = 58	2.91 (1.55) n = 32	3.07 (1.37) n = 41	2.11 (1.44) n = 64	2.84 (1.56) n = 256
Time 3	2.78 (1.61) n = 60	3.32 (1.66) n = 57	2.81 (1.45) n = 32	3.17 (1.46) n = 41	2.48 (1.51) n = 61	2.90 (1.57) n = 251
Time 4	2.85 (1.63) n = 59	3.29 (1.50) n = 55	3.17 (1.07) n = 29	3.13 (1.32) n = 39	2.51 (1.60) n = 59	2.95 (1.50) n = 241
Time 5	2.87 (1.60) n = 55	3.18 (1.54) n = 55	2.76 (1.33) n = 29	3.16 (1.52) n = 38	2.16 (1.36) n = 55	2.81 (1.52) n = 232
Time 6	2.90 (1.78) n = 50	3.17 (1.60) n = 53	NA	NA	NA	3.04 (1.68) n = 104
Negative Affect						
Time 1	4.89 (2.48) n = 64	3.40 (1.56) n = 63	4.73 (2.30) n = 33	2.64 (1.64) n = 43	5.60 (2.32) n = 67	4.34 (2.35) n = 270
Time 2	4.79 (2.06) n = 62	3.06 (1.88) n = 59	4.56 (2.50) n = 32	2.68 (1.88) n = 41	5.08 (2.16) n = 64	4.10 (2.28) n = 258
Time 3	4.50 (2.20) n = 60	3.17 (1.74) n = 57	4.59 (2.11) n = 32	2.34 (1.77) n = 41	4.95 (2.33) n = 61	3.97 (2.26) n = 251
Time 4	4.68 (2.35) n = 60	3.21 (1.99) n = 56	4.07 (1.93) n = 29	2.38 (1.60) n = 39	5.36 (2.31) n = 59	4.06 (2.34) n = 243
Time 5	4.09 (2.32) n = 56	3.18 (1.88) n = 56	4.17 (1.98) n = 29	2.21 (1.51) n = 38	5.12 (2.25) n = 57	3.83 (2.26) n = 236
Time 6	3.94 (2.14) n = 51	3.06 (1.73) n = 54	NA	NA	NA	3.50 (1.98) n = 106
Psychotic Features						
Time 1	1.95 (1.54) n = 64	1.46 (1.51) n = 63	1.91 (1.55) n = 33	0.95 (1.25) n = 43	2.31 (1.88) n = 65	1.76 (1.64) n = 268
Time 2	1.90 (1.68) n = 61	1.24 (1.28) n = 59	1.62 (1.74) n = 32	0.85 (1.11) n = 41	2.17 (2.05) n = 64	1.61 (1.69) n = 257
Time 3	2.05 (1.82) n = 60	1.23 (1.27) n = 57	1.59 (1.62) n = 32	0.93 (1.33) n = 41	2.38 (1.72) n = 60	1.70 (1.66) n = 250
Time 4	1.80 (1.64) n = 60	1.38 (1.45) n = 56	1.00 (.92) n = 29	1.03 (1.20) n = 39	2.30 (1.87) n = 59	1.60 (1.59) n = 243
Time 5	2.04 (1.70) n = 56	1.45 (1.55) n = 55	1.28 (1.13) n = 29	0.79 (1.23) n = 38	2.14 (1.92) n = 57	1.63 (1.66) n = 235
Time 6	1.74 (1.82) n = 51	1.33 (1.24) n = 54	NA	NA	NA	1.52 (1.56) n = 106
Social Withdrawal						
Time 1	3.14 (1.74) n = 64	2.28 (1.42) n = 63	2.91 (1.88) n = 33	2.05 (1.68) n = 43	3.46 (1.88) n = 67	2.82 (1.78) n = 270
Time 2	3.53 (1.82) n = 62	2.51 (1.68) n = 59	2.75 (1.87) n = 32	2.20 (1.42) n = 41	3.61 (1.81) n = 64	3.01 (1.81) n = 258
Time 3	3.30 (1.82) n = 60	2.43 (1.69) n = 56	2.91 (1.61) n = 32	2.05 (1.38) n = 40	3.73 (1.66) n = 60	2.96 (1.76) n = 248
Time 4	3.52 (1.69) n = 60	2.68 (1.65) n = 56	2.72 (1.41) n = 29	2.31 (1.73) n = 39	3.48 (1.66) n = 58	3.02 (1.71) n = 242
Time 5	3.13 (1.77) n = 56	2.55 (1.62) n = 56	2.66 (1.37) n = 29	2.26 (1.62) n = 38	3.82 (1.75) n = 56	2.96 (1.74) n = 2356

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Assessment	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Time 6	3.16 (1.76) <i>n</i> = 50	2.57 (1.60) <i>n</i> = 53	NA	NA	NA	2.88 (1.70) <i>n</i> = 104
Suicidal Thinking						
Time 1	0.95 (1.64) <i>n</i> = 62	0.35 (1.12) <i>n</i> = 62	0.53 (1.08) <i>n</i> = 32	0.05 (.21) <i>n</i> = 43	2.64 (2.27) <i>n</i> = 67	1.04 (1.81) <i>n</i> = 266
Time 2	0.93 (1.63) <i>n</i> = 61	0.29 (.90) <i>n</i> = 58	0.53 (1.22) <i>n</i> = 32	0.05 (.31) <i>n</i> = 41	2.41 (2.36) <i>n</i> = 64	0.96 (1.78) <i>n</i> = 256
Time 3	0.83 (1.59) <i>n</i> = 58	0.21 (.86) <i>n</i> = 57	0.62 (1.36) <i>n</i> = 32	0.10 (.49) <i>n</i> = 41	2.31 (2.40) <i>n</i> = 61	0.90 (1.77) <i>n</i> = 249
Time 4	1.03 (1.83) <i>n</i> = 60	0.21 (.97) <i>n</i> = 56	0.52 (1.30) <i>n</i> = 29	0.02 (.16) <i>n</i> = 39	2.39 (2.24) <i>n</i> = 59	0.95 (1.80) <i>n</i> = 243
Time 5	0.87 (1.62) <i>n</i> = 55	0.20 (.80) <i>n</i> = 56	0.52 (1.30) <i>n</i> = 29	0.10 (.51) <i>n</i> = 38	2.02 (2.00) <i>n</i> = 57	0.82 (1.58) <i>n</i> = 235
Time 6	0.86 (1.71) <i>n</i> = 51	0.26 (1.08) <i>n</i> = 53	NA	NA	NA	0.55 (1.44) <i>n</i> = 105

Prison Behavior Rating Scale (PBRs)

The PBRs was developed by Cooke (1998) for correctional staff to rate inmates' behaviors in prison. While the use of the PBRs in U.S. prisons has been limited, we were unable to find another rating scale that could be easily used by correctional staff to record direct observations of inmates' behaviors. The PBRs is a 36-item measure comprising three subscales: Anti-Authority, Anxious-Depressed, and Dull-Confused. Higher scores on the PBRs indicate worse behavior assessments of inmates by officers. Correctional staff use a 4-point rating scale (0—*never/rarely*, 1—*sometimes*, 2—*often*, 3—*most of the time*) to rate the inmates' behaviors within the 4 weeks preceding the assessment. The PBRs was modified to use language that was more relevant for a United States sample; the questionnaire is given in Appendix A.

The PBRs demonstrated adequate internal consistency reliability across the three subscales in a sample of 467 male prisoners: .91 for Anti-Authority, .84 for Anxious-Depressed, and .72 for Dull-Confused (Cooke, 1998). Cooke also demonstrated good test-retest reliability over 2 to 3 weeks, with .76 for Anti-Authority, .86 for Anxious-Depressed, and .82 for Dull-Confused.

Table B9 provides the summary statistics for the study groups on the PBRs scales at each time period. Internal consistency estimates for the PBRs scales at each time period are provided in Table B10. These alphas indicate strong internal consistency with a mean alpha of .90. Correlations between sequential time periods are provided in Table B11. Correlations between the first and second testing intervals tended to have the weakest correlation coefficients ($M = .16$); this period is when many of the participants switched facilities so there was a change in raters who may lack familiarity with the participants. Correlations between PBRs scales and relevant clinician ratings were low (range = .08 to .24, $M = .19$) as they were with self-report assessments (range = -.07 to .27, $M = .10$). Although the PBRs shows strong internal consistency estimates and some evidence for test-retest reliability, it does not relate well with other measures of similar constructs.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table B9. Summary Statistics (*M*, *SD*, *n*) on PBRs Scales by Group and Time

Assessment	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Anti-Authority						
Time 1	7.62 (7.37) <i>n</i> = 63	7.68 (7.33) <i>n</i> = 63	6.05 (5.67) <i>n</i> = 26	6.28 (7.39) <i>n</i> = 36	2.90 (5.51) <i>n</i> = 67	6.05 (6.99) <i>n</i> = 255
Time 2	6.21 (7.40) <i>n</i> = 61	5.48 (6.67) <i>n</i> = 59	7.06 (5.07) <i>n</i> = 28	8.02 (6.96) <i>n</i> = 35	4.59 (5.34) <i>n</i> = 64	5.97 (6.19) <i>n</i> = 247
Time 3	6.17 (7.40) <i>n</i> = 57	3.73 (4.75) <i>n</i> = 56	9.01 (9.05) <i>n</i> = 28	6.92 (6.29) <i>n</i> = 34	4.75 (5.62) <i>n</i> = 51	5.71 (6.68) <i>n</i> = 226
Time 4	4.02 (5.84) <i>n</i> = 59	4.56 (4.88) <i>n</i> = 55	6.85 (7.96) <i>n</i> = 26	8.66 (6.32) <i>n</i> = 31	5.34 (6.24) <i>n</i> = 55	5.43 (6.21) <i>n</i> = 225
Time 5	4.88 (6.42) <i>n</i> = 56	3.32 (4.38) <i>n</i> = 54	7.30 (9.92) <i>n</i> = 27	7.63 (7.37) <i>n</i> = 36	6.00 (7.14) <i>n</i> = 54	5.50 (6.98) <i>n</i> = 227
Time 6	3.90 (5.11) <i>n</i> = 49	2.48 (4.27) <i>n</i> = 54	NA	NA	NA	3.15 (4.72) <i>n</i> = 103
Anxious-Depressed						
Time 1	6.87 (7.26) <i>n</i> = 64	3.56 (4.61) <i>n</i> = 63	4.08 (5.73) <i>n</i> = 26	2.89 (3.59) <i>n</i> = 36	5.25 (5.06) <i>n</i> = 67	4.79 (5.65) <i>n</i> = 256
Time 2	4.39 (5.81) <i>n</i> = 61	2.00 (3.42) <i>n</i> = 59	5.24 (6.58) <i>n</i> = 28	3.59 (4.08) <i>n</i> = 34	5.25 (5.06) <i>n</i> = 256	4.30 (5.32) <i>n</i> = 246
Time 3	3.14 (4.10) <i>n</i> = 57	1.51 (2.72) <i>n</i> = 56	5.81 (3.91) <i>n</i> = 28	3.97 (3.64) <i>n</i> = 33	6.31 (5.46) <i>n</i> = 246	3.76 (4.18) <i>n</i> = 224
Time 4	3.41 (4.49) <i>n</i> = 59	1.51 (3.29) <i>n</i> = 55	3.59 (4.37) <i>n</i> = 26	8.66 (5.06) <i>n</i> = 30	5.73 (4.73) <i>n</i> = 224	3.73 (5.04) <i>n</i> = 224
Time 5	3.45 (4.57) <i>n</i> = 55	1.74 (3.21) <i>n</i> = 54	3.70 (4.90) <i>n</i> = 27	3.51 (5.39) <i>n</i> = 35	6.36 (6.20) <i>n</i> = 224	3.99 (5.40) <i>n</i> = 225
Time 6	4.00 (5.24) <i>n</i> = 48	1.63 (3.06) <i>n</i> = 54	NA	NA	NA	2.74 (4.37) <i>n</i> = 102
Dull-Confused						
Time 1	3.80 (4.04) <i>n</i> = 64	2.14 (2.90) <i>n</i> = 63	6.05 (5.67) <i>n</i> = 26	6.28 (7.39) <i>n</i> = 36	3.51 (3.94) <i>n</i> = 67	2.94 (3.50) <i>n</i> = 256
Time 2	3.08 (4.01) <i>n</i> = 61	1.38 (2.44) <i>n</i> = 58	7.06 (5.07) <i>n</i> = 28	8.02 (6.96) <i>n</i> = 34	4.18 (4.05) <i>n</i> = 64	2.78 (3.50) <i>n</i> = 245
Time 3	2.35 (2.93) <i>n</i> = 57	0.93 (1.23) <i>n</i> = 56	9.01 (9.05) <i>n</i> = 28	6.92 (6.29) <i>n</i> = 33	3.84 (3.63) <i>n</i> = 51	2.36 (2.85) <i>n</i> = 225
Time 4	2.78 (3.55) <i>n</i> = 59	0.96 (1.41) <i>n</i> = 55	6.85 (7.96) <i>n</i> = 26	8.66 (6.32) <i>n</i> = 31	4.01 (3.96) <i>n</i> = 54	2.45 (3.29) <i>n</i> = 225
Time 5	2.64 (3.43) <i>n</i> = 55	1.05 (1.56) <i>n</i> = 53	7.30 (9.92) <i>n</i> = 27	7.63 (7.37) <i>n</i> = 35	4.69 (4.82) <i>n</i> = 54	2.54 (3.65) <i>n</i> = 224
Time 6	3.12 (3.64) <i>n</i> = 49	1.37 (2.64) <i>n</i> = 54	NA	NA	NA	2.20 (3.26) <i>n</i> = 103
Total Score						
Time 1	18.53 (15.94) <i>n</i> = 64	13.53 (12.84) <i>n</i> = 63	13.17 (13.80) <i>n</i> = 26	11.06 (11.93) <i>n</i> = 36	11.90 (13.52) <i>n</i> = 67	13.97 (14.01) <i>n</i> = 256
Time 2	13.98 (14.41) <i>n</i> = 61	9.00 (11.15) <i>n</i> = 59	15.12 (12.58) <i>n</i> = 28	13.64 (10.66) <i>n</i> = 34	15.43 (12.16) <i>n</i> = 64	13.25 (12.54) <i>n</i> = 246
Time 3	11.90 (13.52) <i>n</i> = 57	6.23 (7.60) <i>n</i> = 56	17.77 (12.88) <i>n</i> = 28	13.27(9.58) <i>n</i> = 33	14.65 (11.77) <i>n</i> = 51	12.05 (11.58) <i>n</i> = 225
Time 4	10.55 (11.75) <i>n</i> = 59	7.10 (7.44) <i>n</i> = 55	12.63 (12.74) <i>n</i> = 26	14.45 (11.97) <i>n</i> = 30	15.96 (15.24) <i>n</i> = 54	11.77 (12.34) <i>n</i> = 224
Time 5	11.10 (12.04) <i>n</i> = 55	6.38 (7.35) <i>n</i> = 54	12.26 (14.76) <i>n</i> = 27	13.54 (14.87) <i>n</i> = 35	18.39 (16.48) <i>n</i> = 54	12.24 (13.74) <i>n</i> = 225
Time 6	10.94 (12.11) <i>n</i> = 48	6.38 (7.35) <i>n</i> = 54	NA	NA	NA	8.06 (10.53) <i>n</i> = 102

Table B10. Internal Consistency Estimates (Cronbach's alpha) for PBRS Scales at each Time Period

PBRS Scale	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6
Anti-Authority	.94	.93	.94	.94	.95	.90
Anxious-Depressed	.90	.90	.94	.90	.90	.90
Dull-Confused	.84	.84	.78	.85	.87	.82
Total	.95	.94	.94	.95	.95	.94

Table B11. Test-retest Correlations between Consecutive Time Periods for each PBRS Subscale

Measure	T1-T2	T2-T3	T3-T4	T4-T5	T5-T6
Anti-Authority	.24	.36	.33	.55	.59
Anxious-Depressed	.14	.33	.46	.58	.48
Dull-Confused	.08	.38	.39	.55	.31
Total	.16	.38	.42	.66	.51

Prison Symptom Inventory (PSI)

The PSI was created by the research staff for this study to measure variables that were not assessed by other existing psychological measures but were thought to be important in association with long-term segregation. Using the literature concerning the impact of AS on psychological functioning (e.g., Grassian, 1983; Haney, 2003), questions were written to assess symptoms associated with this form of confinement, including nervousness, headaches, lethargy, chronic tiredness, trouble sleeping, a sense of impending breakdown, perspiring hands, heart palpitations, dizziness, nightmares, trembling hands, and fainting. Furthermore, questions about exercise, grooming, and safety issues within administration segregation were included in the PSI. The scale has 39 items, rated on a 6-point scale (0- never true to 5- always true). Questions were grouped into the following nine areas: fear level, safety, panic disorder, sensitivity to external stimuli, physical hygiene, physical well-being and exercise, mental well-being, mutism, and attitudes about administrative segregation. The questionnaire is given in Appendix A.

Three subscales were used as part of the major constructs of interest: panic disorder as a measure of anxiety, sensitivity to external stimuli as a measure of hypersensitivity, and physical well-being and exercise as a measure of somatization. Fear level, safety, and attitudes about segregation subscales were used as predictors of how people changed over time, rather than as outcome variables. Analyses comparing groups on all of the PSI scales are included in Appendix C. Table B12 provides the summary statistics for the study groups on PSI scales at each time period.

Table B12. Summary Statistics (M, SD, n) on PSI Scales by Group and Time

Assessment	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Attitudes about Segregation						
Time 1	2.95 (3.19) n = 57	1.68 (2.57) n = 56	2.46 (3.00) n = 28	1.00 (2.10) n = 30	4.58 (3.47) n = 67	2.81 (3.24) n = 238
Time 2	2.97 (3.54) n = 61	1.68 (2.63) n = 56	2.24 (2.76) n = 25	1.54 (2.42) n = 26	5.55 (3.30) n = 60	3.09 (3.42) n = 228
Time 3	3.02 (3.36) n = 55	1.04 (2.39) n = 56	2.04 (2.30) n = 25	1.22 (1.60) n = 27	4.62 (3.45) n = 55	2.58 (3.18) n = 218
Time 4	2.60 (3.21) n = 55	1.29 (2.43) n = 55	1.91 (2.45) n = 22	1.83 (2.58) n = 24	4.96 (3.42) n = 46	2.61 (3.20) n = 202
Time 5	3.12 (3.51) n = 52	1.45 (2.32) n = 55	2.54 (2.67) n = 22	1.30 (1.98) n = 20	5.24 (3.57) n = 45	2.89 (3.32) n = 194

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Assessment	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Time 6	2.57 (3.45) n = 49	1.45 (2.32) n = 52	NA	NA	NA	1.92 (2.98) n = 102
Fear Level						
Time 1	6.25 (4.62) n = 64	4.17 (3.46) n = 63	4.94 (3.78) n = 33	3.51 (3.03) n = 43	7.63 (4.15) n = 67	5.51 (4.18) n = 270
Time 2	5.50 (4.18) n = 62	3.51 (2.54) n = 59	4.53 (3.37) n = 32	3.46 (2.60) n = 41	7.14 (4.96) n = 64	5.01 (4.00) n = 258
Time 3	5.71 (4.10) n = 58	3.91 (2.73) n = 57	4.88 (3.40) n = 32	3.53 (2.07) n = 41	7.66 (4.00) n = 61	5.31 (3.74) n = 250
Time 4	5.22 (3.83) n = 59	3.91 (2.96) n = 56	4.91 (3.08) n = 29	3.26 (2.53) n = 39	6.76 (4.09) n = 59	4.94 (3.63) n = 241
Time 5	5.50 (3.50) n = 58	3.87 (2.75) n = 56	4.79 (3.21) n = 29	3.39 (2.49) n = 38	6.81 (4.30) n = 57	5.00 (3.58) n = 236
Time 6	5.43 (3.43) n = 51	4.28 (2.72) n = 54	NA	NA	NA	4.84 (3.11) n = 106
Hypersensitivity to External Stimuli						
Time 1	10.54 (4.02) n = 64	9.62 (3.92) n = 63	11.00 (5.38) n = 33	8.44 (3.70) n = 43	9.61 (3.94) n = 67	9.82 (4.16) n = 270
Time 2	10.10 (4.52) n = 62	7.81 (3.86) n = 59	11.06 (3.83) n = 32	8.20 (4.09) n = 41	10.11 (3.96) n = 64	9.40 (4.22) n = 258
Time 3	10.71 (4.65) n = 58	8.33 (4.11) n = 57	11.34 (3.95) n = 32	7.76 (4.13) n = 41	9.72 (4.63) n = 61	9.52 (4.50) n = 249
Time 4	9.99 (4.64) n = 58	9.22 (4.49) n = 56	10.15 (3.78) n = 29	8.00 (3.20) n = 39	9.22 (4.49) n = 59	9.55 (4.10) n = 241
Time 5	9.54 (4.09) n = 56	9.03 (4.09) n = 56	10.65 (4.43) n = 29	7.60 (3.62) n = 38	9.03 (4.09) n = 57	9.32 (4.10) n = 236
Time 6	9.37 (4.33) n = 51	9.02 (3.59) n = 54	NA	NA	NA	9.20 (3.93) n = 106
Mental Well-Being						
Time 1	4.95 (2.48) n = 64	4.48 (2.48) n = 63	5.39 (2.54) n = 33	4.00 (2.43) n = 43	5.19 (2.39) n = 67	4.80 (2.48) n = 270
Time 2	4.88 (2.24) n = 61	3.69 (2.55) n = 59	5.09 (2.61) n = 32	3.24 (2.34) n = 41	5.33 (2.53) n = 63	4.48 (2.56) n = 256
Time 3	4.33 (2.42) n = 58	3.98 (2.41) n = 57	4.97 (2.47) n = 32	3.22 (2.31) n = 41	5.44 (2.61) n = 61	4.42 (2.55) n = 249
Time 4	4.71 (2.43) n = 58	3.96 (2.26) n = 56	4.25 (2.78) n = 28	2.77 (1.56) n = 39	5.58 (2.44) n = 59	4.38 (2.48) n = 240
Time 5	4.13 (2.24) n = 55	3.59 (2.25) n = 56	4.66 (2.54) n = 29	2.66 (2.29) n = 38	5.21 (2.24) n = 56	4.08 (2.43) n = 234
Time 6	4.02 (2.01) n = 51	3.68 (2.52) n = 54	NA	NA	NA	3.86 (2.27) n = 106
Mutism						
Time 1	3.67 (2.19) n = 64	2.65 (1.70) n = 63	3.61 (2.07) n = 33	2.40 (1.50) n = 43	3.81 (1.96) n = 67	3.26 (1.98) n = 270
Time 2	4.44 (2.21) n = 62	2.98 (1.97) n = 59	3.16 (1.87) n = 32	2.20 (1.50) n = 41	4.11 (1.72) n = 64	3.51 (2.04) n = 258
Time 3	4.14 (2.29) n = 57	2.98 (1.81) n = 57	3.19 (1.89) n = 32	2.32 (1.56) n = 41	4.44 (2.28) n = 61	3.52 (2.16) n = 248
Time 4	4.41 (2.44) n = 58	2.96 (1.74) n = 56	3.28 (1.74) n = 28	2.59 (1.44) n = 39	3.95 (2.05) n = 57	3.53 (2.06) n = 238
Time 5	4.09 (2.08) n = 55	3.00 (1.80) n = 55	3.21 (1.76) n = 29	2.26 (1.60) n = 38	3.60 (1.94) n = 57	3.31 (1.95) n = 234
Time 6	4.00 (2.19) n = 50	2.75 (1.69) n = 52	NA	NA	NA	3.38 (2.04) n = 103

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Assessment	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Panic Disorder						
Time 1	9.09 (10.18) n = 64	3.89 (5.49) n = 63	5.03 (4.88) n = 33	2.77 (3.77) n = 43	10.98 (7.72) n = 67	6.84 (7.84) n = 270
Time 2	7.87 (8.62) n = 62	3.71 (7.72) n = 59	5.43 (5.12) n = 32	2.37 (3.45) n = 41	9.61 (8.70) n = 64	6.18 (7.62) n = 258
Time 3	8.76 (9.17) n = 60	3.46 (4.20) n = 57	6.11 (5.02) n = 32	3.00 (4.68) n = 41	9.55 (8.68) n = 61	6.47 (7.65) n = 251
Time 4	7.32 (7.72) n = 60	3.79 (5.44) n = 56	4.99 (5.06) n = 29	2.22 (3.28) n = 39	9.94 (8.68) n = 59	6.04 (7.19) n = 243
Time 5	6.50 (8.80) n = 56	3.45 (4.34) n = 56	4.08 (5.13) n = 29	1.82 (3.49) n = 38	8.07 (7.45) n = 57	5.10 (6.78) n = 236
Time 6	5.34 (7.25) n = 51	3.60 (5.78) n = 54	NA	NA	NA	4.41 (6.54) n = 106
Physical Hygiene						
Time 1	5.39 (5.01) n = 64	4.00 (4.03) n = 63	4.64 (2.69) n = 33	2.07 (2.54) n = 43	8.26 (4.64) n = 67	5.16 (4.67) n = 270
Time 2	5.80 (4.75) n = 61	4.06 (3.87) n = 59	3.66 (3.26) n = 32	1.74 (2.69) n = 41	7.59 (5.60) n = 64	4.93 (4.77) n = 257
Time 3	5.67 (4.80) n = 58	3.44 (3.59) n = 57	4.19 (4.10) n = 32	2.17 (3.38) n = 41	6.19 (4.55) n = 61	4.52 (4.40) n = 249
Time 4	5.48 (5.43) n = 58	3.25 (3.92) n = 56	3.21 (3.92) n = 29	1.92 (2.67) n = 39	5.61 (4.99) n = 59	4.14 (4.64) n = 241
Time 5	5.73 (5.01) n = 56	3.12 (3.57) n = 56	3.62 (3.70) n = 29	1.08 (1.99) n = 38	5.22 (5.07) n = 57	3.98 (4.46) n = 236
Time 6	4.90 (4.75) n = 51	3.18 (3.68) n = 54	NA	NA	NA	4.12 (4.41) n = 106
Physical Well-being and Exercise						
Time 1	15.89 (7.76) n = 64	10.43 (5.55) n = 63	15.79 (6.28) n = 33	9.21 (5.61) n = 43	18.93 (6.44) n = 67	14.29 (7.40) n = 270
Time 2	17.10 (7.70) n = 62	9.85 (6.23) n = 59	13.68 (7.37) n = 32	7.26 (5.53) n = 41	18.90 (6.01) n = 64	13.90 (7.91) n = 258
Time 3	17.14 (7.05) n = 58	10.30 (6.18) n = 57	13.84 (6.47) n = 32	8.12 (5.02) n = 41	18.98 (6.78) n = 61	14.12 (7.60) n = 249
Time 4	16.13 (7.47) n = 58	10.28 (6.14) n = 56	13.07 (5.92) n = 29	7.44 (4.36) n = 39	18.46 (6.67) n = 59	13.56 (7.48) n = 241
Time 5	15.39 (7.49) n = 56	9.54 (6.25) n = 56	13.58 (6.78) n = 29	7.08 (4.19) n = 38	17.60 (7.01) n = 57	12.97 (7.57) n = 236
Time 6	13.95 (7.09) n = 51	9.26 (6.72) n = 54	NA	NA	NA	12.97 (7.57) n = 106

Internal consistency estimates are provided in Table B13 for all the PSI subscales. Cronbach's alphas ranged between -.02 and .90 for the three scales related to the study constructs. Panic Disorder had strong reliability estimates ranging between .88 and .90. Hypersensitivity to External Stimuli demonstrated poor internal consistency with alpha estimates ranging between .27 and .41 ($M = .34$). The internal consistency was adequate for the Physical Well-being and Exercise subscale with values ranging from .72 to .76 ($M = .74$). Test-retest correlation coefficients are provided in Table B14 and indicate stability between time periods across all subscales with correlations ranging between .45 and .83 ($M = .67$).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table B13. Internal Consistency Estimates at each Time Period for each PSI Subscale

Measure	# Items	Items*	T1	T2	T3	T4	T5	T6
Attitudes about Segregation	2	r14, 39	.78	.74	.76	.75	.71	.67
Fear Level	4	3, 12, 21, r38	.60	.58	.48	.50	.46	.37
Hypersensitivity to External Stimuli	5	1, 7, r31, r34, 37	.41	.43	.49	.35	.35	.27
Mental Well-being	2	26, r35	.31	.38	.36	.33	.30	-.02
Mutism	2	22, r32	.47	.37	.50	.52	.37	.31
Panic Disorder	9	2, 6, 10, 13, 16, 17, 20, 25, 30	.89	.89	.89	.88	.90	.90
Physical Hygiene	5	4, r9, 18, r23, r29	.62	.65	.61	.68	.66	.60
Physical Well-being and Exercise	8	r5, 8, r11, r15, 19, 24, 27, r28	.72	.76	.73	.75	.76	.74
Safety	2	33, 36	.82	.84	.76	.81	.84	.90
Total	39	all	.90	.90	.90	.90	.89	.88

*r before a number indicates that the item is reversed scored.

Table B14. Test-retest correlations between consecutive time periods for each PSI Subscale

Measure	T1-T2	T2-T3	T3-T4	T4-T5	T5-T6
Attitudes about Segregation	.66	.78	.78	.80	.79
Fear Level	.46	.64	.64	.57	.60
Hypersensitivity to External Stimuli	.45	.62	.58	.58	.53
Mental Well-being	.57	.62	.62	.57	.51
Mutism	.50	.59	.65	.64	.67
Panic Disorder	.68	.71	.78	.75	.74
Physical Hygiene	.67	.68	.73	.65	.64
Physical Well-being and Exercise	.73	.78	.83	.77	.77
Safety	.72	.71	.74	.82	.69

For the three subscales that related to study constructs, correlations with other measures were calculated as assessments of convergent validity. The PSI subscales demonstrated adequate validity estimates with other self-report measures of the constructs (i.e., anxiety, hypersensitivity, and somatization) with correlations ranging between .41 and .61 ($M = .50$) and had lower correlations with staff reports (range = .18 to .39, $M = .29$).

Profile of Mood States (POMS)

Developed by McNair, Lorr, and Droppleman (1971, 1992), the POMS is intended to assess respondents across six mood factors: Tension-Anxiety (heightened musculoskeletal tension), Anger-Hostility (anger and antipathy towards others), Fatigue-Inertia (weariness, inertia, low energy), Depression-Dejection (depression and sense of inadequacy), Vigor-Activity (vigorousness, ebullience, high energy), and Confusion-Bewilderment (bewilderment and muddle-headedness). The POMS is a 65-item self-report measure; higher scores on the POMS indicate more negative feelings held over the past week (McNair & Heuchert, 2006). Respondents rate each item on how well it describes them in the past week, using a 5-point rating scale (0 – *not at all* to 4 – *extremely*). Completion of the POMS takes approximately 15 to 20 minutes and requires an 8th grade reading level (Lorr, McNair, Heuchert, & Droppleman, 2003; McNair et al., 1992).

Acceptable levels of internal consistency (as measured by Kuder-Richardson Formula 20) for the subscales were found in a sample of 350 male psychiatric outpatients, ranging from .86 to .95 (McNair & Heuchert, 2006; Norcross, Guadagnoli, & Prochaska, 1984). Test-retest reliability was assessed in psychiatric outpatients over the course of 3 to 110 days, with a median number of 20 days between tests. Stability coefficients were found to range between .65 and .74, with a median of .69 (McNair & Heuchert, 2006).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or views expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Normative data are available for a variety of populations but not for a prison population. Means for outpatients ranged from 10.0 ($SD = 6.5$) to 26.0 ($SD = 15.8$; McNair & Heuchert, 2006; Norcross et al., 1984). In a nonclinical, community sample comprised of males only, mean scores across subscales ranged from 5.6 ($SD = 4.1$) to 19.8 ($SD = 6.8$; McNair & Heuchert, 2006; Nyenhuis, Yamamoto, Luchetta, Terrien, & Parmentier, 1999). In a nonclinical, male college student sample, mean scores across subscales ranged from 8.6 ($SD = 4.6$) to 15.6 ($SD = 6.0$; McNair & Heuchert, 2006; Nyenhuis et al., 1999). Convergent validity has also been assessed for the POMS and found to be acceptable for the total and subscale scores (Nyenhuis et al., 1999).

Table B15 provides the summary statistics for the study groups on the POMS subscales at each time period. Estimates of internal consistency reliability were strong with Cronbach's alphas ranging between .89 and .96 ($M = .93$). Correlations between sequential time periods indicated stability over time with coefficients ranging between .54 and .80 ($M = .68$). Convergent validity estimates with other self-report measures of the same construct ranged between .35 and .81 ($M = .57$), and with staff reports the coefficients ranged between .14 and .38 ($M = .25$).

Table B15. Summary Statistics (M , SD , n) on POMS Subscales by Group and Time

Assessment	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Anger-Hostility						
Time 1	20.92 (11.86) $n = 64$	14.78 (9.21) $n = 63$	19.28 (11.10) $n = 33$	10.37 (8.80) $n = 43$	18.03 (12.51) $n = 66$	16.88 (11.41) $n = 269$
Time 2	17.46 (11.29) $n = 61$	11.07 (9.36) $n = 58$	17.74 (10.34) $n = 32$	7.54 (6.76) $n = 41$	17.80 (12.07) $n = 64$	14.54 (11.06) $n = 256$
Time 3	16.88 (12.16) $n = 59$	12.23 (10.88) $n = 57$	18.61 (10.48) $n = 31$	7.67 (8.05) $n = 41$	17.74 (10.69) $n = 61$	14.72 (11.30) $n = 249$
Time 4	17.22 (13.07) $n = 60$	12.18 (10.75) $n = 56$	16.87 (11.56) $n = 29$	7.54 (7.46) $n = 39$	18.64 (12.34) $n = 59$	14.81 (12.01) $n = 243$
Time 5	16.02 (12.01) $n = 56$	12.02 (10.64) $n = 55$	15.17 (9.80) $n = 29$	6.71 (7.61) $n = 38$	17.63 (11.84) $n = 57$	13.86 (11.32) $n = 235$
Time 6	14.03 (11.11) $n = 51$	11.65 (10.62) $n = 54$	NA	NA	NA	12.80 (10.88) $n = 105$
Depression-Dejection						
Time 1	25.13 (15.64) $n = 64$	17.17 (13.77) $n = 63$	25.34 (16.57) $n = 33$	12.27 (11.39) $n = 43$	28.92 (14.33) $n = 67$	22.19 (15.51) $n = 270$
Time 2	22.25 (14.17) $n = 61$	13.55 (12.14) $n = 58$	21.56 (13.46) $n = 32$	8.28 (8.66) $n = 41$	26.33 (15.39) $n = 64$	18.98 (14.65) $n = 256$
Time 3	20.69 (14.91) $n = 59$	13.27 (13.42) $n = 57$	23.18 (15.04) $n = 31$	8.56 (12.16) $n = 41$	25.17 (15.23) $n = 61$	18.40 (15.52) $n = 249$
Time 4	21.18 (14.85) $n = 60$	13.90 (12.77) $n = 56$	20.38 (15.02) $n = 29$	7.36 (7.20) $n = 39$	8.56 (12.16) $n = 59$	18.19 (14.80) $n = 243$
Time 5	19.35 (14.48) $n = 56$	12.28 (11.24) $n = 56$	19.03 (13.43) $n = 29$	6.90 (7.73) $n = 38$	7.36 (7.20) $n = 57$	16.97 (14.21) $n = 236$
Time 6	18.49 (13.01) $n = 51$	11.86 (11.94) $n = 54$	NA	NA	NA	15.08 (12.85) $n = 105$
Fatigue-Inertia						
Time 1	10.10 (6.92) $n = 64$	5.99 (6.18) $n = 63$	10.17 (7.26) $n = 33$	4.56 (4.78) $n = 43$	11.26 (6.76) $n = 66$	8.54 (6.92) $n = 269$
Time 2	9.36 (6.65) $n = 61$	4.60 (4.36) $n = 58$	9.44 (5.52) $n = 32$	3.39 (4.15) $n = 41$	10.19 (6.28) $n = 64$	7.54 (6.21) $n = 256$
Time 3	8.84 (7.23) $n = 59$	4.82 (5.32) $n = 57$	10.45 (7.28) $n = 31$	3.63 (4.86) $n = 41$	9.94 (7.02) $n = 61$	7.54 (6.92) $n = 249$
Time 4	9.07 (6.99) $n = 60$	5.20 (5.06) $n = 56$	8.21 (6.98) $n = 29$	3.00 (3.49) $n = 39$	10.55 (7.26) $n = 59$	7.46 (6.73) $n = 243$

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or views expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Assessment	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Time 5	7.59 (6.84) <i>n</i> = 56	5.18 (5.41) <i>n</i> = 56	7.94 (6.87) <i>n</i> = 29	3.05 (4.39) <i>n</i> = 38	11.06 (6.98) <i>n</i> = 57	7.17 (6.75) <i>n</i> = 236
Time 6	6.08 (5.14) <i>n</i> = 51	4.42 (4.85) <i>n</i> = 54	NA	NA	NA	5.23 (5.04) <i>n</i> = 105
Tension-Anxiety						
Time 1	15.83 (8.31) <i>n</i> = 64	10.36 (7.26) <i>n</i> = 63	17.09 (8.71) <i>n</i> = 33	8.60 (6.64) <i>n</i> = 43	17.21 (8.49) <i>n</i> = 67	13.90 (8.61) <i>n</i> = 270
Time 2	14.78 (8.28) <i>n</i> = 62	8.06 (6.01) <i>n</i> = 59	13.97 (7.16) <i>n</i> = 32	6.48 (4.72) <i>n</i> = 41	15.74 (8.46) <i>n</i> = 64	12.06 (8.12) <i>n</i> = 258
Time 3	14.26 (8.84) <i>n</i> = 60	8.40 (7.06) <i>n</i> = 57	14.48 (7.63) <i>n</i> = 31	6.66 (6.02) <i>n</i> = 41	15.47 (8.29) <i>n</i> = 61	12.00 (8.47) <i>n</i> = 250
Time 4	13.04 (8.08) <i>n</i> = 60	8.37 (7.13) <i>n</i> = 56	13.09 (9.11) <i>n</i> = 29	6.51 (4.99) <i>n</i> = 39	16.16 (8.82) <i>n</i> = 59	11.68 (8.49) <i>n</i> = 243
Time 5	12.46 (8.12) <i>n</i> = 56	7.73 (6.64) <i>n</i> = 56	12.45 (7.25) <i>n</i> = 29	6.10 (5.10) <i>n</i> = 38	6.51 (4.99) <i>n</i> = 57	10.91 (7.74) <i>n</i> = 236
Time 6	11.47 (7.73) <i>n</i> = 51	7.85 (6.34) <i>n</i> = 54	NA	NA	NA	9.56 (7.24) <i>n</i> = 106

Saint Louis University Mental Status (SLUMS) Examination

The SLUMS Examination (Tariq, Tumosa, Chibnall, Perry, & Morley, 2006) is an 11-item screening tool designed to assess mild neurocognitive impairment and dementia. It assesses orientation, memory, attention, and executive functions. Scores on the SLUMS Examination can range from 0 to 30, with higher scores indicating better cognitive functioning. Cut-offs for mild neurocognitive impairment and dementia are provided for persons with varying degrees of education (i.e., more than high school, less than high school; Tariq et al., 2006). Administration takes approximately 7 minutes. While the SLUMS Examination is similar to the Mini Mental Status Exam (i.e., both measures screen for cognitive impairment), the SLUMS Examination may be better for assessing milder cognitive problems, because it is a more sensitive measure (Tariq et al., 2006). Due to its more sensitive nature and its associated ability to detect very mild forms of neurocognitive problems, the SLUMS Examination was selected for this study.

Summary statistics are available for several different populations, including nonclinical populations and older individuals. The mean for the total scale was found to be between 25.7 (*SD* = 2.8) and 26.9 (*SD* = 2.00; Tariq et al., 2006) for a nonclinical adult population, while means ranged from 26.9 (*SD* = 2.5) to 28.1 (*SD* = 2.3; Heeter, Winn, Winn, & Bozoki, 2008) for older adults between 60 and 80 years of age.

Table B16 provides the summary statistics for the study groups on the SLUMS test at each time period. The SLUMS' internal consistency estimates for the present study were low (range = .48 to .60, *M* = .52) which may be reasonable given that this is a screening measure and assesses several cognitive functions. Test-retest reliability estimates were stronger with correlations ranging between .63 and .78 (*M* = .71). Convergent validity was estimated by assessing the relationship of the SLUMS to the Trails B/A task and correctional staff's ratings on the PBRS Dull-Confused subscale. Convergent validity coefficients were small with correlations to the Trails task ranging from .13 to .31 (*M* = .21) and to the PBRS Dull-Confused subscale ranging between .03 and .18 (*M* = .10).

Table B16. Summary Statistics (*M, SD, n*) on SLUMS Score by Group and Time

Time	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
1	20.80 (5.43) <i>n</i> = 64	21.73 (3.34) <i>n</i> = 63	21.52 (4.04) <i>n</i> = 33	23.38 (3.73) <i>n</i> = 43	20.54 (3.73) <i>n</i> = 67	21.45 (4.24) <i>n</i> = 270
2	21.16 (4.77) <i>n</i> = 62	22.64 (3.63) <i>n</i> = 59	23.09 (3.68) <i>n</i> = 32	24.12 (3.23) <i>n</i> = 41	21.49 (4.33) <i>n</i> = 63	22.30 (4.16) <i>n</i> = 257
3	22.26 (4.59) <i>n</i> = 60	24.02 (3.25) <i>n</i> = 57	23.88 (2.88) <i>n</i> = 32	24.49 (3.49) <i>n</i> = 41	22.85 (4.37) <i>n</i> = 61	23.37 (3.95) <i>n</i> = 251
4	22.92 (4.31) <i>n</i> = 60	24.38 (3.03) <i>n</i> = 56	23.34 (3.67) <i>n</i> = 29	24.79 (3.68) <i>n</i> = 39	23.38 (3.91) <i>n</i> = 59	23.84 (3.80) <i>n</i> = 243
5	23.59 (4.04) <i>n</i> = 56	24.25 (3.34) <i>n</i> = 56	24.93 (3.24) <i>n</i> = 29	24.82 (3.24) <i>n</i> = 38	23.26 (4.05) <i>n</i> = 57	24.03 (3.69) <i>n</i> = 236
6	23.94 (4.57) <i>n</i> = 49	25.30 (2.88) <i>n</i> = 54	NA	NA	NA	24.62 (3.82) <i>n</i> = 104

State-Trait Anxiety Inventory (STAI)

The STAI (Spielberger, Gorsuch, & Lushene, 1970) partitions anxiety into that which is attributable to the condition one is in (i.e., state) and into the inherent anxiety of an individual (i.e., trait). It is a 40-item self-report inventory that includes two 20-item subscales. The first subscale assesses state anxiety and is answered on a 4-point scale (1—*not at all*, 2—*somewhat*, 3—*moderately so*, 4—*very much so*); the second subscale assesses trait anxiety and is also answered on a 4-point scale (1—*almost never*, 2—*sometimes*, 3—*often*, 4—*almost always*).

Internal consistency is acceptable for the STAI with coefficients between .81 and .92 and a median .84 (Metzger, 1976). In a variety of nonclinical samples (i.e., college students, high school students, military recruits, working adults), the median alpha coefficient was .60 (Novy, Nelson, Goodwin, & Rowzee, 1993). Across males of three different ethnicities (i.e., White, Black, Latino), alpha coefficients were found to be between .93 and .95 for state anxiety and between .92 and .95 for trait anxiety (Novy et al., 1993). Internal consistency measures are high in prison populations (.83; Zinger et al., 2001). Overall, this inventory is valuable in its ability to distinguish between types of anxiety and because normative data exist for a prisoner population (Spielberger et al., 1970).

Test-retest reliability has been variable for the two subscales on the STAI. In a replication study by Joesting (1976), the STAI was administered both before and after a class examination. Correlations between the two tests were .66 for trait anxiety and .60 for state anxiety (Joesting, 1976). For a 104-day test-retest assessment, test-retest reliability ranged from .73 to .84 (Spielberger et al., 1970). Furthermore, test-retest reliability in another nonclinical sample was found to be .16, .26, and .15 for state anxiety assessed for different intervals (3 months, 8 months, 11 months; Nixon & Steffek, 1977). Test-retest reliability was also assessed for trait anxiety in the same nonclinical sample and was found to be .48, .54, and .29 for trait anxiety assessed for the same three intervals (3 months, 8 months, 11 months; Nixon & Steffek, 1977). In another study with college students, test-retest reliability was found to be .97 for trait anxiety and .45 for state anxiety (Metzger, 1976).

Means for state anxiety in nonclinical populations seem to range from 32.90 (*SD* = 11.10) to 49.20 (*SD* = 11.89), while means for trait anxiety in nonclinical populations ranged from 35.60 (*SD* = 9.90) to 45.89 (*SD* = 12.96; Joesting, 1976; Nixon & Steffek, 1977; Novy et al., 1993; Nyenhuis et al., 1999).

The STAI has shown to be a valid measure, demonstrating convergent validity from .52 to .85 and good discriminant validity (Spielberger, 1983). Novy et al. (1993) found moderate to high correlations between the BDI and the STAI State/Trait scale (range = .59 to .81 for BDI and STAI State; range = .44 to .71 for BDI and STAI Trait) and between the BHS and the STAI State/Trait (range = .67 to .92 for BHS and STAI State; range = .26 to .76 for BHS and STAI Trait).

Table B17 provides the summary statistics for the study groups on the STAI scales at each time period. Cronbach's alpha coefficients ranged between .93 and .95 ($M = .94$) for the two subscales, which indicates strong internal consistency estimates. Correlations between sequential time periods (range = .65 to .82, $M = .73$) suggest good stability over 3 month intervals with trait anxiety showing slightly stronger correlations ($M = .79$) than state anxiety ($M = .68$). Convergent validity with other self-report measures of anxiety indicated good validity with coefficients ranging between .37 and .85 ($M = .64$); however, correlations with staff reports of anxiety were lower (ranging from .07 to .49 with a mean of .23).

Table B17. Summary Statistics (M , SD , n) on STAI Subscales by Group and Time

Assessment	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
State Anxiety						
Time 1	46.90 (12.16) $n = 62$	42.05 (11.42) $n = 63$	47.64 (13.52) $n = 33$	39.39 (12.04) $n = 43$	50.14 (12.72) $n = 67$	45.46 (12.80) $n = 268$
Time 2	45.83 (12.43) $n = 62$	38.43 (10.39) $n = 59$	44.53 (13.14) $n = 32$	36.68 (10.98) $n = 41$	48.42 (13.39) $n = 64$	43.16 (12.86) $n = 258$
Time 3	45.45 (12.40) $n = 60$	37.89 (12.05) $n = 57$	47.29 (11.68) $n = 32$	34.65 (7.35) $n = 41$	48.49 (12.78) $n = 61$	42.83 (12.78) $n = 251$
Time 4	44.01 (13.38) $n = 60$	37.50 (10.89) $n = 56$	45.08 (11.90) $n = 29$	33.80 (8.69) $n = 39$	49.41 (12.48) $n = 59$	42.31 (12.93) $n = 243$
Time 5	42.60 (12.98) $n = 56$	37.46 (11.54) $n = 56$	43.76 (12.07) $n = 29$	33.81 (9.40) $n = 38$	48.29 (11.70) $n = 57$	41.48 (12.65) $n = 236$
Time 6	43.28 (12.23) $n = 51$	36.89 (9.98) $n = 54$	NA	NA	NA	40.09 (11.52) $n = 106$
Trait Anxiety						
Time 1	48.41 (12.36) $n = 62$	42.78 (11.11) $n = 63$	49.70 (12.79) $n = 33$	37.82 (9.99) $n = 43$	54.45 (11.48) $n = 67$	47.06 (12.79) $n = 269$
Time 2	47.89 (11.91) $n = 62$	38.77 (10.25) $n = 59$	46.59 (11.84) $n = 32$	35.93 (10.76) $n = 41$	52.64 (12.28) $n = 64$	44.92 (12.97) $n = 258$
Time 3	47.49 (12.33) $n = 60$	38.40 (10.82) $n = 57$	47.59 (10.05) $n = 32$	34.44 (10.15) $n = 41$	51.50 (12.94) $n = 61$	44.27 (13.10) $n = 250$
Time 4	45.75 (13.11) $n = 60$	39.27 (10.10) $n = 56$	45.78 (10.69) $n = 29$	34.06 (8.86) $n = 39$	52.77 (11.10) $n = 59$	44.09 (12.70) $n = 243$
Time 5	44.65 (12.85) $n = 56$	38.70 (11.02) $n = 56$	44.06 (10.01) $n = 29$	32.74 (8.98) $n = 38$	52.28 (11.65) $n = 57$	43.09 (12.93) $n = 236$
Time 6	43.95 (12.10) $n = 51$	37.54 (10.62) $n = 53$	NA	NA	NA	40.70 (11.71) $n = 105$

Structured Inventory of Malingered Symptomatology (SIMS)

The SIMS (Widows & Smith, 2005) is a 75-item screening measure intended to detect feigned symptoms of psychopathology and cognitive functioning in clinical and forensic settings. A total score and scores on five subscales—Psychosis (bizarre or unusual psychotic symptoms), Neurologic Impairment (illogical or highly atypical neurological symptoms), Amnestic Disorders (symptoms of memory impairment), Low Intelligence (general cognitive incapacity or intellectual deficit), and affective disorders (atypical symptoms of depression and anxiety)—are obtained (Widows & Smith, 2005). The subscales are comprised of 15 items each; comple-

tion of this measure takes approximately 10 to 15 minutes (Widows & Smith, 2005). Participants answer whether statements are applicable to them or are generally considered true (T); if a statement does not describe them or cannot be considered true, an F is circled as the answer choice. The SIMS assesses whether respondents endorse atypical, improbable, inconsistent, or illogical symptoms. Scores above the cutoff mark suggest probable malingering but may also suggest genuine psychopathology. For this study, we used elevations above these cutoff scores as an indicator of possible malingering.

Internal consistency estimates have ranged from .24 to .86 for subscales and were found to be .72 to .88 for total scores (Merckelbach & Smith, 2003; Smith, as cited in Widows & Smith, 2005). Three-week test-retest reliability in honest responders was found to be .72 in a Dutch sample.

Also reported in the manual are validity studies that indicated the SIMS to be a valid screening device for malingering. The SIMS total score correlated strongly with validity scales of the MMPI, including the F scale ("faking bad"; $r = .84$) and F-K index scores ("honesty"; $r = .81$; Widows & Smith, 2005). A moderate correlation ($r = .45$) was found between the SIMS total score and the 16 Personality Factor Questionnaire Faking Bad scale (Widows & Smith, 2005). Furthermore, the SIMS total score was highly correlated with other commonly used indexes of malingering, such as the MMPI-2 validity scales (range = .44 to .51), the Structured Interview of Report Symptoms (SIRS) scales (.43 < r < .80), and the M Test (.46 < r < .67; Heinze & Purisch, 2001). A study by Edens, Poythress, and Watkins-Clay (2009) indicated that the SIMS correlated highly with the SIRS (.81) and Personality Assessment Inventory NIM (.84) and that it correlated moderately with the Personality Assessment Inventory MAL (.68) and RDF (.45) scales. Furthermore, the SIMS total scores have been found to correlate significantly ($p < .01$) with the BDI (.64) as well as the STAI Trait (.55). Thus, the SIMS scales seem to be related to both validity and psychopathology measures.

Mean scores on the SIMS were given in Lewis, Simcox, and Berry's (2002) study. Mean scores across subscales for a forensic sample ranged from 1.2 ($SD = 2.1$) to 5.2 ($SD = 2.6$) and the mean total score for the sample was found to be 14.5 ($SD = 8.8$; Lewis et al., 2002). Edens, Poythress, and Watkins-Clay (2009) also found that the SIMS nearly always correctly classified non-malingering inmates but that there were more errors with mentally ill, such that caution against classification of inmates with mental illness as malingerers is warranted. While it is suggested to administer follow-up tests once an elevated score has been found on the SIMS, this measure by itself is yet another way to gain a more comprehensive picture of the inmates in this study, be it in regards to the degree of their malingering or their psychopathology.

Table B18 provides the summary statistics for the study groups on the SIMS scales at each time period. Internal consistency estimates ranged between .50 and .93 ($M = .76$) with the lowest alphas for the Affective Disorder subscale ($M = .55$) and the Low Intelligence subscale ($M = .59$). Table B19 provides the internal consistency estimates for the SIMS scales at each time period. Test-retest coefficients were quite variable ranging between .06 and .83 ($M = .48$) with total scores showing the least variability (range = .54 to .79, $M = .68$). There were not correlations with other malingering variables to assess validity; however, correlations between subscales were computed for each time period. Correlations between subscales ranged between .34 and .98 with variability in which measures demonstrated the weakest and strongest correlations at each time period.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or views expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table B18. Summary Statistics (*M, SD, n*) on POMS Subscales by Group and Time

Assessment	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Affective Disorders						
Time 1	5.97 (2.53) <i>n</i> = 64	4.53 (2.13) <i>n</i> = 63	6.12 (2.60) <i>n</i> = 33	3.52 (1.89) <i>n</i> = 43	6.88 (2.25) <i>n</i> = 67	5.49 (2.56) <i>n</i> = 270
Time 2	6.07 (2.31) <i>n</i> = 62	5.94 (12.44) <i>n</i> = 60	8.90 (16.38) <i>n</i> = 33	3.17 (1.85) <i>n</i> = 41	8.31 (11.74) <i>n</i> = 65	6.50 (10.36) <i>n</i> = 261
Time 3	6.59 (2.52) <i>n</i> = 60	6.36 (12.59) <i>n</i> = 58	6.31 (2.26) <i>n</i> = 32	3.60 (2.07) <i>n</i> = 41	8.05 (12.09) <i>n</i> = 62	6.37 (8.72) <i>n</i> = 253
Time 4	6.09 (2.57) <i>n</i> = 60	4.94 (2.42) <i>n</i> = 56	6.27 (2.78) <i>n</i> = 29	3.51 (1.83) <i>n</i> = 39	6.19 (2.60) <i>n</i> = 59	5.46 (2.64) <i>n</i> = 243
Time 5	5.62 (2.42) <i>n</i> = 56	4.87 (2.38) <i>n</i> = 56	5.95 (2.29) <i>n</i> = 29	3.34 (1.82) <i>n</i> = 38	6.33 (2.47) <i>n</i> = 57	5.29 (2.51) <i>n</i> = 236
Time 6	6.05 (2.31) <i>n</i> = 51	4.96 (2.57) <i>n</i> = 54	NA	NA	NA	5.49 (2.48) <i>n</i> = 106
Amnestic Disorders						
Time 1	3.16 (3.78) <i>n</i> = 64	1.27 (1.53) <i>n</i> = 63	2.88 (2.94) <i>n</i> = 33	0.70 (1.12) <i>n</i> = 43	4.30 (2.60) <i>n</i> = 67	2.58 (2.93) <i>n</i> = 270
Time 2	3.51 (3.67) <i>n</i> = 62	2.97 (12.74) <i>n</i> = 60	5.42 (17.02) <i>n</i> = 33	1.01 (2.50) <i>n</i> = 41	5.98 (12.24) <i>n</i> = 65	3.85 (10.79) <i>n</i> = 261
Time 3	3.32 (3.63) <i>n</i> = 60	2.76 (12.95) <i>n</i> = 58	2.60 (2.74) <i>n</i> = 32	0.49 (.90) <i>n</i> = 41	5.66 (12.59) <i>n</i> = 62	3.21 (9.12) <i>n</i> = 253
Time 4	2.92 (3.52) <i>n</i> = 60	1.70 (2.54) <i>n</i> = 56	2.66 (3.38) <i>n</i> = 29	0.80 (2.31) <i>n</i> = 39	3.95 (3.73) <i>n</i> = 59	2.51 (3.34) <i>n</i> = 243
Time 5	2.63 (3.29) <i>n</i> = 56	1.27 (2.37) <i>n</i> = 56	2.62 (3.45) <i>n</i> = 29	0.66 (.85) <i>n</i> = 38	3.46 (3.53) <i>n</i> = 57	2.19 (3.07) <i>n</i> = 236
Time 6	2.48 (3.07) <i>n</i> = 51	1.37 (2.63) <i>n</i> = 54	NA	NA	NA	1.89 (2.89) <i>n</i> = 106
Low Intelligence						
Time 1	4.04 (12.24) <i>n</i> = 64	2.33 (1.59) <i>n</i> = 63	2.28 (1.44) <i>n</i> = 33	1.77 (1.56) <i>n</i> = 43	2.97 (2.17) <i>n</i> = 67	2.80 (6.17) <i>n</i> = 270
Time 2	2.53 (2.19) <i>n</i> = 62	4.13 (12.56) <i>n</i> = 60	5.45 (16.90) <i>n</i> = 33	1.60 (1.94) <i>n</i> = 41	4.14 (12.10) <i>n</i> = 65	3.52 (10.50) <i>n</i> = 261
Time 3	3.09 (2.53) <i>n</i> = 60	4.09 (12.80) <i>n</i> = 58	2.06 (1.72) <i>n</i> = 32	1.34 (1.51) <i>n</i> = 41	4.27 (12.43) <i>n</i> = 62	3.19 (8.82) <i>n</i> = 253
Time 4	2.55 (2.33) <i>n</i> = 60	2.23 (2.03) <i>n</i> = 56	2.52 (2.01) <i>n</i> = 29	1.59 (1.98) <i>n</i> = 39	2.56 (1.95) <i>n</i> = 59	2.32 (2.09) <i>n</i> = 243
Time 5	2.53 (2.34) <i>n</i> = 56	2.58 (1.94) <i>n</i> = 56	2.18 (1.97) <i>n</i> = 29	1.50 (1.50) <i>n</i> = 38	2.61 (2.17) <i>n</i> = 57	2.35 (2.06) <i>n</i> = 236
Time 6	2.85 (2.44) <i>n</i> = 51	2.37 (2.10) <i>n</i> = 54	NA	NA	NA	2.62 (2.27) <i>n</i> = 106
Neurological Impairment						
Time 1	4.85 (12.36) <i>n</i> = 64	2.24 (2.09) <i>n</i> = 63	2.88 (2.42) <i>n</i> = 33	1.44 (1.45) <i>n</i> = 43	4.24 (2.92) <i>n</i> = 67	3.31 (6.44) <i>n</i> = 270
Time 2	3.04 (3.23) <i>n</i> = 62	3.70 (12.66) <i>n</i> = 60	5.58 (16.89) <i>n</i> = 33	1.54 (2.18) <i>n</i> = 41	5.74 (12.33) <i>n</i> = 65	3.95 (10.70) <i>n</i> = 261
Time 3	3.27 (3.08) <i>n</i> = 60	3.80 (12.88) <i>n</i> = 58	3.13 (2.79) <i>n</i> = 32	1.30 (1.77) <i>n</i> = 41	5.37 (12.54) <i>n</i> = 62	3.57 (9.00) <i>n</i> = 253
Time 4	2.97 (3.10) <i>n</i> = 60	2.54 (2.79) <i>n</i> = 56	2.59 (2.10) <i>n</i> = 29	1.31 (1.56) <i>n</i> = 39	3.90 (3.52) <i>n</i> = 59	2.78 (2.94) <i>n</i> = 243
Time 5	2.68 (2.98) <i>n</i> = 56	1.93 (2.21) <i>n</i> = 56	3.14 (2.86) <i>n</i> = 29	1.11 (1.41) <i>n</i> = 38	3.83 (3.72) <i>n</i> = 57	2.58 (2.95) <i>n</i> = 236
Time 6	2.87 (2.83) <i>n</i> = 51	2.15 (2.82) <i>n</i> = 54	NA	NA	NA	2.49 (2.82) <i>n</i> = 106

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Assessment	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Psychosis						
Time 1	2.87 (3.52) <i>n</i> = 64	1.10 (1.66) <i>n</i> = 63	1.97 (2.47) <i>n</i> = 33	0.42 (.66) <i>n</i> = 43	4.55 (3.59) <i>n</i> = 67	2.37 (3.12) <i>n</i> = 270
Time 2	2.67 (3.06) <i>n</i> = 62	2.42 (12.73) <i>n</i> = 60	4.48 (17.06) <i>n</i> = 33	0.80 (2.09) <i>n</i> = 41	5.43 (12.35) <i>n</i> = 65	3.24 (10.75) <i>n</i> = 261
Time 3	2.68 (3.06) <i>n</i> = 60	2.66 (12.94) <i>n</i> = 58	1.38 (1.62) <i>n</i> = 32	0.69 (.94) <i>n</i> = 41	5.37 (12.67) <i>n</i> = 62	2.85 (9.06) <i>n</i> = 253
Time 4	2.55 (3.10) <i>n</i> = 60	1.18 (2.12) <i>n</i> = 56	1.79 (2.37) <i>n</i> = 29	0.51 (1.02) <i>n</i> = 39	3.73 (3.92) <i>n</i> = 59	2.11 (3.03) <i>n</i> = 243
Time 5	2.07 (2.84) <i>n</i> = 56	0.86 (1.54) <i>n</i> = 56	1.76 (3.11) <i>n</i> = 29	0.50 (.86) <i>n</i> = 38	3.44 (3.76) <i>n</i> = 57	1.82 (2.87) <i>n</i> = 236
Time 6	2.26 (3.11) <i>n</i> = 51	1.02 (2.34) <i>n</i> = 54	NA	NA	NA	1.63 (2.79) <i>n</i> = 106
Total						
Time 1	18.79 (15.41) <i>n</i> = 64	11.46 (5.85) <i>n</i> = 63	16.13 (8.36) <i>n</i> = 33	7.85 (4.38) <i>n</i> = 43	22.96 (9.56) <i>n</i> = 67	16.05 (11.25) <i>n</i> = 270
Time 2	17.82 (11.97) <i>n</i> = 62	12.55 (12.69) <i>n</i> = 60	17.83 (16.58) <i>n</i> = 33	8.13 (9.43) <i>n</i> = 41	23.49 (15.63) <i>n</i> = 65	16.50 (14.33) <i>n</i> = 261
Time 3	18.94 (12.00) <i>n</i> = 60	12.83 (12.82) <i>n</i> = 58	15.47 (7.88) <i>n</i> = 32	7.41 (4.78) <i>n</i> = 41	22.30 (16.32) <i>n</i> = 62	16.06 (13.15) <i>n</i> = 253
Time 4	17.07 (12.16) <i>n</i> = 60	12.59 (8.97) <i>n</i> = 56	15.81 (9.60) <i>n</i> = 29	7.72 (5.17) <i>n</i> = 39	20.33 (12.53) <i>n</i> = 59	15.18 (11.17) <i>n</i> = 243
Time 5	15.53 (10.87) <i>n</i> = 56	11.51 (7.43) <i>n</i> = 56	15.63 (10.48) <i>n</i> = 29	7.11 (4.06) <i>n</i> = 38	19.68 (13.31) <i>n</i> = 57	14.23 (10.78) <i>n</i> = 236
Time 6	16.53 (10.90) <i>n</i> = 51	11.88 (9.33) <i>n</i> = 54	NA	NA	NA	14.13 (10.29) <i>n</i> = 106

Table B19. Internal Consistency Estimates (Cronbach's alpha) for SIMS Scales at each Time Period

SIMS Scale	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6
Affective disorder	.55	.60	.58	.56	.52	.50
Amnesia	.80	.86	.86	.88	.87	.87
Low Intelligence	.52	.52	.62	.63	.60	.65
Neurological Impairment	.75	.82	.79	.81	.83	.83
Psychosis	.85	.85	.84	.87	.86	.87
Total	.90	.92	.92	.93	.92	.91

Trail Making Test (TMT)

The TMT (Reitan, 1958) measures neurocognitive deficits related to attention, speed, and mental flexibility. There are two tasks (A and B) and the length of time to complete each task was recorded as total score for each task. Completion time on this measure varies widely but is generally around 5 to 10 minutes for both tasks (Strauss, Sherman, & Spreen, n.d.). While individuals connect only numbers in ascending order on Trails A, they have to connect numbers and letters alternately in ascending order for Trails B (Tombaugh, 2004). We computed two derived scores—ratio of times on the two tasks (B/A) and the difference between times on the two tasks (B – A); these derived scores provide an indication of the time difference between Trails A and Trails B (Tombaugh, 2004); however, for the analysis in the report the Trails ratio (B/A) was used to assess change over time. The TMT has been shown to be sensitive to neurocognitive deficits (Sherrill-Pattison, Donders, & Thompson, 2000); it is important to consider age and education of participants when interpreting scores though.

The TMT has demonstrated adequate test-retest reliability over a 14- to 24-week period (*Mdn* = 20 weeks; Trails A *r* = .46; Trails B *r* = .44; Matarazzo, Wiens, Matarazzo, & Goldstein, 1974) although stability may be impacted by population groups and time intervals. Practice effects might be a problem (McCaffrey, Ortega, & Haase, 1993) although research has shown that practice effects between administrations separated by at least 3 months may be negligible (e.g., Basso, Bornstein, & Lang, 1999). Another study found that the TMT's 3-week test-retest reliability was moderate to high, with Trails A having a correlation of .55 and Trails B having a correlation of .75 (Bornstein, Baker, & Douglass, 1987).

Normative data are available on nonclinical populations, separated by age group (Tombaugh, 2004). For people aged 18 to 59, mean times on the Trails A ranged from 22.93 (*SD* = 6.87; 18-24 years) to 35.10 (*SD* = 10.94; 55-59 years); mean times on the Trails B ranged from 48.97 (*SD* = 12.69; 18-24 years) to 78.84 (*SD* = 19.09; 55-59 years; Tombaugh, 2004). Additionally, Matarazzo et al. (1974) found means for the Trails A and B to be 21.76 (*SD* = 5.65) and 54.17 (*SD* = 12.54), respectively. Descriptive statistics were also provided on the two scores that will be derived in the current study. While B-A was found to have a mean of 39.7 (*SD* = 21.5), B/A was found to have a mean of 2.1 (*SD* = .6) for an older adult, community-dwelling sample (Sánchez-Cubillo et al., 2009). Convergent validity is adequate for the total scores on each of the trials as well as for the B-A score (Sánchez-Cubillo et al., 2009). The ratio score of B/A did not show significant correlations with any of the other assessed cognitive measures in Sánchez-Cubillo et al.'s (2009) study but Periañez et al. (2007) suggested that the B/A might be a purer measure of executive functioning.

After consultation with a neuropsychologist, it was decided to use a derived score by taking the ratio of time to complete Task B with time to complete Task A; however, information about all Trails scores are provided in this section. Table B20 provides the summary statistics for the study groups on the Trails tasks at each time period. Correlations between sequential time periods for the entire sample are given in Table B21. Correlations between the Trails tasks at each time period ranged between .21 and .97 (absolute values of correlations are given; the Trails B/A was always negatively correlated with Trails A time). Table B21 also provides the mean correlation of each task with the other tasks over time. The Trails tasks were correlated with performance on the SLUMS and the PBRs Dull-Confused subscale. Correlation coefficients were small with the SLUMS (range = .13 to .31, *M* = .21) and with the correctional officer ratings on the PBRs Dull-Confused (range = -.12 to .09, *M* = -.01), indicating that these measures are assessing distinct aspects of cognitive functioning.

Table B20. Summary Statistics (*M*, *SD*, *n*) on TMT Scores by Group and Time

Assessment	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Task A Time						
Time 1	29.36 (14.18) <i>n</i> = 62	27.20 (8.76) <i>n</i> = 61	25.65 (7.70) <i>n</i> = 33	24.28 (6.00) <i>n</i> = 43	32.36 (19.80) <i>n</i> = 67	28.34 (13.51) <i>n</i> = 266
Time 2	29.70 (17.34) <i>n</i> = 61	24.46 (7.00) <i>n</i> = 59	24.09 (8.22) <i>n</i> = 32	22.25 (4.84) <i>n</i> = 41	29.46 (13.36) <i>n</i> = 64	26.54 (12.10) <i>n</i> = 256
Time 3	29.45 (21.46) <i>n</i> = 60	23.06 (7.36) <i>n</i> = 57	24.52 (6.79) <i>n</i> = 32	21.70 (5.97) <i>n</i> = 41	28.13 (15.10) <i>n</i> = 61	25.78 (14.01) <i>n</i> = 251
Time 4	29.78 (19.01) <i>n</i> = 59	22.63 (7.94) <i>n</i> = 56	21.99 (7.32) <i>n</i> = 29	21.04 (4.17) <i>n</i> = 39	26.64 (11.00) <i>n</i> = 59	25.02 (12.27) <i>n</i> = 242
Time 5	27.54 (15.32) <i>n</i> = 56	23.48 (8.91) <i>n</i> = 56	21.34 (6.67) <i>n</i> = 29	20.97 (3.94) <i>n</i> = 38	27.57 (15.45) <i>n</i> = 57	24.78 (12.11) <i>n</i> = 235
Time 6	27.27 (15.95) <i>n</i> = 48	20.91 (5.59) <i>n</i> = 54	NA	NA	NA	23.81 (12.02) <i>n</i> = 103

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Assessment	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Task B Time						
Time 1	84.70 (55.39) n = 61	83.46 (42.82) n = 62	82.46 (34.64) n = 33	68.94 (23.83) n = 43	96.29 (58.26) n = 67	84.50 (47.67) n = 266
Time 2	77.74 (44.98) n = 60	71.09 (29.85) n = 59	66.30 (27.77) n = 31	70.63 (35.41) n = 40	81.57 (34.51) n = 63	74.62 (35.86) n = 253
Time 3	75.42 (45.66) n = 60	66.33 (26.20) n = 57	66.11 (27.25) n = 32	63.41 (24.79) n = 41	78.26 (38.12) n = 61	70.90 (34.93) n = 251
Time 4	71.63 (37.92) n = 59	62.78 (25.66) n = 56	58.29 (22.17) n = 29	56.78 (19.63) n = 38	74.69 (37.81) n = 59	66.38 (31.78) n = 241
Time 5	69.43 (32.73) n = 56	62.27 (30.01) n = 56	54.98 (20.41) n = 29	58.81 (19.96) n = 38	72.71 (41.64) n = 57	65.04 (32.04) n = 236
Time 6	62.75 (32.75) n = 48	55.84 (20.71) n = 54	NA	NA	NA	59.09 (27.13) n = 102
B – A Time						
Time 1	55.27 (47.23) n = 61	56.33 (40.60) n = 60	56.81 (31.14) n = 33	44.66 (21.75) n = 43	63.93 (46.29) n = 67	56.17 (40.64) n = 264
Time 2	48.82 (35.29) n = 60	46.63 (26.48) n = 59	42.40 (21.97) n = 31	48.41 (34.89) n = 40	52.11 (27.20) n = 63	48.28 (29.82) n = 253
Time 3	45.98 (31.68) n = 60	43.27 (24.97) n = 57	41.59 (23.35) n = 32	41.71 (22.39) n = 41	50.13 (30.18) n = 61	45.12 (27.45) n = 251
Time 4	41.84 (27.51) n = 59	40.15 (23.39) n = 56	36.30 (19.15) n = 29	35.81 (18.01) n = 38	48.04 (30.67) n = 59	41.35 (25.45) n = 241
Time 5	41.89 (23.20) n = 56	38.78 (25.33) n = 56	33.64 (18.59) n = 29	37.58 (18.81) n = 37	45.14 (32.39) n = 57	40.24 (25.25) n = 235
Time 6	34.48 (22.88) n = 48	34.93 (18.53) n = 54	NA	NA	NA	35.19 (20.59) n = 102
B/A Ratio						
Time 1	2.95 (1.11) n = 61	3.19 (1.54) n = 60	3.30 (1.21) n = 33	2.89 (.88) n = 43	3.04 (1.12) n = 67	3.06 (1.20) n = 264
Time 2	2.84 (1.19) n = 60	2.94 (1.01) n = 59	2.80 (.74) n = 31	3.27 (1.84) n = 40	2.80 (.74) n = 63	2.95 (1.20) n = 253
Time 3	2.68 (.82) n = 60	3.02 (1.24) n = 57	2.71 (.84) n = 32	2.97 (1.00) n = 41	2.71 (.84) n = 61	2.87 (1.05) n = 251
Time 4	2.59 (1.00) n = 59	2.92 (1.15) n = 56	2.74 (.84) n = 29	2.72 (.82) n = 38	2.74 (.84) n = 59	2.77 (.99) n = 241
Time 5	2.64 (.74) n = 56	2.73 (1.00) n = 56	2.67 (.90) n = 29	2.87 (1.05) n = 37	2.67 (.90) n = 57	2.72 (.99) n = 235
Time 6	2.40 (.74) n = 48	2.73 (.86) n = 54	NA	NA	NA	2.57 (.82) n = 102

Table B21. Test-Retest Correlation Coefficients for Trails Tasks at Consecutive Testing Intervals

Trails Task	Time 1 to 2	Time 2 to 3	Time 3 to 4	Time 4 to 5	Time 5 to 6	r*
Trails A	.70	.65	.83	.81	.87	.41
Trails B	.66	.70	.73	.76	.75	.71
Trails B – A	.58	.63	.58	.63	.59	.69
Trails B/A	.36	.37	.44	.44	.39	.50

*means correlation of task with other Trails tasks over time.

Trauma Symptom Inventory (TSI)

The TSI (Briere, 1995) is a 100-item self-report assessment of posttraumatic stress and other psychological consequences of traumatic events, including but not limited to rape, child abuse, spouse abuse, physical assault, combat, major accidents, and natural disasters. Respondents use a 4-point rating scale (0—never to

3—often) to report on the experience of 100 events that could have occurred within the last 6 months. Scores are obtained on three validity scales (Atypical Response, Response Level, and Inconsistent Response) and 10 clinical symptom domains (Anxious Arousal, Depression, Anger/Irritability, Intrusive Experiences, Defensive Avoidance, Dissociation, Sexual Concerns, Dysfunctional Sexual Behavior, Impaired Self-Reference, and Tension Reduction Behavior). Greater scores indicate more symptoms associated with trauma (Fernandez, 1998). It has been found that a fifth- to seventh-grade reading level is required to complete the TSI; it takes approximately 20 minutes to complete this assessment (Fernandez, 1998).

The TSI's clinical subscales have demonstrated internal reliability with different samples. Alpha coefficients ranged between .74 and .91 for the standardization sample (nonclinical) with a median alpha coefficient of .88; alphas ranged from .69 to .90 with a median alpha of .86 for the university sample; furthermore, internal consistency reliability ranged from .74 to .90 with a median alpha coefficient of .89 for a clinical sample (Briere, 1995). The three validity scales were also found to have internal consistency reliabilities between .51 and .80 across the standardization as well as a military sample (Briere, 1995). There are no known test-retest reliability estimates or studies completed with prison populations.

Normative data are available by gender and age groups for the general population as well as for a clinical sample separated by gender (Briere, 1995). Means on the subscales ranged from 2.32 ($SD = 4.20$) to 7.69 ($SD = 6.03$) for nonclinical, younger males (i.e., 18-54; Briere, 1995). Clinical samples were separated by trauma history; means for males in the group without trauma history ranged from 2.24 ($SD = 3.17$) to 9.45 ($SD = 5.90$), whereas means for males in the group with trauma history ranged from 6.30 ($SD = 7.44$) to 16.32 ($SD = 5.27$; Briere, 1995).

Reasonable convergent validity was found between the TSI and other measures, such as the BSI, Symptom Checklist, Impact of Event Scale, and the Personality Assessment Inventory (Briere, 1995). More specifically, three of the TSI's clinical subscales that are most closely associated with subscales of the BSI were found to have high correlations: Anxious Arousal (TSI) and Anxiety (BSI) had a high correlation of .75, Anger/Irritability (TSI) and Hostility (BSI) correlated at .77, and Defensive Avoidance (TSI) and Depression (BSI) had a correlation of .82 (Briere, 1995). The correlations of the TSI subscales and the subscales from two posttraumatic stress scales (i.e., Impact of Event Scale, Symptom Checklist) were found to range between .35 and .74, indicating convergent validity (Briere, 1995). The Sexual Concern subscale of the TSI was moderately correlated ($r = .53$) with a measure tapping into sexual concerns (Briere, 1995). Furthermore, the Dysfunctional Sexual Behavior subscale was moderately correlated ($r = .32$) with a question on another measure on how many sexual partners the person had over the course of the past 12 months as well as with four questions assessing the likelihood of the participant engaging in sex with an attractive stranger ($r = .32$), sex with any stranger ($r = .19$), posing for pornography ($r = .22$), and sex for money ($r = .17$; Briere, 1995). The Borderline Personality Disorder subscale of the PAI was found to correlate moderately with the Impaired Self-Reference scale ($r = .65$) as well as the Tension Reduction Behavior scale ($r = .54$; Briere, 1995).

Table B22 provides the summary statistics for the study groups on the TSI Total Score. The internal consistency estimates were similar across groups and high ($M = .97$).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or views expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table B22. Summary Statistics (*M*, *SD*, *n*) on TSI Total Score by Group and Time

Statistic	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
<i>M</i>	98.43	53.66	87.86	40.64	119.55	83.12
<i>SD</i>	58.85	42.04	50.06	30.34	58.40	58.17
<i>n</i>	62	60	33	41	60	262
Cronbach's α	.98	.97	.96	.96	.97	.98

NORMATIVE COMPARISONS

Because we used standardized assessments in this study, it is possible to compare scores for the study sample to normative data. In this section, comparisons were made between each study groups' mean and the normative mean using a one sample *t* test. Normative values were taken from the test manuals when available or were gathered from the literature. Normative data from general adult populations were typically used; if male norms were available they were used. If only normative data for clinical samples were available then outpatient norms were used. One sample *t* tests indicated that in general, for all groups except the GP NMI group, scores were elevated above the normative data when participants entered the study and tended to stay that way. There were also fewer elevations on the SIMS malingering subscales with study groups frequently scoring similarly to "honest responders." Table B23 provides a visual representation of the significant differences by group at each time period on each measure. Red shading indicates that the group mean is significantly different from the normative mean in the direction of more psychological or cognitive problems, whereas green shading indicates that the group mean is significantly better than the normative mean. No shading indicates the groups were statistically similar to the normative data.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or views expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table B23. Significant Differences of Study Groups from Normative Means

Measure	Norm Mean	CSP MI						CSP NMI						GP MI					GP NMI					SCCF				
		1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
BHS-college	2.32																											
BHS-clinical	6.04																											
BPRS Total	49.29																											
BSI Anxiety	1.56																											
BSI Depression	1.26																											
BSI Hostility	1.70																											
BSI Interpersonal Sensitivity	0.96																											
BSI Obsessive-Compulsive	2.22																											
BSI Paranoid Ideation	1.65																											
BSI Phobic Anxiety	0.55																											
BSI Psychotic	0.75																											
BSI Somatization	1.61																											
BSI GSI	0.25																											
PAS Acting Out	2.45																											
PAS Alienation	1.70																											
PAS Anger Control	1.73																											
PAS Health Problems	1.13																											
PAS Hostile Control	2.52																											
PAS Negative Affect	2.84																											
PAS Psychotic Features	0.71																											
PAS Social Withdrawal	2.17																											
PAS Suicidal Thinking	0.37																											
PAS Total	16.66																											
POMS Anger-Hostility	7.10																											
POMS Depression-Dejection	7.50																											
POMS Fatigue-Inertia	7.30																											
POMS Tension-Anxiety	7.10																											
POMS Vigor	<19.80																											
POMS Total	14.80																											
SIMS Affective Disorders	5.2																											
SIMS Amnestic Disorders	2.5																											
SIMS Low Intelligence	3.2																											
SIMS Neurological Impairment	2.4																											
SIMS Psychosis	1.2																											
SIMS Total	14.5																											
SLUMS	<25.70																											
STAI-State	35.72																											
STAI-Trait	34.89																											
Trails A Time	22.93																											
Trails B Time	48.97																											
Trails B – A	29																											
Trails B/A	2.18																											

*There are six assessments for CSP groups and five for the other three groups.

COMPOSITE SCORES

A composite score was developed for seven of the eight primary constructs by standardizing scores from the scales on the self-report assessments. Clinician and correctional officer ratings are not included in composites so that comparisons between self-report and staff reports can be made. Self-report scores were stan-

standardized so that comparisons between different measures could be made more easily and to create a single measure for constructs assessed by multiple self-report assessments. Scores were standardized by centering on the mean of the entire sample at the first assessment and dividing by the standard deviation. A composite score was computed by standardizing each assessment and averaging the standardized scores across the individual assessments as the composite score. Internal consistency reliability estimates, test-retest correlations, and validity coefficient estimates for these composites and associated subscales are presented within the discussion of each construct. Composite means and standard deviations are reported in the main body of the report.

Anxiety Construct

Anxiety was measured by eight self-report variables assessed at each time period. The self-report measures used to create the anxiety composite score were the State and Trait subscales of the STAI; the Anxiety, Obsessive-Compulsive, and Phobic Anxiety subscales of the BSI; the Negative Affect subscale of the PAS; the Tension-Anxiety subscale of the POMS; and the Panic Disorder subscale of the PSI. This construct was also assessed with ratings by correctional staff (PBRs Anxious-Depressed) and clinicians (BPRS Anxiety-Depression).

Internal consistency reliabilities were computed for each assessment period for the entire sample. The internal consistency estimates are provided in Table B24. The reliability estimates were strong, indicating good internal consistency at each time period for the composite. Reliabilities for individual scales were similar across testing intervals, and they were similar to internal consistency estimates from normative samples; only the BPRS showed low internal consistency estimates.

Table B24. Internal Consistency Estimates for Anxiety Construct

Measure	# of Items	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6
Anxiety Composite	8	.89	.90	.91	.91	.90	.89
BSI Anxiety	6	.86	.90	.91	.91	.90	.90
BSI Obsessive-Compulsive	6	.88	.89	.90	.91	.88	.88
BSI Phobic Anxiety	5	.83	.86	.86	.87	.82	.88
PAS Negative Affect	3	.68	.65	.65	.70	.65	.60
POMS Tension-Anxiety	9	.91	.91	.92	.92	.91	.89
PSI Panic Disorder	9	.89	.89	.89	.89	.90	.90
STAI State	10	.93	.94	.94	.94	.94	.93
STAI Trait	10	.93	.94	.94	.94	.95	.93
PBRs Anxious-Depressed	14	.90	.90	.94	.90	.90	.90
BPRS Anxious-Depressed	5	.55	NA	.60	NA	.66	NA

Test-retest correlations for the anxiety composite are provided in Table B25 and indicate stable constructs for three month assessments.

Table B25. Test-Retest Correlation Coefficients for Anxiety Composite at Consecutive Time Periods for each Study Group

Interval	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Time 1-2	.76	.57	.71	.80	.49	.56
Time 2-3	.73	.86	.82	.84	.71	.82
Time 3-4	.86	.86	.75	.70	.73	.84
Time 4-5	.83	.76	.77	.76	.71	.83
Time 5-6	.75	.83	NA	NA	NA	.80

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

The validity coefficients between anxiety measures are provided in Table B26 and indicate good convergent validity for all measures except for the PBRS Anxious-Depressed subscale. The pattern of relationships is similar across time periods.

Table B26. Correlations between Anxiety Construct Measures at each Time Period

Time 1	BSI OC	BSI PA	PAS NA	POMS TA	PSI PD	STAI-S	STAI-T	PBRS AD	BPRS AD
BSI Anxiety	.79	.75	.65	.81	.71	.60	.68	.05	.42
BSI Obsessive-Compulsive		.66	.69	.70	.64	.55	.66	.12	.35
BSI Phobic Anxiety			.50	.58	.56	.46	.50	.05	.33
PAS Negative Affect				.71	.46	.62	.76	.12	.35
POMS Tension-Anxiety					.52	.70	.76	.10	.37
PSI Panic Disorder						.37	.41	.06	.35
STAI-State							.78	.07	.36
STAI-Trait								.11	.37
PBRS Anxious Depressed									.33
BPRS Anxious-Depressed									
Time 2	BSI OC	BSI PA	PAS NA	POMS TA	PSI PD	STAI-S	STAI-T	PBRS AD	BPRS AD
BSI Anxiety	.78	.78	.62	.82	.73	.63	.67	.10	
BSI Obsessive-Compulsive		.69	.60	.69	.68	.59	.68	.04	
BSI Phobic Anxiety			.55	.59	.65	.54	.56	.09	
PAS Negative Affect				.66	.39	.63	.79	.13	
POMS Tension-Anxiety					.56	.72	.78	.11	
PSI Panic Disorder						.44	.46	.06	
STAI-State							.83	.18	
STAI-Trait								.17	
PBRS Anxious-Depressed									
Time 3	BSI OC	BSI PA	PAS NA	POMS TA	PSI PD	STAI-S	STAI-T	PBRS AD	BPRS AD
BSI Anxiety	.83	.83	.69	.83	.73	.67	.70	.04	.41
BSI Obsessive-Compulsive		.75	.69	.73	.72	.66	.70	.04	.39
BSI Phobic Anxiety			.59	.64	.69	.55	.58	.06	.36
PAS Negative Affect				.73	.48	.68	.80	.14	.40
POMS Tension-Anxiety					.58	.74	.79	.16	.40
PSI Physical Symptoms						.48	.48	-.02	.34
STAI-State							.85	.12	.40
STAI-Trait								.14	.49
PBRS Anxious-Depressed									.12
BPRS Anxiety-Depression									
Time 4	BSI OC	BSI PA	PAS NA	POMS TA	PSI PD	STAI-S	STAI-T	PBRS AD	BPRS AD
BSI Anxiety	.81	.78	.63	.82	.69	.69	.72	.20	
BSI Obsessive-Compulsive		.66	.64	.75	.69	.71	.72	.12	
BSI Phobic Anxiety			.52	.59	.58	.54	.63	.15	
PAS Negative Affect				.70	.48	.66	.77	.13	
POMS Tension-Anxiety					.60	.77	.78	.17	
PSI Physical Symptoms						.51	.51	.19	
STAI-State							.84	.15	
STAI-Trait								.18	
PBRS Anxious-Depressed									
Time 5	BSI OC	BSI PA	PAS NA	POMS TA	PSI PD	STAI-S	STAI-T	PBRS AD	BPRS AD
BSI Anxiety	.80	.79	.66	.79	.72	.62	.65	.20	.33
BSI Obsessive-Compulsive		.74	.69	.77	.67	.63	.70	.22	.33
BSI Phobic Anxiety			.62	.60	.58	.51	.58	.19	.26
PAS Negative Affect				.68	.47	.62	.76	.21	.38
POMS Tension-Anxiety					.61	.78	.78	.27	.40

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

PSI Physical Symptoms		.44	.47	.20	.32				
STAI-State			.83	.19	.36				
STAI-Trait				.21	.46				
PBR5 Anxious-Depressed					.26				
BPRS Anxiety-Depression									
Time 6	BSI OC	BSI PA	PAS NA	POMS TA	PSI PD	STAI-S	STAI-T	PBR5 AD	BPRS AD
BSI Anxiety	.77	.74	.62	.81	.74	.57	.64	.12	
BSI Obsessive-Compulsive		.67	.55	.64	.66	.55	.61	.14	
BSI Phobic Anxiety			.49	.52	.62	.37	.49	.05	
PAS Negative Affect				.69	.45	.65	.77	.13	
POMS Tension-Anxiety					.63	.76	.77	.18	
PSI Physical Symptoms						.44	.49	-.01	
STAI-State							.85	.09	
STAI-Trait								.08	
PBR5 Anxious-Depressed									

Note: Only Times 1, 3, and 5 had the BPRS administered. Time 6 includes only the CSP NMI and CSP MI groups.

Cognitive Impairment Construct

Cognitive impairment was assessed by two individually administered tests and ratings by the researcher. The SLUMS was used to assess orientation, memory, attention, and executive function. The TMT was used to assess attention. The time required to complete the Trails A (connect sequential numbers) and B (connect alternating numbers and letters) tasks were collected, and the ratio of times (B/A) was used as the attention measure. Because these measures were not correlated (see descriptions of measures above), we did not combine these scores into a composite measure of cognitive impairment and each was used individually. The correctional staff completed ratings on the PBR5 Dull-Confused scale.

Internal consistency estimates for the SLUMS were provided earlier in the description of the measure. Table B27 provides the correlations between consecutive testing periods and Table B28 has the correlations between the cognitive assessments, including the correctional officer ratings. The correlations between consecutive time periods are moderate in size and indicate stability across 3 month assessment periods. There are some variations in size of coefficients across groups. The validity coefficients are small and indicate that these assessments are likely measuring unique aspects of cognitive function.

Table B27. Test-Retest Correlation Coefficients for Cognitive Impairment Measures at Consecutive Time Intervals by Study Group

Interval	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
SLUMS Scores						
Time 1-2	0.72	0.55	0.64	0.38	0.59	0.63
Time 2-3	0.63	0.74	0.65	0.59	0.78	0.70
Time 3-4	0.75	0.77	0.67	0.70	0.76	0.75
Time 4-5	0.75	0.42	0.67	0.64	0.75	0.67
Time 5-6	0.84	0.67	NA	NA	NA	0.78
Trails B/A Scores						
Time 1-2	0.51	0.32	0.29	0.48	0.40	0.36
Time 2-3	0.20	0.63	0.40	0.32	0.33	0.37
Time 3-4	0.31	0.55	0.58	0.64	0.28	0.44
Time 4-5	0.59	0.33	0.66	0.70	0.30	0.44
Time 5-6	0.47	0.35	NA	NA	NA	0.39

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table B28. Correlations between Cognitive Impairment Measures at each Time Period

Time 1	TMT B/A	PBR5 DC	Time 4	TMT B/A	PBR5 DC
SLUMS	-.26	-.12	SLUMS	-.14	-.10
TMT B/A		.09	TMT B/A		-.01
PBR5 Dull-Confused			PBR5 Dull-Confused		
Time 2	TMT B/A	PBR5 DC	Time 5	TMT B/A	PBR5 DC
SLUMS	-.170	-.081	SLUMS	-.30	-.18
TMT B/A		-.019	TMT B/A		-.05
PBR5 Dull-Confused			PBR5 Dull-Confused		
Time 3	TMT B/A	PBR5 DC	Time 6	TMT B/A	PBR5 DC
SLUMS	-.26	-.11	SLUMS	-.13	-.03
TMT B/A		.04	TMT B/A		-.12
PBR5 Dull-Confused			PBR5 Dull-Confused		

Note: Only times 1, 3, and 5 had the BPRS administered. Time 6 includes only the CSP NMI and CSP MI groups.

Depression/Hopelessness Construct

The depression construct was assessed using five self-report measures. The subscales used to create this composite were the BHS, the BSI Depression subscale, the PAS Negative Affect and Suicidal subscales, and the POMS Depression-Dejection subscale. Table B29 provides estimates of internal consistency reliability for each subscale and the composite at each time period. There is evidence for adequate internal consistency for the composite ($M = .75$; range = .71 to .77). Internal consistency estimates for the subscales were similar to reliabilities found with normative samples with subscales with fewer items demonstrating lower alphas.

Table B29. Internal Consistency Estimates for Depression Construct by Time Interval

Measure	# of items	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6
Depression Composite	5	.74	.76	.76	.77	.76	.71
BHS Total	20	.92	.94	.94	.94	.94	.92
BSI Depression	6	.87	.89	.90	.89	.90	.86
PAS Negative Affect	3	.68	.65	.65	.70	.65	.60
PAS Suicidal Thinking	2	.86	.91	.94	.95	.90	.94
POMS Depression-Dejection	15	.95	.95	.96	.95	.95	.93
PBR5 Anxious-Depressed	14	.90	.90	.94	.90	.90	.90
PBR5 Anxiety-Depression	5	.55	NA	.60	NA	.66	NA

Table B30 provides estimates of test-retest reliability. The correlations between consecutive assessments for the depression composite were strong ($M = .76$, range = .57 to .90) indicating good stability over time. Although there was some variability in estimates across groups and times, there is reasonable stability estimates for each group.

Table B30. Test-Retest Correlation Coefficients for Depression Composite at Consecutive Time Intervals for each Study Group

Interval	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Time 1-2	.77	.69	.57	.79	.65	.59
Time 2-3	.68	.82	.74	.85	.73	.80
Time 3-4	.88	.90	.73	.69	.72	.84
Time 4-5	.77	.74	.82	.71	.77	.83
Time 5-6	.70	.79	NA	NA	NA	.76

Table B31 provides the correlations between the measures of the depression construct. The validity coefficients between assessments of the depression construct indicate good convergent validity with the excep-

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or views expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

tion of the PBRs Anxious Depressed subscale. The magnitude and general pattern of relationships between measures were similar across assessment periods.

Table B31. Correlations between Depression-Hopelessness Measures at each Time Period

Time 1	BSI Dep	PAS NA	PAS ST	POMS DD	PBRs AD	BPRS AD
BHS	.71	.61	.47	.60	.12	.32
BSI Depression		.71	.59	.85	.10	.43
PAS Negative Affect			.46	.65	.12	.35
PAS Suicidal Thinking				.44	-.01	.34
POMS Depression-Dejection					.45	.13
PBRs Anxious Depressed						-.03
BPRS Anxiety Depression						
Time 2	BSI Dep	PAS NA	PAS ST	POMS DD	PBRs AD	BPRS AD
BHS	.70	.54	.45	.61	.10	
BSI Depression		.63	.62	.88	.14	
PAS Negative Affect			.46	.67	.13	
PAS Suicidal Thinking				.54	.18	
POMS Depression-Dejection					.12	
PBRs Anxious Depressed						
Time 3	BSI Dep	PAS NA	PAS ST	POMS DD	PBRs AD	BPRS AD
BHS	.74	.61	.43	.65	-.02	.33
BSI Depression		.69	.60	.88	.07	.37
PAS Negative Affect			.41	.71	.15	.40
PAS Suicidal Thinking				.53	.11	.30
POMS Depression-Dejection					.10	.44
PBRs Anxious Depressed						.12
BPRS Anxiety Depression						
Time 4	BSI Dep	PAS NA	PAS ST	POMS DD	PBRs AD	BPRS AD
BHS	.77	.61	.43	.69	.13	
BSI Depression		.67	.58	.88	.18	
PAS Negative Affect			.42	.66	.13	
PAS Suicidal Thinking				.46	.23	
POMS Depression-Dejection					.18	
PBRs Anxious Depressed						
Time 5	BSI Dep	PAS NA	PAS ST	POMS DD	PBRs AD	BPRS AD
BHS	.75	.61	.39	.66	.19	.27
BSI Depression		.69	.51	.89	.25	.42
PAS Negative Affect			.39	.68	.21	.38
PAS Suicidal Thinking				.48	.29	.39
POMS Depression-Dejection					.27	.42
PBRs Anxious Depressed						.26
BPRS Anxiety Depression						
Time 6	BSI Dep	PAS NA	PAS ST	POMS DD	PBRs AD	BPRS AD
BHS	.58	.46	.45	.45	.14	
BSI Depression		.62	.51	.84	.20	
PAS Negative Affect			.35	.62	.13	
PAS Suicidal Thinking				.42	.41	
POMS Depression-Dejection					.12	
PBRs Anxious Depressed						

Note: Only times 1, 3, and 5 had the BPRS administered. Time 6 includes only the CSP NMI and CSP MI groups.

Hostility/Anger Control Construct

The hostility/anger control construct was assessed using five self-report measures: the BSI Hostility subscale; the Anger Control, Hostile Control, and Acting Out subscales on the PAS; and the POMS Anger-Hostility subscale. Ratings by correctional staff (PBRs Anti-authority) and clinicians (BPRS Hostility) also assess the hostility construct. Table B32 provides the internal consistency reliabilities for the subscales and composite. The composite internal consistency was low ($M = .57$) for the six time periods. Subscale reliabilities were lower than expected for the PAS subscales and the BPRS, although the smaller internal consistency estimates were for scales with a small number of items and these reliability estimates are similar to other literature.

Table B32. Internal Consistency Estimates for Hostility Construct by Time Interval

Measure	# of items	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6
Hostility Composite	5	.54	.56	.57	.61	.55	.60
BSI Hostility	5	.85	.84	.87	.90	.87	.88
PAS Anger Control	2	.53	.60	.51	.56	.52	.33
PAS Hostile Control	2	.52	.47	.42	.37	.36	.61
PAS Acting Out	3	.27	.28	.30	.39	.39	.46
POMS Anger-Hostility	12	.92	.93	.94	.94	.94	.94
PBRs Anti-Authority	13	.94	.93	.94	.94	.95	.90
BPRS Hostility	3	.57	NA	.61	NA	.51	NA

The correlations between consecutive time periods are given in Table B33. These estimates of test-retest reliability indicate that the hostility composite is stable between 3 month assessment periods. Groups are fairly similar in the magnitude of correlation coefficients between testing periods.

Table B33. Test-Retest Correlation Coefficients for Hostility Composite at Consecutive Time Intervals for each Study Group

Interval	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Time 1-2	.73	.75	.83	.78	.69	.71
Time 2-3	.67	.82	.76	.77	.80	.77
Time 3-4	.80	.84	.76	.80	.77	.80
Time 4-5	.76	.78	.56	.79	.77	.76
Time 5-6	.66	.72	NA	NA	NA	.69

Table B34 provides the correlations between measures of the hostility-anger control construct for each time period. The validity coefficients were quite variable across all measures and time periods (ranging between $-.11$ and $.84$). Although scores on these measures tend to be stable, the different assessments may be tapping into quite different aspects of hostility given these variable and lower correlations. Examination of the content of the PAS items suggested that these items tapped into useful domains for understanding hostile and acting out behavior of the participants and thus these measures were kept, even though this leads to lower internal consistency estimates. Removing these items from the composite did increase internal consistency estimates but did not substantially change the study results, thus all subscales were kept as part of the composite (additional results are available from the authors upon request).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or views expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table B34. Correlations between Hostility-Anger Control Measures at each Time Period

Time 1	PAS AO	PAS AC	PAS HC	POMS AH	PBRs AA	BPRs H
BSI Hostility	.20	.54	.17	.69	.06	.27
PAS Acting Out		.18	.14	.10	-.11	.09
PAS Anger Control			.28	.44	.14	.10
PAS Hostile Control				.17	.12	.19
POMS Anger-Hostility					.09	.19
PBRs Anti-Authority						.15
BPRs Hostility						
Time 2	PAS AO	PAS AC	PAS HC	POMS AH	PBRs AA	BPRs H
BSI Hostility	.15	.60	.19	.76	-.01	
PAS Acting Out		.15	.14	.12	.08	
PAS Anger Control			.24	.56	.06	
PAS Hostile Control				.14	.12	
POMS Anger-Hostility					.04	
PBRs Anti-Authority						
Time 3	PAS AO	PAS AC	PAS HC	POMS AH	PBRs AA	BPRs H
BSI Hostility	.15	.45	.15	.50	.02	.36
PAS Acting Out		.15	.15	.16	.06	.05
PAS Anger Control			.24	.48	.11	.29
PAS Hostile Control				.11	.08	.13
POMS Anger-Hostility					.05	.28
PBRs Anti-Authority						.26
BPRs Hostility						
Time 4	PAS AO	PAS AC	PAS HC	POMS AH	PBRs AA	BPRs H
BSI Hostility	.24	.55	.33	.84	.14	
PAS Acting Out		.10	.13	.18	.03	
PAS Anger Control			.33	.51	-.02	
PAS Hostile Control				.27	.09	
POMS Anger-Hostility					.12	
PBRs Anti-Authority						
Time 5	PAS AO	PAS AC	PAS HC	POMS AH	PBRs AA	BPRs H
BSI Hostility	.19	.47	.22	.77	.18	.25
PAS Acting Out		.18	.14	.16	-.02	.04
PAS Anger Control			.32	.46	.15	.20
PAS Hostile Control				.16	.003	.04
POMS Anger-Hostility					.14	.27
PBRs Anti-Authority						.27
BPRs Hostility						
Time 6	PAS AO	PAS AC	PAS HC	POMS AH	PBRs AA	BPRs H
BSI Hostility	.22	.46	.36	.81	.07	
PAS Acting Out		.23	.34	.10	.03	
PAS Anger Control			.39	.47	.08	
PAS Hostile Control				.24	-.06	
POMS Anger-Hostility					.13	
PBRs Anti-Authority						

Note: Only times 1, 3, and 5 had the BPRs administered. Time 6 includes only the CSP NMI and CSP MI groups.

Hypersensitivity Construct

The hypersensitivity construct was measured by two self-report measures—the External Stimulus subscale of the PSI and the Interpersonal Sensitivity subscale of the BSI. This construct is assessing two different aspects of hypersensitivity—environmental and interpersonal. Internal consistency reliabilities for the subs-

cales computed for each assessment period for the entire sample indicate that there is substantial variability in the internal consistency estimates (see Table B35); however, examination of each scale shows that the BSI has strong internal consistency estimates whereas the PSI has low estimates. The PSI was created by the researchers, and its purpose was to capture variables not measured by existing measures, thus it may not be a unidimensional construct. The internal estimates of the composite are lower than might be hoped for and are evidence for the lack of a homogeneous construct.

Although internal consistency estimates were low, the composite demonstrated modest estimates of test-retest reliability (see Table B36) and the correlations between these two subscales provided evidence of convergent validity (see Table B37). Thus these scales were analyzed as a composite for the major analyses completed in the report. Results for analyses done on each individual variable are available from the researchers upon request.

Table B35. Internal Consistency Estimates for Hypersensitivity Construct by Time Interval

Measure	# of items	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6
Hypersensitivity Composite	2	.55	.58	.61	.51	.48	.47
BSI Interpersonal Hypersensitivity	4	.81	.71	.86	.84	.84	.83
PSI External Stimulus	5	.22	.32	.28	.39	.34	.27

Table B36. Test-Retest Correlation Coefficients for Hypersensitivity Composite at Consecutive Time Intervals for each Study Group

Interval	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Time 1-2	.56	.59	.56	.67	.21	.46
Time 2-3	.59	.74	.78	.80	.60	.71
Time 3-4	.71	.71	.63	.65	.57	.70
Time 4-5	.68	.64	.75	.72	.60	.71
Time 5-6	.67	.65	NA	NA	NA	.68

Table B37. Correlations between Hypersensitivity Measures at each Time Period

Time:	1	2	3	4	5	6
BSI IS with PSI ES	.38	.41	.44	.34	.32	.31

Note: Time 6 includes only the CSP NMI and CSP MI groups.

Psychosis Construct

The psychosis construct was assessed by three self-report measures—the Paranoid Ideation and Psychotic subscales of the BSI and the Psychotic Features subscale of the PAS—and clinician ratings (BPRS Thought Disorder). Table B38 provides the Cronbach's alphas for each subscale and the psychosis composite. Internal consistency estimates were good for this composite and its components. Internal consistency estimates for the subscales were similar to those found with normative samples.

Table B38. Internal Consistency Estimates for Psychosis Construct by Time Interval

Measure	# of items	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6
Psychosis Composite	3	.73	.78	.80	.79	.80	.76
BSI Paranoid Ideation	5	.78	.80	.82	.83	.82	.82
BSI Psychoticism	5	.77	.78	.80	.77	.79	.75
PAS Psychotic Features	2	.62	.72	.71	.71	.79	.73
BPRS Thought Disorder	5	.64	NA	.52	NA	.57	NA

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table B39 provides correlations between sequential time periods. Examination of these test-retest reliability estimates indicates good stability between assessment periods. The study groups are similar in magnitude of correlations.

Table B39. Test-Retest Correlation Coefficients for Psychosis Composite at Consecutive Time Intervals for each Study Group

Interval	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Time 1-2	.59	.61	.67	.74	.55	.55
Time 2-3	.52	.83	.75	.87	.73	.75
Time 3-4	.76	.74	.70	.81	.77	.80
Time 4-5	.71	.69	.78	.82	.68	.75
Time 5-6	.64	.79	NA	NA	NA	.71

Table B40 provides estimates of validity coefficients between the measures of the psychosis construct. Correlations indicate reasonable convergent validity for this sample, although correlations are stronger between self-report assessments than between self- and staff-report assessments.

Table B40. Correlations between Psychosis Measures at each Time Period

Time	Measure	BSI Psy	PAS Psy	BPRS TD
1	BSI Paranoid Ideation	.71	.57	.22
	BSI Psychoticism		.35	.27
	PAS Psychotic Features			.32
	BPRS Thought Disorder			
2	BSI Paranoid Ideation	.74	.66	
	BSI Psychoticism		.48	
	PAS Psychotic Features			
3	BSI Paranoid Ideation	.77	.76	.31
	BSI Psychoticism		.56	.30
	PAS Psychotic Features			.25
	BPRS Thought Disorder			
4	BSI Paranoid Ideation	.78	.71	
	BSI Psychoticism		.47	
	PAS Psychotic Features			
5	BSI Paranoid Ideation	.79	.72	.15
	BSI Psychoticism		.53	.13
	PAS Psychotic Features			.14
	BPRS Thought Disorder			
6	BSI Paranoid Ideation	.72	.64	
	BSI Psychoticism		.42	
	PAS Psychotic Features			

Note: Only times 1, 3, and 5 had the BPRS administered. Time 6 includes only the CSP NMI and CSP MI groups.

Somatization Construct

The somatization construct was measured by four self-report assessments, including the Somatization subscale of the BSI, the Health Problems subscale of the PAS, the POMS Fatigue-Inertia subscale, and the Physical Well-Being subscale of the PSI. The mean Cronbach’s alpha across somatization measures and time periods was .79 (see Table B41) and for the composite the mean alpha was .77, indicating adequate internal consistency for this sample. Estimates were similar across time periods.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table B41. Internal Consistency Estimates for Somatization Construct by Time Interval

Measure	# of items	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6
Somatization Composite	3	.78	.79	.78	.78	.78	.73
BSI Somatization	7	.87	.85	.88	.89	.88	.88
PAS Health Problems	2	.56	.65	.65	.59	.65	.59
POMS Fatigue-Inertia	7	.91	.90	.94	.92	.92	.90
PSI Physical Well-Being	8	.72	.76	.73	.75	.76	.74

Table B42 provides test-retest reliability estimates. Correlations between consecutive time periods indicate strong stability across time. Reliability estimates are similar across study groups. Table B43 provides estimates of convergent validity. The correlations between the measures of somatization are reasonable for both self-report assessments and clinician ratings. Correlations show the same basic pattern and magnitude at each time period.

Table B42. Test-Retest Correlation Coefficients for Somatization Composite at Consecutive Time Intervals for each Study Group

Interval	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Time 1-2	.81	.69	.82	.79	.60	.62
Time 2-3	.74	.80	.81	.86	.77	.83
Time 3-4	.80	.74	.74	.76	.81	.84
Time 4-5	.81	.71	.67	.82	.58	.77
Time 5-6	.77	.67	NA	NA	NA	.85

Table B43. Correlations between Somatization Measures at each Time Period

Time	Measure	PSI PE	PAS HP	POMS F	BPRS AD
1	BSI Somatization	.54	.54	.59	.34
	PSI Physical Exercise		.55	.61	.43
	PAS Health Problems			.42	.35
	POMS Fatigue				.40
	BPRS Anxiety Depression				
2	BSI Somatization	.62	.55	.61	
	PSI Physical Exercise		.63	.64	
	PAS Health Problems			.49	
	POMS Fatigue				
3	BSI Somatization	.56	.53	.65	.37
	PSI Physical Exercise		.59	.56	.39
	PAS Health Problems			.46	.44
	POMS Fatigue				.40
4	BSI Somatization	.58	.52	.67	
	PSI Physical Exercise		.62	.60	
	PAS Health Problems			.40	
	POMS Fatigue				
5	BSI Somatization	.56	.54	.65	.27
	PSI Physical Exercise		.57	.62	.44
	PAS Health Problems			.47	.28
	POMS Fatigue				.34
6	BSI Somatization	.44	.40	.64	
	PSI Physical Exercise		.54	.54	
	PAS Health Problems			.38	
	POMS Fatigue				

Note: Only times 1, 3, and 5 had the BPRS administered. Time 6 includes only the CSP NMI and CSP MI groups.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Withdrawal/Alienation Construct

The withdrawal/alienation construct was assessed using two PAS subscales—Alienation and Social Withdrawal—and clinicians’ ratings on the BPRS Withdrawal subscale. Internal consistency reliabilities were computed for each assessment period for the entire sample and are provided in Table B44. The Cronbach’s alphas indicate adequate internal consistency estimates for the self-report measures but are lower for the clinicians’ ratings. Estimates are of similar magnitude across time periods.

Table B44. Internal Consistency Estimates for Withdrawal-Alienation Construct by Time Interval

Measure	# of items	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6
Withdrawal Composite	2	.63	.71	.67	.70	.71	.62
PAS Alienation	2	.79	.74	.75	.72	.72	.72
PAS Social Withdrawal	2	.69	.74	.72	.78	.75	.83
BPRS Withdrawal	6	.47	NA	.49	NA	.40	NA

Table B45 provides the estimates of test-retest reliability for the withdrawal composite. Correlations between sequential time periods were strong indicating good stability. Reliabilities were similar across testing intervals, although there was some variability.

Table B45. Test-Retest Correlation Coefficients for Withdrawal Composite at Consecutive Time Intervals for each Study Group

Interval	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Time 1-2	.65	.67	.55	.64	.87	.60
Time 2-3	.72	.83	.75	.65	.50	.73
Time 3-4	.76	.83	.61	.69	.52	.73
Time 4-5	.63	.81	.60	.49	.69	.71
Time 5-6	.67	.76	NA	NA	NA	.72

The convergent validity estimates are provided in Table B46. The correlations between the self-report assessments (both PAS subscales) indicate strong correlation coefficients; however, the correlations of the self-report with the clinician reports are low. The same pattern is shown across all time periods.

Table B46. Correlations between Withdrawal-Alienation Measures at each Time Period

Time	Measure	PAS-SW	BPRS-W
1	PAS Alienation	.46	.17
	PAS Social Withdrawal		.16
	BPRS Withdrawal		
2	PAS Alienation	.55	
	PAS Social Withdrawal		
3	PAS Alienation	.50	.06
	PAS Social Withdrawal		.14
	BPRS Withdrawal		
4	PAS Alienation	.53	
	PAS Social Withdrawal		
5	PAS Alienation	.54	.18
	PAS Social Withdrawal		.19
	BPRS Withdrawal		
6	PAS Alienation	.45	
	PAS Social Withdrawal		

Note: Only times 1, 3, and 5 had the BPRS administered. Time 6 includes only the CSP NMI and CSP MI groups.

SUMMARY

Using standardized measures allowed us to assess if the scores we obtained in this sample were reliable and valid; we were also able to compare scores from our sample to known values. The self-report assessment variables in general tended to perform similarly across study groups, across time, and to normative samples when examining internal consistency and test-retest reliability. Self-report measures of similar constructs tended to demonstrate convergent validity. In order to combine measures, a composite was developed for each construct of interest, except cognitive impairment, by standardizing scores from the first assessment and computing the mean across measures of the same construct. These composites demonstrated adequate research reliability (internal consistency and test-retest).

This study included self-report and staff report information to allow for convergence using different sources of information. Clinician ratings were gathered using the well-known BPRS. The scores from this measure had a floor effect with scores much lower than normative data (used with clinical populations, including patients who were not in crisis). These scores had lower than expected reliability and validity estimates, although some subscales correlated modestly with self-report measures. There are several possible reasons for this, none of which have been tested in this study—participants were not forthcoming with clinicians, changing of locations may lead to unfamiliarity between participant and clinicians; lack of familiarity with measures by clinicians; clinicians have been desensitized to extreme behaviors in a prison setting so participants seem to be functioning well; clinicians are accurate but participants are exaggerating. Although the floor effect is a concern, because we hypothesized the scores to increase over time on the BPRS, the measure should be able to assess if the mean scores are getting worse over time.

We also used the only measure we could find that allowed for ratings by correctional staff. The only study available on the PBRS as a reference described the development of this measure. The PBRS assessments showed good internal consistency reliability; however, test-retest reliabilities were low and there was little evidence of convergent validity.

Mean scores for each sample were compared to means from normative samples or published research. These comparisons to general adult populations (typically, but sometimes psychiatric populations were used) tended to show that all groups except the general prison participants without mental illness (GP NMI) had statistically significant elevations for the majority of measures across the study assessment periods. This finding suggests that participants entering administrative segregation have significant mental health issues to start, which underscores the importance of assessing individuals over time to explore changes that conditions of confinement might engender.

In summary, the demonstration of good psychometric properties of the data would suggest that responses are not given randomly or haphazardly and that participants are responding in a consistent fashion.

APPENDIX B REFERENCES

- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed). Washington, DC: Author.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). Washington, DC: Author.
- Basso, M. R., Bornstein, R. A., & Lang, J. M. (1999). Practice effects on commonly used measures of executive function across twelve months. *The Clinical Neuropsychologist*, *13*, 283-292. doi: 10.1076/clin.13.3.283.1743
- Beck, A. T. (1986). Hopelessness as a predictor of eventual suicide. *Annals of the New York Academy of Sciences*, *487*, 90-96. doi: 10.1111/j.1749-6632.1986.tb27888.x
- Beck, A. T., Brown, G., Berchick, R. J., Stewart, B. L., & Steer, R. A. (1990). Relationship between hopelessness and ultimate suicide: A replication with psychiatric outpatients. *American Journal of Psychiatry*, *147*, 190-195. Retrieved from <http://ajp.psychiatryonline.org/>
- Beck, A. T., & Steer, R. A. (1993). *Beck Hopelessness Scale: Manual*. San Antonio, TX: The Psychological Corporation: Harcourt Brace & Company.
- Beck, A. T., Steer, R. A., Kovacs, M., & Garrison, B. (1985). Hopelessness and eventual suicide: A 10 year prospective study of patients hospitalized with suicidal ideation. *American Journal of Psychiatry*, *142*, 559-563. Retrieved from <http://www.lexisnexis.com/>
- Beck, A. T., Weissman, A., Lester, D., & Trexler, L. (1974). The measurement of pessimism: The hopelessness scale. *Journal of Consulting and Clinical Psychology*, *42*, 861-865. doi: 10.1037/h0037562
- Biancosino, B., Barbui, C., Pera, V., Osti, M., Rocchi, D., Marmai, L., & Grassi, L. (2004). Patient opinions on the benefits of treatment programs in residential psychiatric care. *The Canadian Journal of Psychiatry*, *49*, 613-619. Retrieved from <http://server03.cpa-apc.org:8080/Publications/Archives/CJP/2004/september/biancosino.pdf>
- Bieling, P. J., Beck, A. T., & Brown, G. K. (2000). The Sociotropy-Autonomy Scale: Structure and implications. *Cognitive Therapy and Research*, *24*, 763-780. doi: 10.1023/A:1005599714224
- Bornstein, R. A., Baker, G. B., & Douglass, A. B. (1987). Short-term retest reliability of the Halstead-Reitan Battery in a normal sample. *The Journal of Nervous and Mental Disease*, *175*, 229-232. doi: 10.1097/00005053-198704000-00007
- Boulet, J., & Boss, M. W. (1991). Reliability and validity of the Brief Symptom Inventory. *Psychological Assessment*, *3*, 433-437. doi: 10.1037/1040-3590.3.3.433
- Briere, J. (1995). *Trauma Symptom Inventory: Professional manual*. Lutz, FL: Psychological Assessment Resources, Inc.
- Brodsky, S. F., & Mason, J. L. (1991). Internal consistency of the Brief Symptom Inventory for counseling-center clients. *Psychological Reports*, *68*, 94. doi: 10.2466/PRO.68.1.94-94

- Brown, G. K., Beck, A. T., Steer, R. A., & Grisham, J. R. (2000). Risk factors for suicide in psychiatric outpatients: A 20-year prospective study. *Journal of Consulting and Clinical Psychology, 68*, 371-377. doi: 10.1037/0022-006X.68.3.371
- Brown, S., Chhina, N., & Dye, S. (2008). The psychiatric intensive care unit: A prospective survey of patient demographics and outcomes at seven English PICUs. *Journal of Psychiatric Intensive Care, 4*, 17-27. doi: 10.1017/S1742646408001295
- Burger, K. G., Calsyn, R. J., Morse, G. A., Klinkenberg, W. D., & Trusty, M. L. (1997). Factor structure of the Expanded Brief Psychiatric Rating Scale. *Journal of Clinical Psychology, 53*, 451-454. doi: 10.1002/(SICI)1097-4679(199708)53:5<451::AID-JCLP5>3.0.CO;2-Q
- Clum, G. A., & Yang, B. (1995). Additional support for the reliability and validity of the Modified Scale for Suicide Ideation. *Psychological Assessment, 7*, 122-125. doi: 10.1037/1040-3590.7.1.122
- Cochran, C. D., & Hale, W. D. (1985). College norms on the Brief Symptom Inventory. *Journal of Clinical Psychology, 41*, 777-779. doi: 10.1002/1097-4679(198511)41:6<777::AID-JCLP2270410609>3.0.CO;2-2
- Cooke, D. J. (1998). The development of the Prison Behavior Rating Scale. *Criminal Justice and Behavior, 25*, 482-506. doi: 10.1177/0093854898025004005
- Coolidge, F. L. (n.d. a). *An introduction to the Coolidge Correctional Inventory (CCI)*. Retrieved from <http://www.uccs.edu/~faculty/fcoolidge/>
- Coolidge, F. L. (n.d. b). *Coolidge Axis Two Inventory-Revised (CATI+): Manual*. Retrieved from <http://www.uccs.edu/~faculty/fcoolidge/>
- Coolidge, F. L. (2004). *Coolidge Correctional Inventory*. Colorado Springs, CO: Author.
- Coolidge, F. L., Segal, D. L., Klebe, K. J., Cahill, B. S., & Whitcomb, J. M. (2009). Psychometric properties of the Coolidge Correctional Inventory in a sample of 3,962 prison inmates. *Behavioral Sciences and the Law, 27*, 713-726. doi: 10.1002/bsl.896
- Cundick, B. P. (1975). Review of the Brief Symptom Inventory. In J. Close Conoley & J. J. Kramer (Eds.), *The tenth mental measurement yearbook* (pp. 111-112). Lincoln, NE: University of Nebraska Press.
- Derogatis, L. (1993). *Brief Symptom Inventory: Administration, scoring, and procedures manual*. Minneapolis, MN: NCS Pearson, Inc.
- Durham, T. W. (1982). Norms, reliability, and item analysis of the Hopelessness Scale in general psychiatric, forensic psychiatric, and college populations. *Journal of Clinical Psychology, 38*, 597-600. doi: 10.1002/1097-4679(198207)38:3<597::AID-JCLP2270380321>3.0.CO;2-6
- Dyce, J. A. (1996). Factor structure of the Beck Hopelessness Scale. *Journal of Clinical Psychology, 52*, 555-558. doi: 10.1002/(SICI)1097-4679(199609)52:5<555::AID-JCLP10>3.0.CO;2-D
- Edens, J. F., Poythress, N. G., & Watkins-Clay, M. M. (2009). Detection of malingering in psychiatric unit and general population prison inmates: A comparison of the PAI, SIMS, and SIRS. *Journal of Personality Assessment, 88*, 33-42. doi: 10.1207/s15327752jpa8801_05

- Fernandez, E. (1998). Review of the Beck Hopelessness Scale [revised]. In J. C. Impara & B. S. Plake (Eds.), *The thirteenth mental measurement yearbook* (pp. 123-125). Lincoln, NE: University of Nebraska Press.
- Fliege, H., Kocalevent, R. D., Walter, O. B., Beck, S., Gratz, K. L., Guitierrez, P. M., & Klapp, B. F. (2006). Three assessment tools for deliberate self-harm and suicide behavior: Evaluation and psychopathological correlates. *Journal of Psychosomatic Research, 61*, 113-121. doi: 10.1016/j.jpsychores.2005.10.006
- Gondolf, E. W. (2008). *Supplemental mental health treatment for batterer program participants* (Document no. 223030). Retrieved from <http://www.ncjrs.gov/pdffiles1/nij/grants/223030.pdf>
- Grassian, S. (1983). Psychopathological effects of solitary confinement. *American Journal of Psychiatry, 140*, 1450-1454. Retrieved from <http://ajp.psychiatryonline.org/index.dtl>
- Gratz, K. L. (2001). Measurement of deliberate self-harm: Preliminary data on the deliberate self-harm inventory. *Journal of Psychopathology and Behavioral Assessment, 23*, 253-263. doi: 10.1023/A:1012779403943
- Gratz, K. L., & Chapman, A. L. (2007). The role of emotional responding and childhood maltreatment in the development and maintenance of deliberate self-harm among male undergraduates. *Psychology of Men & Masculinity, 8*, 1-14. doi: 10.1037/1524-9220.8.1.1
- Gray, R., Bressington, D., Lathlean, J., & Mills, A. (2008). Relationship between adherence, symptoms, treatment attitudes, satisfaction, and side effects in prisoners taking antipsychotic medication: *The Journal of Forensic Psychiatry & Psychology, 19*, 335-351. doi: 10.1080/14789940802113493
- Haney, C. (2003). Mental health issues in long-term solitary and "supermax" confinement. *Crime and Delinquency, 49*, 124-156. doi: 10.1177/0011128702239239
- Harrison, K. S., & Rogers, R. (2007). Axis I screens and suicide risk in jails: A comparative analysis. *Assessment, 14*, 171-180. doi: 10.1177/1073191106296483
- Hayes, J. A. (1997). What does the Brief Symptom Inventory measure in college and university counseling center clients? *Journal of Counseling Psychology, 44*, 360-367. doi: 10.1037/0022-0167.44.4.360
- Heeter, C., Winn, B., Winn, J., & Bozoki, A. (2008). *The challenge of challenge: Avoiding and embracing difficulty in a memory game* (Unpublished manuscript). Retrieved from http://meaningfulplay.msu.edu/proceedings2008/mp2008_paper_31.pdf
- Heinze, M. C., & Purisch, A. D. (2001). Beneath the mask: Use of psychological tests to detect and subtype malingering in criminal defendants. *Journal of Forensic Psychology Practice, 1*, 23-52. doi: 10.1300/J158v01n04_02
- Holden, R. R., Book, A. S., Edwards, M. J., Wasylkiw, L., & Starzyk, K. B. (2003). Experimental faking in self-reported psychopathology: Unidimensional or multidimensional? *Personality and Individual Differences, 35*, 1107-1117. doi: 10.1016/S0191-8869(02)00321-5
- Holden, R. R., & Fekken, C. (1988). Test-retest reliability of the Hopelessness Scale and its items in a university population. *Journal of Clinical Psychology, 44*, 40-43. doi: 10.1002/1097-4679(198801)44:1<40::AID-JCLP2270440108>3.0.CO;2-P

- Joesting, J. (1976). Test-retest reliabilities of state-trait anxiety inventory in an academic setting: replication. *Psychological Reports, 38*, 318. Retrieved from <http://ejournals.ammonsscientific.com/>
- Kellett, S., Beail, N., Newman, D. W., & Frankish, P. (2003). Utility of the Brief Symptom Inventory in the assessment of psychological distress. *Journal of Applied Research in Intellectual Disabilities, 16*, 127-134. doi: 10.1046/j.1468-3148.2003.00152.x
- Kovacs, M., Beck, A. T., & Weissman, A. (1975). Hopelessness: An indicator of suicidal risk. *Suicide, 5*, 98-103.
- Lewis, J. L., Simcox, A. M., & Berry, D. T. R. (2002). Screening for feigned psychiatric symptoms in a forensic sample by using the MMPI-2 and the Structured Inventory of Malingered Symptomatology. *Psychological Assessment, 14*, 170-176. doi: 10.1037/1040-3590.14.2.170
- Long, J. D., Harring, J. R., Brekke, J. S., Test, M.A., & Greenberg, J. (2007). Longitudinal construct validity of Brief Symptom Inventory subscales in schizophrenia. *Psychological Assessment, 19*, 298-308. doi: 10.1037/1040-3590.19.3.298
- Lorr, M., McNair, D. M., Heuchert, J.W. P., & Droppleman, L. F. (2003). POMS: profile of mood states. Toronto, Canada: Multi-Health Systems.
- Matarazzo, J., Wiens, A., Matarazzo, R., & Goldstein, S. (1974). Psychometric and clinical test-retest reliability of the Halstead Impairment Index in a sample of healthy, young, normal men. *Journal of Nervous and Mental Disease, 158*, 37-49. doi: 10.1097/00005053-197401000-00006
- McCaffrey, R. J., Ortega, A., & Haase, R. F. (1993). Effects of repeated neuropsychological assessments. *Archives of Clinical Neuropsychology, 8*, 519-524. doi: 10.1016/0887-6177(93)90052-3
- McNair, D. M., & Heuchert, J. W. P. (2006). *POMS: Profile of Mood States, technical update*. North Tonawanda, NY: Multi-Health Systems Inc.
- McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). *Manual: Profile of Mood States*. San Diego, CA: Educational and Industrial Testing Service.
- McNair, D. M., Lorr, M., & Droppleman, L. F. (1992). *Manual: Profile of Mood States, Revised 1992*. San Diego, CA: Educational and Industrial Testing Service.
- Merckelbach, H., & Smith, G. P. (2003). Diagnostic accuracy of the structured inventory of malingered symptomatology (SIMS) in detecting instructed malingering. *Archives of Clinical Neuropsychology, 18*, 145-152. doi: 10.1093/arclin/18.2.145
- Metzger, R. L. (1976). A reliability and validity study of the State-Trait Anxiety Inventory. *Journal of Clinical Psychology, 32*, 276-278. doi: 10.1002/1097-4679(197604)32:2<276::AID-JCLP2270320215>3.0.CO;2-G
- Morey, L. C. (1991). *Personality Assessment Inventory: Professional manual*. Odessa, FL: Psychological Assessment Resources, Inc.
- Morey, L. C. (1997). *Personality Assessment Screener: Professional manual*. Lutz, FL: Psychological Assessment Resources, Inc.

- Morlan, K. K., & Tan, S. (1998). Comparison of the Brief Psychiatric Rating Scale and the Brief Symptom Inventory. *Journal of Clinical Psychology, 54*, 885-894. doi: 10.1002/(SICI)1097-4679(199811)54:7<885::AID-JCLP3>3.0.CO;2-E
- Nicholls, T. L., Lee, Z., Corrado, R. R., & Ogloff, J. R. P. (2004). Women inmates' mental health needs: Evidence of the validity of the Jail Screening Assessment Tool (JSAT). *International Journal of Forensic Mental Health, 3*, 167-184. Retrieved from <http://www.iafmhs.org/iafmhs.asp>
- Nixon, G. F., & Steffek, J. C. (1977). Reliability of the State-Trait Anxiety Inventory. *Psychological Reports, 40*, 357-358. Retrieved from <http://ejournals.ammonsscientific.com/>
- Norcross, J. C., Guadagnoli, E., & Prochaska, J. O. (1984). Factor structure of the Profile of Mood States (POMS): Two partial replications. *Journal of Clinical Psychology, 40*, 1270-1277. doi: 10.1002/1097-4679(198409)40:5<1270::AID-JCLP2270400526>3.0.CO;2-7
- Novy, D. M., Nelson, D. V., Goodwin, J., & Rowzee, R. D. (1993). Psychometric comparability of the State-Trait Anxiety Inventory for different ethnic subpopulations. *Psychological Assessment, 5*, 343-349. doi: 10.1037/1040-3590.5.3.343
- Nyenhuis, D. L., Yamamoto, C., Luchetta, T., Terrien, A., & Parmentier, A. (1999). Adult and geriatric normative data and validation of the Profile of Mood States. *Journal of Clinical Psychology, 55*, 79-86. doi: 10.1002/(SICI)1097-4679(199901)55:1<79::AID-JCLP8>3.0.CO;2-7
- Palmer, E. J., & Connelly, R. (2005). Depression, hopelessness and suicide ideation among vulnerable prisoners. *Criminal Behaviour and Mental Health, 15*, 164-170. doi: 10.1002/cbm.4
- Periáñez, J. A., Ríos-Lago, M., Rodríguez-Sánchez, J. M., Adrover-Roig, D., Sánchez-Cubillo, I., Crespo-Facorro, B., Quemada, J. I., ... (2007). Trail Making Test in traumatic brain injury, schizophrenia, and normal ageing: Sample comparisons and normative data. *Archives of Clinical Neuropsychology, 22*, 433-447. doi: 10.1016/j.acn.2007.01.022
- Perlick, D. A., Rosenheck, R. A., Clarkin, J. F., Sirey, J., & Raue, P. (1999). Symptoms predicting inpatient service use among patients with bipolar affective disorder. *Psychiatric Services, 50*, 806-812. Retrieved from <http://ps.psychiatryonline.org/index.dtl>
- Piersma, H. L., Reaume, W. M., & Boes, J. L. (1994). The Brief Symptom Inventory (B5I) as an outcome measure for adult psychiatric inpatients. *Journal of Clinical Psychology, 50*, 555-563. doi: 10.1002/1097-4679(199407)50:4<555::AID-JCLP2270500410>3.0.CO;2-G
- Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills, 8*, 271-276. doi: 10.2466/PMS.8.7.271-276
- Sánchez-Cubillo, I., Periáñez, J. A., Adrover-Roig, D., Rodríguez-Sánchez, J. M., Ríos-Lago, M., Tirapu, J., & Barceló, F. (2009). Construct validity of the Trail Making Test: Role of task-switching, working memory, inhibition/inference control, and visuomotor abilities. *Journal of International Neuropsychological Society, 15*, 438-450. doi: 10.1017/S1355617709090626

- Segal, S. P., & Silverman, C. (2002). Determinants of client outcomes in self-help agencies. *Psychiatric Services, 53*, 304-309. doi: 10.1176/appi.ps.53.3.304
- Sherrill-Pattison, S., Donders, J., & Thompson, E. (2000). Influence of demographic variables on neuropsychological test performance after traumatic brain injury. *The Clinical Neuropsychologist, 14*, 496-503. doi: 10.1076/clin.14.4.496.7196
- Spielberger, C., Gorsuch, R., & Lushene, R. (1970). *The State-Trait Anxiety Inventory: Test manual for Form X*. Palo Alto, CA: Consulting Psychologists Press.
- Spielberger, C., Gorsuch, R., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *The State-Trait Anxiety Inventory for adults: Sampler set*. Palo Alto, CA: Consulting Psychologists Press.
- Steed, L. (2001). Further validity and reliability evidence for Beck Hopelessness Scale scores in a nonclinical sample. *Educational and Psychological Measurement, 61*, 303-316. doi: 10.1177/00131640121971121
- Strauss, E., Sherman, E. M. S., & Spreen, O. (Eds.). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed., pp. 3-28). New York, NY: Oxford University Press.
- Tariq, S. H., Tumosa, N., Chibnall, J. T., Perry, M. H., & Morley, J. E. (2006). Comparison of the Saint Louis University Mental Status Examination and the Mini-Mental State Examination for detecting dementia and mild neurocognitive disorder—a pilot study. *American Journal of Geriatric Psychiatry, 14*, 900-910. doi: 10.1097/01.JGP.0000221510.33817.86
- Thomas, A., Donnell, A. J., & Young, T. R. (2004). Factor structure and differential validity of the Expanded Brief Psychiatric Rating Scale. *Assessment, 11*, 177-187. doi: 10.1177/1073191103262893
- Tombaugh, T. N. (2004). Trail Making Test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology, 19*, 203-214. doi: 10.1016/S0887-6177(03)00039-8
- Ventura, J., Lukoff, D., Nuechterlein, K. H., Liberman, R. P., Green, M. F., & Shaner, A. (1993). Brief Psychiatric Rating Scale (BPRS) expanded version (4.0): Scales, anchor points, and administration manual. *International Journal of Methods in Psychiatric Research, 3*, 227-244.
- Whitcomb, J. M. (2006). *Psychometric Properties of the Coolidge Correctional Inventory (CCI) in a sample of 4872 prison inmates* (Unpublished master's thesis). University of Colorado at Colorado Springs, CO.
- Widows, M. R., & Smith, G. P. (2005). *Structured Inventory of Malingered Symptomatology: Professional manual*. Lutz, FL: Psychological Assessment Resources, Inc.
- Zinger, I., Wichmann, C., & Andrews, D. A. (2001). The psychological effects of 60 days in administrative segregation. *Canadian Journal of Criminology, 43*, 47-83. Retrieved from <http://www.cjca-acjp.ca/en/cjc.html>

APPENDIX C

PRISON SYMPTOM INVENTORY ANALYSES

The Prison Symptom Inventory (PSI) was developed by the researchers to assess potential responses to AS confinement that were not covered by other measures. These variables were identified through examination of the professional literature and include nervousness, headaches, lethargy, chronic tiredness, trouble sleeping, a sense of impending breakdown, perspiring hands, heart palpitations, dizziness, nightmares, trembling hands, and fainting. Additionally, items about exercise, grooming, and safety issues within AS were included. The 39-item inventory is given in Appendix A and information about the psychometric properties of the scale are provided in Appendix B.

PSI items were grouped by the researchers into subscales which were thought to measure specific constructs. Three of these subscales (Hypersensitivity to External Stimuli, Panic Disorder, and Physical Well-Being and Exercise) related to constructs assessed by the composites and were included in the composite analyses reported in the main body of the report. In this appendix we provide results from the analyses comparing the study groups on all of the PSI subscales to address the major hypotheses. Table C1 gives a list of the subscales along with items and possible range of scores. Higher scores on the PSI subscales indicate more negative behaviors except on the Attitudes about Segregation subscale where higher scores indicate a preference for AS. Table C2 (a replication of Table B12) provides the summary statistics on each subscale for each study group.

Table C1. PSI Subscales and Range of Possible Scores

Subscale	# Items	Items	Range of Possible Scores
Attitudes about Segregation	2	r14, 39	0 – 10
Fear Level	4	3, 12, 21, r38	0 – 20
Hypersensitivity to External Stimuli	5	1, 7, r31, r34, 37	0 – 25
Mental Well-Being	2	26, r35	0 – 10
Mutism	2	22, r32	0 – 10
Panic Disorder	9	2, 6, 10, 13, 16, 17, 20, 25, 30	0 – 45
Physical Hygiene	5	4, r9, 18, r23, r29	0 – 25
Physical Well-Being and Exercise	8	r5, 8, r11, r15, 19, 24, 27, r28	0 – 40
Safety	2	33, 36	0 – 10

Note. Item numbers with an r indicate that the item is reversed coded.

Table C2. Summary Statistics (*M*, *SD*, *n*) on PSI Scales by Group and Time

Assessment	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Attitudes about Segregation						
1	2.95 (3.19) <i>n</i> = 57	1.68 (2.57) <i>n</i> = 56	2.46 (3.00) <i>n</i> = 28	1.00 (2.10) <i>n</i> = 30	4.58 (3.47) <i>n</i> = 67	2.81 (3.24) <i>n</i> = 238
2	2.97 (3.54) <i>n</i> = 61	1.68 (2.63) <i>n</i> = 56	2.24 (2.76) <i>n</i> = 25	1.54 (2.42) <i>n</i> = 26	5.55 (3.30) <i>n</i> = 60	3.09 (3.42) <i>n</i> = 228
3	3.02 (3.36) <i>n</i> = 55	1.04 (2.39) <i>n</i> = 56	2.04 (2.30) <i>n</i> = 25	1.22 (1.60) <i>n</i> = 27	4.62 (3.45) <i>n</i> = 55	2.58 (3.18) <i>n</i> = 218
4	2.60 (3.21) <i>n</i> = 55	1.29 (2.43) <i>n</i> = 55	1.91 (2.45) <i>n</i> = 22	1.83 (2.58) <i>n</i> = 24	4.96 (3.42) <i>n</i> = 46	2.61 (3.20) <i>n</i> = 202
5	3.12 (3.51) <i>n</i> = 52	1.45 (2.32) <i>n</i> = 55	2.54 (2.67) <i>n</i> = 22	1.30 (1.98) <i>n</i> = 20	5.24 (3.57) <i>n</i> = 45	2.89 (3.32) <i>n</i> = 194
6	2.57 (3.45) <i>n</i> = 49	1.45 (2.32) <i>n</i> = 52	NA	NA	NA	1.92 (2.98) <i>n</i> = 102

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Assessment	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Fear Level						
1	6.25 (4.62) n = 64	4.17 (3.46) n = 63	4.94 (3.78) n = 33	3.51 (3.03) n = 43	7.63 (4.15) n = 67	5.51 (4.18) n = 270
2	5.50 (4.18) n = 62	3.51 (2.54) n = 59	4.53 (3.37) n = 32	3.46 (2.60) n = 41	7.14 (4.96) n = 64	5.01 (4.00) n = 258
3	5.71 (4.10) n = 58	3.91 (2.73) n = 57	4.88 (3.40) n = 32	3.53 (2.07) n = 41	7.66 (4.00) n = 61	5.31 (3.74) n = 250
4	5.22 (3.83) n = 59	3.91 (2.96) n = 56	4.91 (3.08) n = 29	3.26 (2.53) n = 39	6.76 (4.09) n = 59	4.94 (3.63) n = 241
5	5.50 (3.50) n = 58	3.87 (2.75) n = 56	4.79 (3.21) n = 29	3.39 (2.49) n = 38	6.81 (4.30) n = 57	5.00 (3.58) n = 236
6	5.43 (3.43) n = 51	4.28 (2.72) n = 54	NA	NA	NA	4.84 (3.11) n = 106
Hypersensitivity to External Stimuli						
1	10.54 (4.02) n = 64	9.62 (3.92) n = 63	11.00 (5.38) n = 33	8.44 (3.70) n = 43	9.61 (3.94) n = 67	9.82 (4.16) n = 270
2	10.10 (4.52) n = 62	7.81 (3.86) n = 59	11.06 (3.83) n = 32	8.20 (4.09) n = 41	10.11 (3.96) n = 64	9.40 (4.22) n = 258
3	10.71 (4.65) n = 58	8.33 (4.11) n = 57	11.34 (3.95) n = 32	7.76 (4.13) n = 41	9.72 (4.63) n = 61	9.52 (4.50) n = 249
4	9.99 (4.64) n = 58	9.22 (4.49) n = 56	10.15 (3.78) n = 29	8.00 (3.20) n = 39	9.22 (4.49) n = 59	9.55 (4.10) n = 241
5	9.54 (4.09) n = 56	9.03 (4.09) n = 56	10.65 (4.43) n = 29	7.60 (3.62) n = 38	9.03 (4.09) n = 57	9.32 (4.10) n = 236
6	9.37 (4.33) n = 51	9.02 (3.59) n = 54	NA	NA	NA	9.20 (3.93) n = 106
Mental Well-Being						
1	4.95 (2.48) n = 64	4.48 (2.48) n = 63	5.39 (2.54) n = 33	4.00 (2.43) n = 43	5.19 (2.39) n = 67	4.80 (2.48) n = 270
2	4.88 (2.24) n = 61	3.69 (2.55) n = 59	5.09 (2.61) n = 32	3.24 (2.34) n = 41	5.33 (2.53) n = 63	4.48 (2.56) n = 256
3	4.33 (2.42) n = 58	3.98 (2.41) n = 57	4.97 (2.47) n = 32	3.22 (2.31) n = 41	5.44 (2.61) n = 61	4.42 (2.55) n = 249
4	4.71 (2.43) n = 58	3.96 (2.26) n = 56	4.25 (2.78) n = 28	2.77 (1.56) n = 39	5.58 (2.44) n = 59	4.38 (2.48) n = 240
5	4.13 (2.24) n = 55	3.59 (2.25) n = 56	4.66 (2.54) n = 29	2.66 (2.29) n = 38	5.21 (2.24) n = 56	4.08 (2.43) n = 234
6	4.02 (2.01) n = 51	3.68 (2.52) n = 54	NA	NA	NA	3.86 (2.27) n = 106
Mutism						
1	3.67 (2.19) n = 64	2.65 (1.70) n = 63	3.61 (2.07) n = 33	2.40 (1.50) n = 43	3.81 (1.96) n = 67	3.26 (1.98) n = 270
2	4.44 (2.21) n = 62	2.98 (1.97) n = 59	3.16 (1.87) n = 32	2.20 (1.50) n = 41	4.11 (1.72) n = 64	3.51 (2.04) n = 258
3	4.14 (2.29) n = 57	2.98 (1.81) n = 57	3.19 (1.89) n = 32	2.32 (1.56) n = 41	4.44 (2.28) n = 61	3.52 (2.16) n = 248
4	4.41 (2.44) n = 58	2.96 (1.74) n = 56	3.28 (1.74) n = 28	2.59 (1.44) n = 39	3.95 (2.05) n = 57	3.53 (2.06) n = 238
5	4.09 (2.08) n = 55	3.00 (1.80) n = 55	3.21 (1.76) n = 29	2.26 (1.60) n = 38	3.60 (1.94) n = 57	3.31 (1.95) n = 234
6	4.00 (2.19) n = 50	2.75 (1.69) n = 52	NA	NA	NA	3.38 (2.04) n = 103

Assessment	CSP MI	CSP NMI	GP MI	GP NMI	SCCF	All
Panic Disorder						
1	9.09 (10.18) n = 64	3.89 (5.49) n = 63	5.03 (4.88) n = 33	2.77 (3.77) n = 43	10.98 (7.72) n = 67	6.84 (7.84) n = 270
2	7.87 (8.62) n = 62	3.71 (7.72) n = 59	5.43 (5.12) n = 32	2.37 (3.45) n = 41	9.61 (8.70) n = 64	6.18 (7.62) n = 258
3	8.76 (9.17) n = 60	3.46 (4.20) n = 57	6.11 (5.02) n = 32	3.00 (4.68) n = 41	9.55 (8.68) n = 61	6.47 (7.65) n = 251
4	7.32 (7.72) n = 60	3.79 (5.44) n = 56	4.99 (5.06) n = 29	2.22 (3.28) n = 39	9.94 (8.68) n = 59	6.04 (7.19) n = 243
5	6.50 (8.80) n = 56	3.45 (4.34) n = 56	4.08 (5.13) n = 29	1.82 (3.49) n = 38	8.07 (7.45) n = 57	5.10 (6.78) n = 236
6	5.34 (7.25) n = 51	3.60 (5.78) n = 54	NA	NA	NA	4.41 (6.54) n = 106
Physical Hygiene						
1	5.39 (5.01) n = 64	4.00 (4.03) n = 63	4.64 (2.69) n = 33	2.07 (2.54) n = 43	8.26 (4.64) n = 67	5.16 (4.67) n = 270
2	5.80 (4.75) n = 61	4.06 (3.87) n = 59	3.66 (3.26) n = 32	1.74 (2.69) n = 41	7.59 (5.60) n = 64	4.93 (4.77) n = 257
3	5.67 (4.80) n = 58	3.44 (3.59) n = 57	4.19 (4.10) n = 32	2.17 (3.38) n = 41	6.19 (4.55) n = 61	4.52 (4.40) n = 249
4	5.48 (5.43) n = 58	3.25 (3.92) n = 56	3.21 (3.92) n = 29	1.92 (2.67) n = 39	5.61 (4.99) n = 59	4.14 (4.64) n = 241
5	5.73 (5.01) n = 56	3.12 (3.57) n = 56	3.62 (3.70) n = 29	1.08 (1.99) n = 38	5.22 (5.07) n = 57	3.98 (4.46) n = 236
6	4.90 (4.75) n = 51	3.18 (3.68) n = 54	NA	NA	NA	4.12 (4.41) n = 106
Physical Well-Being and Exercise						
1	15.89 (7.76) n = 64	10.43 (5.55) n = 63	15.79 (6.28) n = 33	9.21 (5.61) n = 43	18.93 (6.44) n = 67	14.29 (7.40) n = 270
2	17.10 (7.70) n = 62	9.85 (6.23) n = 59	13.68 (7.37) n = 32	7.26 (5.53) n = 41	18.90 (6.01) n = 64	13.90 (7.91) n = 258
3	17.14 (7.05) n = 58	10.30 (6.18) n = 57	13.84 (6.47) n = 32	8.12 (5.02) n = 41	18.98 (6.78) n = 61	14.12 (7.60) n = 249
4	16.13 (7.47) n = 58	10.28 (6.14) n = 56	13.07 (5.92) n = 29	7.44 (4.36) n = 39	18.46 (6.67) n = 59	13.56 (7.48) n = 241
5	15.39 (7.49) n = 56	9.54 (6.25) n = 56	13.58 (6.78) n = 29	7.08 (4.19) n = 38	17.60 (7.01) n = 57	12.97 (7.57) n = 236
6	13.95 (7.09) n = 51	9.26 (6.72) n = 54	NA	NA	NA	12.97 (7.57) n = 106

The analyses in this section follows the same structure as in the main body of the report: (1) comparisons between the two CSP groups are made on the six time periods; (2) comparisons between the two NMI groups are completed on the five common time periods; and (3) comparisons between the three MI groups are completed on the five common time periods. Mixed design analysis of variance (ANOVA) techniques are used to compare group differences, change over time, and the interaction between groups and time. We were most interested in whether there are significant interactions which would imply that groups are changing differentially over time. The analyses in this section help address goals 2 and 3 of the project—to assess whether offenders with mental illness decompensate differentially in AS compared to offenders without mental illness and to compare psychological functioning of participants AS to relevant comparisons groups.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

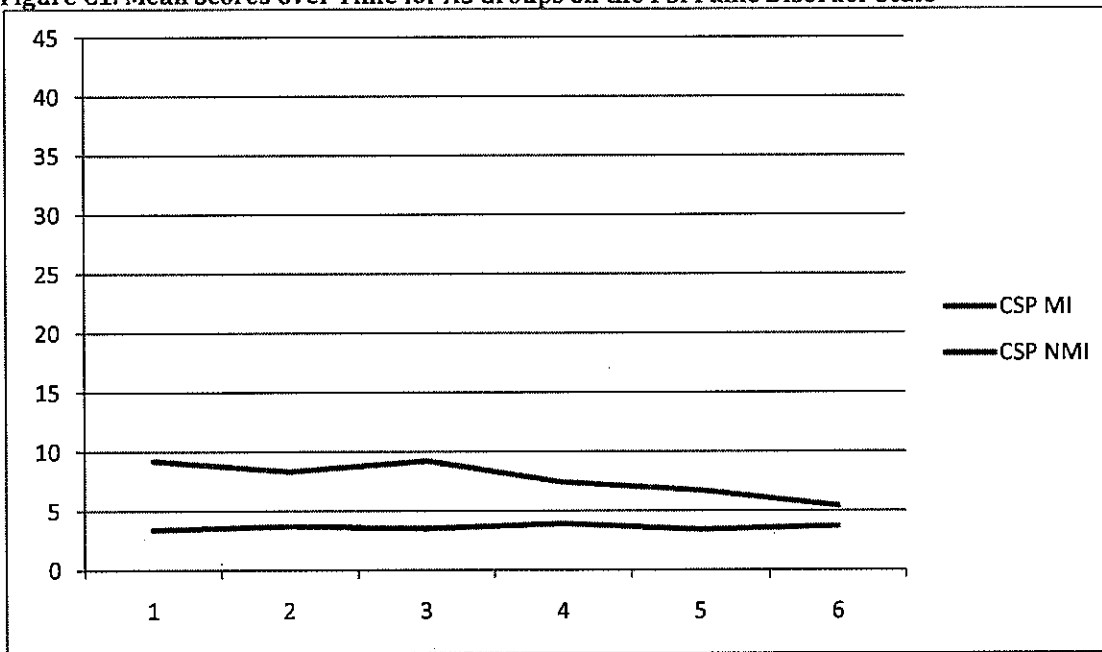
Comparisons between CSP Groups

Table C3 gives the results from the ANOVA analyses comparing mean change over time and mean differences between the two AS groups. For each PSI subscale, the MI group scored significantly higher than the NMI group although the strength of the difference varied over the subscales ($.05 \leq \eta^2 \leq .23$). For the Physical Well-Being and Exercise and Mental Well-Being subscales there were significant time effects, but not significant interaction effects. For both subscales, there was a general decrease over time with scores at the last time period showing significantly lowered mean scores compared to scores at earlier assessment periods. For the Panic Disorder and Hypersensitivity to External Stimuli subscales, there were significant interaction effects. For the Panic Disorder subscale, there was a significant decrease over time for the MI group; however, the NMI group did not change significantly over time. Figure C1 provides a graphical display of this interaction.

Table C3. F Statistics and Partial η^2 Comparing AS Groups across 6 Time Periods

Subscale	Group Main Effect	Time Main Effect	Interaction Effect
Attitudes about Segregation	$F(1, 77) = 3.35, p = .07, \eta^2 = .04$	$F(3.82, 294.12) = 0.63, p = .48, \eta^2 = .01$	$F(3.82, 294.12) = 1.93, p = .11, \eta^2 = .02$
Fear Level	$F(1, 97) = 17.82, p < .001, \eta^2 = .16$	$F(4.36, 422.41) = 0.73, p = .58, \eta^2 = .01$	$F(4.36, 422.41) = 1.54, p = .18, \eta^2 = .02$
Hypersensitivity to External Stimuli	$F(1, 96) = 8.11, p = .005, \eta^2 = .08$	$F(5, 480) = 1.55, p = .17, \eta^2 = .02$	$F(5, 480) = 2.65, p = .02, \eta^2 = .03$
Mental Well-Being	$F(1, 94) = 5.10, p = .03, \eta^2 = .05$	$F(4.70, 441.37) = 3.25, p = .01, \eta^2 = .03$	$F(4.70, 441.37) = 2.44, p = .53, \eta^2 = .01$
Mutism	$F(1, 92) = 17.80, p < .001, \eta^2 = .16$	$F(5, 460) = 1.93, p = .09, \eta^2 = .02$	$F(5, 460) = 0.26, p = .94, \eta^2 = .003$
Panic Disorder	$F(1, 99) = 12.60, p = .001, \eta^2 = .11$	$F(4.00, 396.25) = 2.75, p = .03, \eta^2 = .03$	$F(4.00, 396.25) = 3.10, p = .02, \eta^2 = .03$
Physical Hygiene	$F(1, 95) = 8.76, p = .004, \eta^2 = .08$	$F(5, 475) = 1.84, p = .10, \eta^2 = .02$	$F(5, 475) = 1.46, p = .20, \eta^2 = .02$
Physical Well-Being and Exercise	$F(1, 96) = 27.30, p < .001, \eta^2 = .22$	$F(4.60, 441.67) = 2.45, p = .03, \eta^2 = .02$	$F(4.60, 441.67) = 1.00, p = .42, \eta^2 = .01$

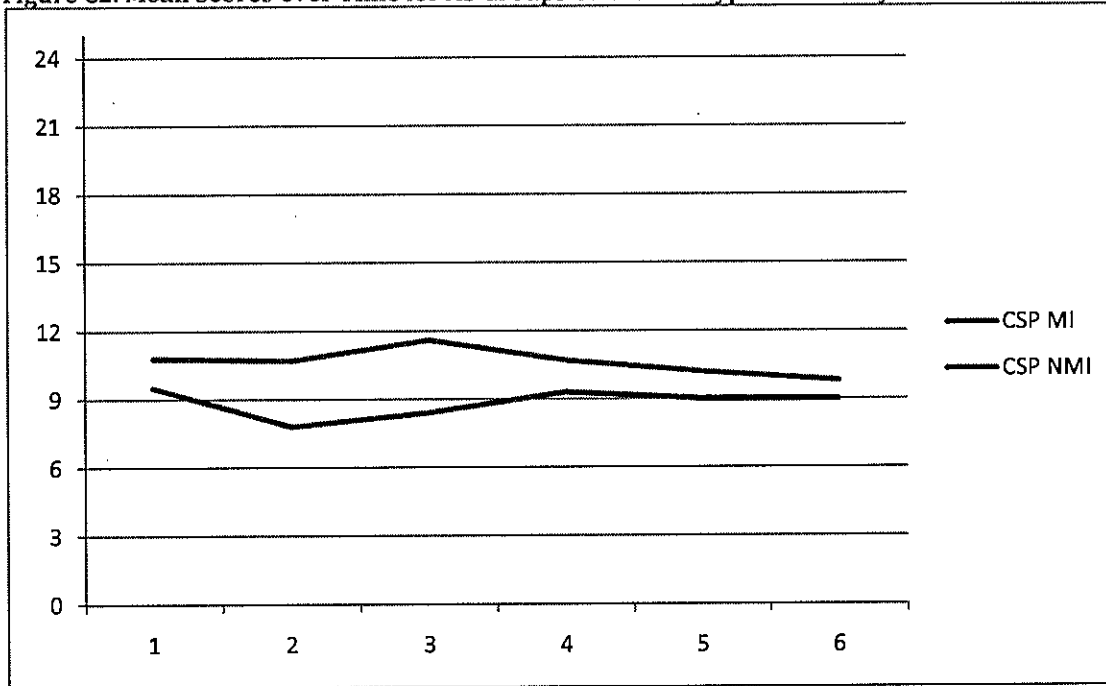
Figure C1. Mean Scores over Time for AS Groups on the PSI Panic Disorder Scale



This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

For the Hypersensitivity to External Stimuli Scale, the interaction is displayed in Figure C2. There was a significant change over time for the NMI group but the MI group did not change significantly over time. For the NMI group, there was a significant decrease in mean scores from time 1 to time 2 but scores were significantly higher at times 4, 5, and 6 than time 2.

Figure C2. Mean Scores over Time for AS Groups on the PSI Hypersensitivity to External Stimuli Scale



Comparisons between NMI groups

Comparisons were made between the CSP NMI and GP NMI groups on the five common time periods. Table C4 provides the results from the mixed design ANOVA analyses comparing mean change over time and mean differences between groups.

Table C4. F Statistics and Partial η^2 Comparing NMI Groups across 5 Time Periods

Subscale	Group/Main Effect	Time Main Effect	Interaction Effect
Attitudes about Segregation	$F(1, 55) = 0.01, p = .93, \eta^2 < .001$	$F(4, 220) = 1.66, p = .16, \eta^2 = .03$	$F(4, 220) = 1.01, p = .40, \eta^2 = .02$
Fear Level	$F(1, 91) = 1.88, p = .17, \eta^2 = .02$	$F(4, 364) = 0.28, p = .89, \eta^2 = .003$	$F(4, 364) = 0.32, p = .85, \eta^2 = .003$
Hypersensitivity to External Stimuli	$F(1, 91) = 1.74, p = .19, \eta^2 = .02$	$F(4, 364) = 2.59, p = .04, \eta^2 = .03$	$F(4, 364) = 1.01, p = .40, \eta^2 = .01$
Mental Well-Being	$F(1, 91) = 4.32, p = .04, \eta^2 = .04$	$F(3.64, 330.90) = 3.97, p = .005, \eta^2 = .04$	$F(3.64, 330.90) = 1.42, p = .23, \eta^2 = .02$
Mutism	$F(1, 90) = 5.76, p = .02, \eta^2 = .06$	$F(4, 360) = 0.70, p = .59, \eta^2 = .01$	$F(4, 360) = 1.24, p = .29, \eta^2 = .01$
Panic Disorder	$F(1, 91) = 4.01, p = .05, \eta^2 = .04$	$F(3.52, 319.86) = 0.28, p = .87, \eta^2 = .003$	$F(3.52, 319.86) = 0.59, p = .65, \eta^2 = .01$
Physical Hygiene	$F(1, 91) = 9.47, p = .003, \eta^2 = .09$	$F(4, 364) = 2.20, p = .07, \eta^2 = .02$	$F(4, 364) = 1.56, p = .18, \eta^2 = .02$
Physical Well-Being and Exercise	$F(1, 91) = 6.36, p = .01, \eta^2 = .06$	$F(3.84, 349.56) = 1.73, p = .14, \eta^2 = .02$	$F(3.84, 349.56) = 0.90, p = .46, \eta^2 = .01$

The CSP NMI group had significantly higher mean scores than the GP NMI group for all PSI subscales except Fear Level, Hypersensitivity to External Stimuli, and Attitudes about Segregation. There were significant main effects of time for Hypersensitivity to External Stimuli and for Mental Well-Being. Follow-up tests for changes in sequential time periods indicated that the first assessment period scores were higher than the

second assessment period scores for both scales with significant time effects. There were no statistically significant interactions between groups and time implying that scores over time were similar in the groups.

Comparisons between MI groups

Comparisons were made between the CSP MI, GP NMI, and SCCF groups on the five common time periods. Table C5 provides the results from the mixed design ANOVA analyses comparing mean change over time and mean differences between groups.

Table C5. F Statistics and Partial η^2 Comparing MI Groups across 5 Time Periods

Subscale	Group Main Effect	Time Main Effect	Interaction Effect
Attitudes about Segregation	$F(2, 91) = 12.56, p < .001, \eta^2 = .22$	$F(3.60, 327.62) = 1.34, p = .26, \eta^2 = .02$	$F(7.20, 327.62) = 1.12, p = .35, \eta^2 = .02$
Fear Level	$F(2, 132) = 6.86, p = .001, \eta^2 = .09$	$F(3.72, 491.72) = 1.27, p = .28, \eta^2 = .01$	$F(7.45, 491.72) = 0.49, p = .86, \eta^2 = .01$
Hypersensitivity to External Stimuli	$F(1, 131) = 0.65, p = .52, \eta^2 = .01$	$F(4, 524) = 0.77, p = .55, \eta^2 = .01$	$F(4, 524) = 0.58, p = .80, \eta^2 = .01$
Mental Well-Being	$F(2, 126) = 3.46, p = .03, \eta^2 = .05$	$F(4, 504) = 2.06, p = .08, \eta^2 = .02$	$F(8, 504) = 1.03, p = .41, \eta^2 = .02$
Mutism	$F(2, 127) = 2.11, p = .12, \eta^2 = .03$	$F(4, 508) = 0.62, p = .65, \eta^2 = .005$	$F(8, 508) = 1.32, p = .23, \eta^2 = .02$
Panic Disorder	$F(2, 135) = 4.65, p = .01, \eta^2 = .06$	$F(3.62, 489.00) = 3.33, p = .01, \eta^2 = .02$	$F(7.24, 489.00) = 0.69, p = .69, \eta^2 = .01$
Physical Hygiene	$F(2, 130) = 4.05, p = .02, \eta^2 = .06$	$F(3.91, 508.46) = 3.13, p = .02, \eta^2 = .02$	$F(7.82, 508.46) = 2.87, p = .004, \eta^2 = .04$
Physical Well-Being and Exercise	$F(2, 131) = 5.73, p = .004, \eta^2 = .08$	$F(3.70, 485.27) = 2.42, p = .05, \eta^2 = .02$	$F(7.41, 485.27) = 1.02, p = .42, \eta^2 = .02$

There were significant group differences on all of the subscales except Hypersensitivity to External Stimuli and Mutism. For the Fear Level and Attitudes toward Segregation subscales, the SCCF group scored significantly higher than the other two groups. For the Panic Disorder, Physical Hygiene, Physical Well-Being and Exercise subscales, the GP MI group scored significantly lower than the other two groups. For the Mental Well-Being subscale, the SCCF group scored significantly higher than the CSP MI group but there were no other significant differences.

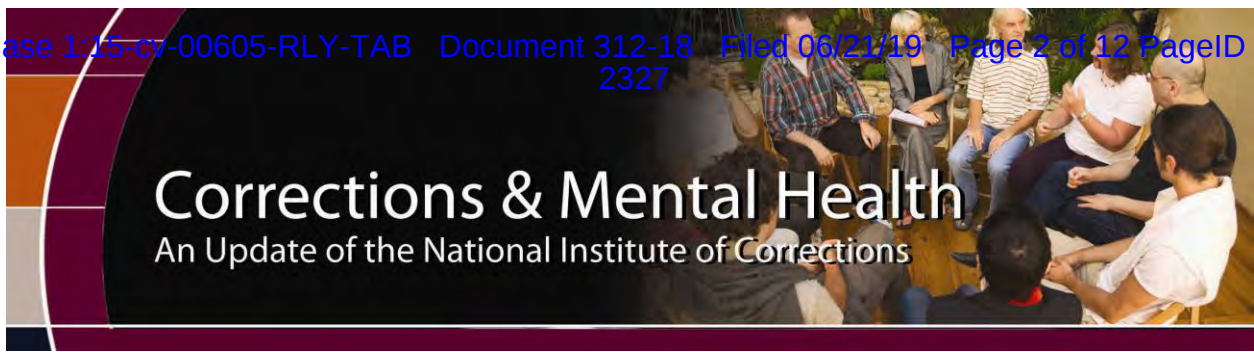
The Panic Disorder and Physical Well-Being and Exercise subscales showed statistically significant changes over time; however for both variables the changes showed improvement over time. Just one of the subscales showed a statistically significant time effect and interaction effect—Physical Hygiene. For this subscale, the SCCF group showed significant decreases across time (i.e., improved hygiene over time) whereas the CSP MI and GP MI groups did not show significant change over time.

SUMMARY

Although there were statistically significant findings, the results did not support the hypotheses of the study. We expected that there would be a worsening over time in reported behavior/sensations and that this change would be worse for inmates with mental illness in AS. However, we found that when significant changes over time occurred, they tended to be in the direction of improvement and this improvement tended to occur more frequently for inmates with mental illness. When making comparisons of the A5 groups to the relevant comparison groups, there was no indication that the segregation groups behavior and attitudes declined over time in comparison to the non-segregated groups.

EXHIBIT 18

Smith, "The Effects of Solitary Confinement"



The effects of solitary confinement: Commentary on One Year Longitudinal Study of the Psychological Effects of Administrative Segregation

By Peter Scharff Smith

Abstract

Solitary confinement is a common practice in many prisons, but it has sparked debates and research on its effects on prisoners. This article examines a recent study on administrative segregation in Colorado in the context of relevant European research on the effects of solitary confinement

Key words: administrative segregation, solitary confinement

The use of large scale solitary confinement became common with the rise of the modern penitentiary during the first half of the 19th century and has remained a feature of Western prison systems. A debate about the effects of solitary confinement was largely settled early in the 20th century, when both experts and practitioners tended to agree that solitary confinement was harmful. Discussions on the effects of solitary confinement resurfaced in the 1950s and the following two decades when sensory deprivation studies were carried out in reaction to, among other things, stories of the brainwashing of U.S. prisoners of war during the Korean War. During the 1980s, solitary confinement regained topicality in the wake of the creation of supermax prisons in the United States. But solitary confinement has also been used, debated, and researched extensively elsewhere. As one example, solitary confinement has been an integral part of Scandinavian pre-trial prison practice for many years (Smith 2006). In 2010, the Colorado Department of Corrections and the Department of Psychology at the University of Colorado issued a new study on solitary confinement. In this article, I will discuss research on the effects of solitary confinement and make some comments on the Colorado study. (*Editor's note: All references to, or quotes from, the Colorado study are from O'Keefe et al., 2010.*)

Colorado Study

The Colorado study is longitudinal and mainly based on self-reported data. The battery of tests used in this study looks impressive and covers the various symptoms and health issues described in the earlier solitary confinement literature, such as anxiety, depression, and suicidal thinking. However, it is clearly important that all these tests were used without in-depth interviews being conducted, and that the self-reported data was not collected by a psychiatrist, a psychologist, or an experienced prison researcher. The main conclusion in the Colorado study was that the results “were largely inconsistent with (...) the bulk of literature that indicates AS is extremely detrimental to inmates.” and that “there was initial improvement in psychological well-being across all study groups, with the bulk of the improvements occurring between the first and second testing periods.” However “all of the study groups, with the exception of the GP NMI (general population, non-mentally ill) group, showed symptoms that were associated with the SHU (special housing unit) syndrome” (i.e. high degrees of psychological disturbance). In this article, I will discuss a number of issues that will help explain the apparent discrepancy between the Colorado study conclusions and the results gathered in other available research.

Why not use the available research?

The Colorado report begins with the claims that the debate on the use of long-term administrative segregation “has suffered from a lack of empirical research” and that “the scant empirical research conducted to date suffers from research bias and serious methodological flaws.” This is a seriously misleading statement. The problem is not that relevant and rigorous empirical research does not exist, but that the authors of the Colorado report haven’t used it. Much of this research is European, but it has been presented and reviewed in international journals, including U.S.-based journals (Smith, 2006 and Haney, 2009).

European studies on the effects of solitary confinement

A growing body of American research is clearly relevant to a discussion of solitary confinement and segregation regimes (see, for example, Lovell, 2008; Cloyes, Lovell, Allen, and Rhodes, 2006; Rhodes, 2004; and Haney, 2008). In the following, I will briefly review some of the European research, which seems to be less known to American readers. This research has not been carried out in supermax prisons in the U.S. (for European supermax research, see King, 2005 and Shalev, 2009) but it is, in fact, research on how prisoners react to being subjected to 22-23 hours of solitary confinement in their cell each day, so it is most certainly relevant. According to the Colorado study, the “defining feature” of administrative segregation in Colorado is single-cell confinement for 23 hours per day.

For various reasons, the use of pre-trial solitary confinement has historically been extensive in Sweden, Norway, and Denmark and has sparked intense debates and also research on the effects of solitary confinement, especially in Denmark and Norway (Smith 2006). In Norway, a 1993 longitudinal study of 63 isolated remand prisoners found widespread health problems after four weeks of solitary confinement, including depression, anxiety, stomach and muscle pains, and an inability to concentrate. The study excluded inmates with obvious withdrawal symptoms and those deemed at risk of suffering from a psychosis (Gamman 2001). A longitudinal follow-up in 1995 with a sample of 54 remand prisoners

included a control group and reported significantly more physical and psychological suffering, including sleeplessness, concentration problems, anxiety, and depression, among the prisoners in solitary confinement, who were also given much more medication than the control group (Gamman, 1995, 2001). The author of this study found that several of the isolated prisoners developed symptoms of a hallucinatory nature, that there were “important differences” between the health of those isolated and those not, and concluded that “the isolated had more symptoms of both psychological and somatic nature” (Gamman, 1995, p. 2245).

In terms of the prevalence of symptoms, 94 percent of those in pre-trial solitary confinement suffered from adverse symptoms after four weeks. More than half suffered from serious symptoms like depression and anxiety, and 13 percent had mutilated themselves (Gamman, 2001). In a third Norwegian study on disciplinary segregation, more than 43 percent of the isolated prisoners suffered adverse symptoms after only an average of 39.7 hours in solitary confinement (Stang et al., 2003).

In Denmark during the 1980's and 90's, extensive research on the effects of solitary confinement was carried out in the form of a number of interview-based studies as well as a so-called “isolation-study,” which was a large-scale longitudinal study consisting of a comprehensive psychiatric and psychological study (1994) and a follow-up study (1997), both with control groups. The Colorado report authors are not aware of some of the most important articles and results from these studies (Sestoft et al., 1998; Andersen, 2004; see also Smith, 2006), and furthermore do not fully incorporate the findings of the two related studies they actually list in their references. The Danish 1994 study involved 367 remand prisoners and reported a significantly higher rate of psychiatric problems among prisoners in isolation. A higher incidence of psychiatric morbidity – mainly adjustment disorders - was found among those in solitary confinement (28 percent) compared to those not in isolation (15 percent). The rate of psychiatric morbidity was highest (43 percent) among a third group of remand prisoners who had been in solitary confinement for more than two months (Andersen et al., 1994). A number of standardized instruments were used to measure health quantitatively. The scores for those in solitary (as a group) were unchanged throughout the isolation period, while those not in isolation “had a gradual improvement on most quantitative mental health scores during this early phase of imprisonment (Andersen, 2004, p. 39)” Those in solitary confinement experienced an improvement in health scores when the solitary confinement conditions were relieved (Andersen 2004). The researchers concluded that the differences between the isolated remand prisoners and the control group were caused “mainly by different conditions of SC and non-SC” (Andersen 2004, p. 39), and that pre-trial detention in isolation compared with pre-trial detention without isolation involved strain and risk of damaging the mental health of the imprisoned individuals (Andersen et al. 1994, 2000).

The 1994 study was longitudinal, incorporated both quantitative and qualitative elements, used standardized instruments to measure health, incorporated in-depth interviews, used highly-skilled researchers, included control groups and a very large number of prisoners in solitary confinement, produced statistically significant results, and verified their results through other objective data regarding the hospitalization of remand prisoners.

Still, the thoroughness of the study caused the research itself to constitute a significant intrusion into the lives of the study's participants (Andersen, 2004). During the first three weeks of imprisonment those in solitary confinement were typically subjected to four or five days of intense interviews and testing (2–4

hours each day, not counting filling out questionnaires, having blood samples taken etc.). These remand prisoners were, in other words, effectively *not* in solitary confinement during those four or five days. This constituted around 20 to 25 percent of the period between the first test and the end of the second test round after approximately three weeks. This must have downgraded the measured differences between the isolated prisoners and the control group significantly, especially since the interviews constituted meaningful social contact in which the well-being and innermost thoughts of the imprisoned individual was in focus (Smith, 2006).

Given this issue, it is not surprising that the second part of the 1994 study - a survey of hospitalization among remand prisoners – gave even more clear-cut results. A sample of 124 remand prisoners who had been transferred to prison hospital revealed that, if “a person remained in SC [solitary confinement] for four weeks the likelihood of being admitted to the prison hospital for a psychiatric reason was about twenty times as high as for a person remanded in NSC [non-solitary confinement] for the same period of time” (Sestoft et al., 1998, p. 103).

A 1997 follow-up study was based on reports (questionnaires) from former participants in the original study, and illustrated how former remand prisoners in solitary confinement found their incarceration significantly more straining than did remand prisoners not in isolation. Thirty-eight percent of those in solitary confinement and 36 percent of those in long-term solitary found their remand imprisonment extraordinarily straining, as opposed to 12 percent of those not in solitary (Andersen et al., 1997). Furthermore, 23 percent of those in solitary confinement and 27 percent of those in long-term solitary reported that they experienced severe psychological reactions after their remand imprisonment, as opposed to nine percent of those not in solitary (Andersen et al., 1997). The authors concluded that from a medical and psychological perspective the practice of pre-trial solitary confinement should be abandoned (Andersen et al., 1997).

A Swiss study on the effects of solitary confinement documented a similar problem surrounding hospitalization of inmates in solitary confinement. The study sample consisted of 203 male patients in a psychiatric clinic in Zurich, of whom 102 were committed from a prison (76 percent of these came directly from solitary confinement). The study concluded that remand prisoners in solitary confinement were much more often hospitalized for psychiatric reasons than were prisoners who came from communal prison conditions (Volkart, Rothenfluth, et al., 1983).

Volkart and colleagues also compared 30 prisoners in solitary confinement with a control group of 28 prisoners in communal imprisonment. The study was cross-sectional and incorporated no longitudinal data. Isolated inmates had spent an average of ninety-one days in solitary confinement while the control group had spent on average 326 days imprisoned. All participants had normal intelligence and their health and personalities were assessed through psychiatric questionnaires. The group of isolated inmates “showed considerably more psychopathological symptoms than the control group [and these] effects were mainly caused by solitary confinement; age, schooling, duration of detention and personality turned out to be of subordinate importance.” (Volkart, Dittrich, et al. 1983, p. 44)

Social contact and contamination across groups

The available research, including the above-mentioned studies, demonstrates that solitary confinement “causes serious health problems for a significant number of inmates. The central harmful feature is that it reduces meaningful social contact to an absolute minimum: a level of social and psychological stimulus that many individuals will experience as insufficient to remain reasonably healthy and relatively well-functioning.” (Smith, 2006, p.503)

This should be a starting point for further research on solitary confinement. Previous research does not show, for example, that the availability of television, radio, or newspapers, or even good material conditions of confinement, will offset the negative impact of solitary confinement on many prisoners, although access to such items and conditions can ameliorate any prison experience to a certain extent. But as the Colorado report concludes, the availability of modern technology, such as videoconferencing, is not always positive for the prisoners since “it also increases the degree of isolation experienced by inmates.”

Therefore, it is unfortunate that the Colorado study does not explore this issue convincingly, i.e., measuring the relative level of psychologically meaningful social contact in administrative segregation (AS), punitive segregation, and general population (GP). If we look closer at the Colorado study it describes basic AS conditions as single-cell confinement for around 23 hours per day. In AS, prisoners are given five 1-hour recreation spells each week, as well as three 15-minute showers (although apparently inmates use less time for showers). Prisoners are escorted to recreation in “full-restraints.” Depending on custody level, inmates are allowed either two 2-hour noncontact visits per month (Level 2) or four 3-hour visits per month (Level 3). Phone calls for those in the Colorado State Penitentiary apparently amounted to only a few minutes daily. If we look at both recreation, visits, and showers, an inmate on level 2 will apparently (assuming he receives visitors) stay at least around 23 hours in his cell on a daily basis, while those on level 3 get two more hours out of their cell on a weekly basis (once again assuming that they receive visits) – i.e. less than 20 minutes less cell time on a daily basis.

In addition to the above, there is some contact with mental health clinicians who do monthly rounds and occasional “mental health sessions” for one to two hours per week. Furthermore inmates in AS go through a “Quality of Life Program,” which includes cognitive classes, but as far as I can see this does not result in increased social contact since these classes, along with some recreational activities, take place over the television.

Punitive segregation, where many inmates stayed prior to AS, is single-cell confinement for 23-24 hours per day, during which inmates only come out for recreation and showers in the living unit. So most inmates stay inside the segregation unit during their entire stay and are “placed in full-restraints” if escorted out of the cell. Inmates in punitive segregation are not allowed to work or participate in any programs or education, and do not have a television.

Descriptions of these conditions indicate that the amount of psychologically meaningful social contact is extremely scarce in both AS and punitive segregation, with the latter regime apparently allowing even less out-of-cell time and social contact. There is, however, one unclear factor. According to the Colorado

report, the inmates in AS can communicate with sign language and they can also yell to each other. Exactly how much and what kind of contact this results in is not described. Furthermore, GP conditions are not described along with the amount of social contact allowed under that regime.

Basically, it is somewhat unclear in the Colorado study how much meaningful social contact inmates in AS had access to during the study. AS conditions suggest that they had very limited access to such contact, although it is not entirely clear what level of communication was allowed through yelling and sign language, where especially the former might potentially yield some level of meaningful contact. Furthermore, it is unclear how much staff contact inmates have, although it is seemingly not a lot.

To confuse matters even more, there was “contamination across groups” meaning that “all offenders in AS were not confined in segregation for their entire period of participation in the study” and inmates in GP may “at some time during their study participation [have] been placed in punitive segregation or even AS.” In fact, when looking at “pure cases” of continuous AS, there were only 26 among the mentally ill and 39 among the non mentally ill, and even more alarming, there were only 13 “pure cases” of continuous GP prison time among the mentally ill GP control group (GP-MI) and 11 “pure cases” of continuous GP prison time among the non mentally ill GP control group (GP-NMI). This means that out of the 33 GP-MI and 43 GP-NMI who participated in the study (some of which later dropped out) only 13 GP-MI and 11 GP-NMI spent their entire study time in GP conditions. So the GP control group was not really a GP control group at all since the majority of these experienced either AS or punitive segregation during their participation in the study, and in addition most – perhaps all – experienced AS immediately prior to their AS hearing, after which they went into GP.

The Colorado researchers looked at their “pure cases” and found no major differences between these and other GP inmates. Then, they disregarded the problem, although such a finding questions the validity of their self-reported data and the setup of the entire study. Under all circumstances, the Colorado study is in fact *not* a study comparing segregation/solitary confinement with non-segregation/solitary confinement, since most of the GP inmates experienced solitary confinement during the study.

Equally important are uncertainties surrounding the levels of meaningful contact the study participants had prior to the start of the study. It is unclear how many participants came from solitary confinement when they entered AS or how much time they spent under such conditions before their initial tests. If some came directly from GP conditions to AS, then it is a problem that we do not know what that means in terms of a change in the level of available, meaningful social contact. We do know that some inmates – although not how many - came directly from punitive segregation and given the way these conditions are described in the Colorado study it seems likely that these inmates experienced better conditions with more meaningful contact when they entered AS. In that case, it is hardly surprising that the study found positive developments between the first and second testing of the inmates.

Were the study participants harmed by solitary confinement prior to the study?

The mental health of the Colorado inmates when they entered AS is very important, as are the conditions they arrived from prior to the start of the study. Needless to say, it puts the Colorado study in different light if many participants were actually in segregation prior to the start of the study. Unfortunately, the Colorado study is somewhat unclear about this.

The Colorado report states that “all study participants classified to AS were waitlisted for and placed in CSP,” which as far as I understand means that they were living in AS conditions when waiting for their AS hearing. The introduction to the report says something slightly different, however, when it states that “in the time leading up to and during their AS hearing, inmates have typically been in segregation.” So some prisoners were apparently not in segregation? The Colorado authors “recognized that significant changes could occur while inmates were held in segregation at their originating facility.” Therefore, they collected a pre-baseline measure “as close to the AS hearing as possible.”

In order to use the study to discuss the effects of solitary confinement, we need to know exactly how many were in segregation prior to the study and, even more importantly, we need to know for how long those subjected to a pre-baseline measure had been in segregation before they were subjected to the pre-baseline measure. This information is crucial and seems lacking in the report. All we are told is that pre-baseline measures were collected “as close to the AS hearing as possible.” But what does this mean in practice? The question, of course, involves the extent to which participants were possibly affected by solitary confinement prior to the start of the study. This is important since we know from other research that reactions to solitary confinement vary from one individual to another, but they “often set in very quickly.” (Thelle & Traeholt, 2003, p.769)

The Colorado report concludes that “all of the study groups, with the exception of the GP-NMI group, showed symptoms that were associated with the SHU syndrome. These elevations were present from the start and were more serious for the mentally ill than non-mentally ill.” So if many study participants had been subjected to segregation prior to the study that would likely explain their symptoms. In other words, the study participants were already damaged by solitary confinement when the study began, and the Colorado study shows us that these prisoners continued to show “symptoms that were associated with the SHU syndrome” during their time in AS.

Furthermore, positive developments between the first and second test could be explained by the transfer from punitive segregation conditions to apparently better AS conditions, which include a more meaningful form of social contact (visits). Seen in this light, the results of the Colorado study are in line with previous research. The AS inmates in Colorado got slightly better when they had access to slightly more meaningful social contact, but they remained in a very bad condition, and continued to show symptoms, as they stayed in solitary confinement.

How was the self-reported data obtained?

According to the Colorado study, all the self-reported data were collected by one field researcher who was a female university employee with CDOC training and badge that allowed her unescorted access to the prison facilities. The field researcher had an undergraduate degree and is not the responsible author. This is a very big difference in contrast to Danish and Norwegian studies, where the actual researchers who designed the studies and wrote the reports were trained psychiatrists and psychologists and also operated as field researchers. They accessed the health of the study participants themselves and did the in-depth interviews. In my opinion, this is the only serious and professional way to design and conduct a study about health in prison, which includes obtaining data directly from prisoners. Sending a “researcher” who is neither a health practitioner nor a PhD-level researcher with experience doing prison research, into a prison in order to access the health of prisoners by collecting self-reported data simply means that the

data are likely to be unreliable. That the field researcher had to report to an employee of the prison system studied (the leading author of the report) is also problematic.

The Colorado report itself describes instances in which the self-reported data appeared questionable. When this occurred, the field researcher apparently asked prisoners to retake the test if they admitted to “not being truthful.” If study participants said they were being honest and the researcher still did not believe them, she “marked the test as questionable.” This validation process seems outright naive. On what grounds did the university’s inexperienced field researcher assess whether or not the prisoners were “being truthful” about their psychological problems and mental health? This obviously requires education, experience, and psychological or medical knowledge. Seen in this light, it is interesting to note that when the Colorado study authors removed persons “with questionable or inconsistent responses” it “did not change the overall effects and results” so they used all the responses in their analysis. This raises serious questions about the field researcher’s capacity to assess whether or not the prisoners were ‘truthful’ and, once again, raises questions about the reliability of all of the self-reported data.

Professional researchers report that it can be difficult to learn about symptoms suffered by isolated inmates since many (male prisoners in particular) try to hide their condition (Smith, 2006). Researchers also explain that it is often extremely difficult, traumatic, and painful for formerly isolated individuals to talk about their experience of solitary confinement: “A few studies seem to explain the fact that some inmates do not complain and seem to adapt more or less peacefully to solitary confinement as a sign of a healthy coping strategy, while others explain this as an unhealthy sign of social withdrawal typically accompanied by severe psychological problems. Such problems often will be discovered only by personal in-depth interviews in a positive (therapeutic) atmosphere.” (Smith, 2006, p. 474; see also Koch, 1982; Toch, 1992; Jackson, 1983)

King, who has interviewed many supermax prisoners, observes that a significant number of these prisoners “found it extremely difficult to bring themselves to talk about their experience” and only after “considerable persistence some prisoners came to regard a researcher from another culture, who treated them with respect and clearly wanted to learn, as an acceptable proxy and began to open up.” (King, 2005, p.130)

Furthermore, the study authors made a mistake by advising inmates that “the purpose of the study was to learn about their adjustment to prison.” It is well known that within a prison community it is important for prisoners to seem capable of adjusting to prison, and those who do not manage to do this are typically placed at the bottom of the prison hierarchy. Approaching study participants with an overall question regarding “their adjustment to prison” in other words makes it likely that they will try to hide possible weaknesses and try to convey the impression that they cope and adjust relatively well. In a prison context, it is not an “open” but a “leading” question.

Crisis events, hospitalization, and objective data

The Colorado researchers describe initial attempts to include “crisis events” such as self-mutilation or suicide attempts recorded by prison clinicians in their study, but they decided not to, because the number of participants who experienced these events allegedly was too small and because crisis events could occur without staff’s knowledge. The authors conclude that the available data “raise more questions than they provide answers.” If we look carefully at these data, however, they certainly raise some questions.

If we compare the number of crisis events among the mentally ill in GP and in AS, we find that throughout the study two persons had two crisis events in the former group, while 10 persons had 26 crisis events in the latter group (one suicide attempt, 14 cases of suicidal/self harm ideation, and 11 cases of self harming behavior). This seems a significant difference with respect to important behaviors that have been identified in past research as among the adverse effects of solitary confinement. The numbers are small, but, still, five times as many prisoners in the AS-MI group had crisis events compared to the GP-MI group, and 13 times as many crisis events occurred in the AS-MI group compared to the GP-MI group. Furthermore, 11 crisis events in the AS-MI group were associated with psychotic symptoms compared to one such crisis event in the GP-MI group.

These data are important in two ways. They suggest that solitary confinement had a negative impact on the health of the mentally ill, but also, even more importantly, they seriously question the reliability of the study’s self-reported data. These crisis event data raise questions about why the difference among the AS-MI and GP-MI groups was not found through the self-reported data. After all, a significant number of participants in the AS-MI group had crisis events and the prevalence of these events were much higher than in the GP-MI group. Furthermore, such crisis events would normally be considered “the tip of the iceberg.” A likely hypothesis would be that a prison environment producing significantly more self-harm and suicidal thoughts than other prison regimes would also reveal many more “lesser” psychological problems. One cannot help asking how and why the Colorado researchers chose to ignore this data, which in fact questions the entire setup of their study?

Conclusion

The Colorado study suffers from several major problems. First, some of the most relevant research available was not used and it was wrongfully claimed that previous research was biased and flawed. Secondly, the way the self-reported data was collected very likely made these data unreliable. Thirdly, the study authors ignored that their crisis data seriously questioned the validity of their self-reported data and in fact suggested that AS might have serious ill effects. Fourth, the majority of the study participants apparently came directly from segregation, and were thus likely to be harmed from solitary confinement before the study started. Finally, the Colorado study in fact did not compare segregation/solitary confinement with non-segregation/solitary confinement since most of the GP participants also went into solitary confinement during the study. Imagine a similar situation with, for example, medical research on the effects of a new type of medicine where it turns out that most of the control group participants also received the new medicine being tested both during the study and prior to study start. It does not make sense. It is therefore extremely difficult to gain any valuable information about the effects of AS and solitary confinement from the Colorado Study.

Author note: Peter Scharff Smith, Ph.D., has studied the history, practice, and effects of solitary confinement for more than a decade. Currently a Senior Research Fellow at the Danish Institute of Human Rights, his published articles on solitary confinement have appeared in, for example, **Crime and Justice**, **Criminal Justice & Behavior**, **Torture**, and **Punishment & Society**. Email: pss@humanrights.dk.

References

- Andersen, H.S. 2004. "Mental Health in Prison Populations: A Review – with Special Emphasis on a Study of Danish Prisoners on Remand." *Acta Psychiatrica Scandinavica Supplementum* 110(424):5-59.
- Andersen, Henrik Steen, Dorte Sestoft, Tommy Lillebæk, Gorm Gabrielsen, and R. Hemmingsen. 2003. "A Longitudinal Study of Prisoners on Remand: Repeated Measures of Psychopathology in the Initial Phase of Solitary versus Nonsolitary Confinement." *International Journal of Law and Psychiatry*, 26, 165–77.
- Andersen, Henrik Steen, Dorte Sestoft, Tommy Lillebæk, Gorm Gabrielsen, R. Hemmingsen, and P. Kramp. 2000. "A Longitudinal Study of Prisoners on Remand: Psychiatric Prevalence, Incidence and Psychopathology in Solitary versus Non-solitary Confinement." *Acta Psychiatrica Scandinavica*, 102(1):19–25.
- Andersen, Henrik Steen, Tommy Lillebæk, and Dorte Sestoft. 1997. *Efterundersøgelsen - en opfølgingsundersøgelse af danske varetoegtsarrestanter*. Copenhagen: Schultz.
- Andersen, Henrik Steen, Tommy Lillebæk, Dorte Sestoft, and Gorm Gabrielsen. 1994. *Isolationsundersøgelsen: Varetoegtsfængsling og psykisk helbred*. Vols. 1–2. Copenhagen: Schultz.
- Cloyes, K., Lovell, D., Allen, D., and Rhodes, L. 2006. "Assessment of Psychosocial Impairment in a Supermaximum Security Unit Sample." *Criminal Justice and Behavior*, 33, 760.
- Gamman, Tor. 2001. "Om bruk av isolasjon under varetektsfengsling." *Nordisk Tidsskrift for Kriminalvidenskab* 88:42–50.
- Gamman, Tor. 1995. "Uheldige helsemessige effekter av isolasjon. En klinisk studie av to grupper av varetektsinnsatte." *Tidsskrift for den norske Loegeförening*, 115:2243–46.
- Haney, C.. 2009. "The Social Psychology of Isolation: Why Solitary Confinement is Psychologically Harmful." *Prison Service Journal*, 181, 12-20.
- Haney, C.. 2008. "A Culture of Harm: Taming the Dynamics of Cruelty in Supermax Prisons." *Criminal Justice and Behavior*, 35, 956-984.

- Jackson, Michael. 1983. *Prisoners of Isolation: Solitary Confinement in Canada*. Toronto: University of Toronto Press.
- King, Roy D. 2005. "The Effects of Supermax Custody." In Liebling and Maruna, eds. *The Effects of Imprisonment*. Portland, OR: Willan Publishing.
- Koch, Ida E. 1982. "Isolationens psykiske og sociale følger." *Månedsskrift for Praktisk Lægegering* 60(June):369–32.
- Lovell, David. 2008. "Patterns of Disturbed Behavior in a Supermax Population." *Criminal Justice and Behavior*, 35, 985.
- O'Keefe, M.L., K.J. Klebe, A. Stucker, K. Strum, and W. Leggett. 2010. *One Year Longitudinal Study of the Psychological Effects of Administrative Segregation*. Colorado Springs, CO: Colorado Department of Corrections.
- Sestoft, Dorte Maria, Henrik Steen Andersen, Tommy Lillebæk, and Gorm Gabrielsen 1998. "Impact of Solitary Confinement on Hospitalization among Danish Prisoners in Custody." *International Journal of Law and Psychiatry* 21:99-108.
- Shalev, Sharon. 2009. *Supermax: Controlling Risk through Solitary Confinement*, Portland, OR: Willan Publishing.
- Smith, Peter S. 2006. "The Effects of Solitary Confinement on Prison Inmates: A Brief History and Review of the Literature." In Michael Tonry, ed. *Crime and Justice*, 34, 441-528.
- Stang, Jan, Jens Moe, Inge Arne Teigset, Bjørn Østberg, and Tron Anders Moger. 2003. "Fanger i sikkerhetscelle - en utfordring." *Tidsskrift for Den norske Lægeforening*. 123(13–14):1844–46.
- Thelle, Hatla, and Anne-Marie Traeholt. 2003. "Protection of Suspects' Rights versus Investigation Needs: The Use of Solitary Confinement in Denmark." In *How to Eradicate Torture: A Sino-Danish Joint Research on the Prevention of Torture*, edited by Morten Kjærum, Xia Yong, Hatla Thelle, and Bixiaoqing. Beijing: Social Sciences Documentation Publishing House.
- Toch, Hans. 1992. *Mosaic of Despair: Human Breakdowns in Prison*. Princeton, NJ: Princeton University Press.
- Volkart, Reto, A. Dittrich, T. Rothenfluth, and W. Paul. 1983. "Eine kontrollierte Untersuchung über psychopathologische Effekte der Einzelhaft." *Schweizerische Zeitschrift für Psychologie und ihre Anwendungen*, 42:25–46.
- Volkart, Reto, Thomas Rothenfluth, Erner Kobelt, Adolf Dittrich, and Klaus Ernst. 1983. "Einzelhaft als Risikofaktor für psychiatrische Hospitalisierung." *Psychiatria Clinica*, 16:365–77.

EXHIBIT 19

Grassian & Kupers, "The Colorado Study vs. The Reality of the Supermax"

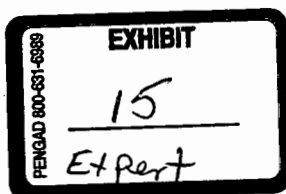
The Colorado Study vs. the Reality of Supermax Confinement

by Stuart Grassian, M.D.,J.D.¹ and Terry Kupers M.D., M.S.P.²

Just about everyone who has taken a serious look at long-term isolated confinement (as in supermaximum security or long-term administrative segregation) has concluded there is serious harm from long-term isolated confinement.³ Most of the published research regarding inmate mental health in solitary has been based upon record review and clinical interview. That body of work is extensive, and it is supported by a plethora of related studies.

There are studies of the 19th American Penitentiary System, studies of the 19th – early 20th Century German medical literature, the extensive research resulting from the Korean War and KGB interrogation practices, the research regarding profound sensory deprivation precipitated by those concerns, as well as the literature regarding other situations of social and perceptual deprivation. There are also studies of medical situations, explorers, the experience of workers wintering over at polar work stations, and so forth. And then there are the many rigorous reports and declarations filed with the courts regarding the harmful effects of long-term isolated confinement, including many by the present authors: Drs. Grassian and Kupers.

One of the most stunning and inescapable statistical facts regarding long-term segregation is that on average, 50% of completed suicides by inmates occur among the 2-8% of prisoners who are housed in isolated confinement.⁴ This fact can mean only two things: either it demonstrates that segregation is psychologically toxic, or else it demonstrates that the more troubled inmates who need psychiatric help are instead placed in a psychiatrically punitive environment. Of course, it is both: the more psychologically



VERMILLION 004760

troubled inmates have less control over their behavior, and the system's response to their unacceptable behaviors is to punish them with isolation. The troubled inmate then psychologically deteriorates in segregation.

The Colorado researcher's data itself includes quite a lot of psychiatric distress and quite a few psychotic and suicidal crises among the subjects with mental illness in administrative segregation during the study period. The researcher's, however, chose to ignore these crises or dismiss them as insignificant. Perhaps because they deemed this tendency towards psychiatric crisis to be to be pre-existing, they did not conclude that the suicidal and psychotic crises that occurred in the course of their study reflected harmful effects of isolated confinement. Importantly, they made this determination without actually interviewing the prisoners, or carefully reviewing their clinical charts. This is very odd, and certainly problematic in terms of clinical science.

We will comment further on about methodology but on the issue of a pre-existing inclination, consider a hypothetical young adult who attempted suicide as an adolescent, maybe after being jilted by a girlfriend, then entered prison, and, while doing a stint in administrative segregation, despaired of ever getting out of isolation and made a serious suicide attempt. Would we dismiss the suicide attempt as a pre-existing proclivity toward self-harm that was not caused by confinement in Ad Seg? Yet that is essentially what these researchers have done regarding the psychiatric symptoms and crises experienced by the subjects they studied.

Both of the authors offered feedback to the Colorado researchers about problems in their study, and our feedback was refused, ignored or rejected. Dr. Grassian was invited by the authors to participate in their presentation of this research at the 2010

American Psychological Association (APA) Annual Meeting, and there he pointed out several seeming fatal flaws in their methodology. Yet the Colorado research team chose not to incorporate or respond to any of these concerns. Further, they refused to provide us with the raw data from their study.

The critique offered here is based upon the report itself, discussions held (with Dr. Grassian) publicly at the presentation at the APA Meeting, presentations and discussion at a conference on supermaximum security units held in Washington, D.C. on November 18, 2010, where Dr. Kupers and Dr. Metzner spoke, and on material gleaned from discovery in *Dunlap v. Zavaras*, USDistCt, Colorado, Civ. No. 09-CV-01196-CMA-MEH, including the transcript of the deposition of the lead author for the Colorado Study, Maureen O'Keefe, as well as e-mail memoranda between the authors and advisors generated from the beginning of the study period and included in discovery.

Research Subjects, Control Group

The research authors argue that in this study, the Ad Seg group with mental illness — the group whose adjustment in Ad Seg is centrally at issue in the research — has a “comparison group”: the group in general population (GP) with mental illness. The authors pride themselves on having thus obtained in this manner virtually a controlled study.⁵ It should be noted however that the researchers excluded all potential subjects who could not read at an eighth grade level. They provide little information as to the number or percentage of potential subjects so excluded, nor of the likely explanations for this illiteracy (how many of these were simply non-English speakers; how many had significant cognitive limitations, etc.).

This omission is quite important. It has been well-documented that illiteracy and cognitive impairment are significant risk factors for psychiatric decompensation in solitary. Thus, the researchers excluded many of the most vulnerable individuals. Similarly, the authors properly excluded inmates who did not agree to participate in the study. Of course, they were right to respect inmates' right to consent, but again the excluded group likely includes many of the inmates suffering the most harm from isolated confinement.

Data Collection and the Problem of Validation

In the Colorado study, the researchers had the subject inmates fill out self-report rating scales. Usually the instructions for utilizing such scales include the recommendation that they not stand alone, but rather be integrated with clinical history and examination. The Colorado researchers, however, did not use any clinical data at all. While this methodology has certain advantages, including ease in accomplishing a study, it has the major difficulty of establishing validity. The question, of course, is whether these self-report scales are a valid measure of the subject inmates' *actual* psychiatric status. In the Colorado study, this is a very dubious proposition.

In general, the instruments employed were validated only for people in life situations extremely different from that of the subject inmates. The instruments have been validated for college students, most of whom were studying psychology, and for outpatients in psychotherapy. It is not surprising that subjects in these two groups filled out the self-reports reasonably thoughtfully and accurately — their self-reports thus being a valid, reasonably accurate, reflection of their clinical state.

But *inmates* are in *no* way similarly placed. In prison, revealing weakness or psychological dysfunction is dangerous, potentially subjecting the inmate to harassment, possibly even to physical danger. Moreover, in deposition,⁶ Ms. O’Keefe, the first study author, was asked what explanation was given the subject inmates as to the purpose of the study. In response, she revealed that the subjects were told that the research was intended to study how inmates were adjusting to prison life. She had no real answer to the follow-up questions — whether she really thought an inmate would think it wise to declare he was adjusting poorly. Anyone with a background in corrections knows that is *not* the kind of information an inmate would likely expose. It could harm him, even surreptitiously, for example at a parole hearing or in hearings to determine whether he could progress to higher levels in Ad Seg. At her deposition, Ms. O’Keefe also admitted⁷ that if an inmate reported suicidal thinking, this would be reported to prison staff. Again, there is stigma attached to mental disorder and displaying weakness in prison, and there is the likelihood of being sent to a very restrictive observation setting, all of which contributes to unbalanced reporting.

There are other problems as well. For example, the graduate student, Alyusha, who actually met with the inmates is apparently an attractive young woman, talking with inmates who had virtually no contact with any such young attractive women. Even the research group itself noted the likely distorting effect of this fact, referring to it as the “Alyusha Effect.” The inmates were likely to be reluctant to reveal weakness to this attractive young woman.

Thus, it cannot be assumed that inmate self-reports are a valid means of assessing psychiatric status. It would not be surprising if these self-reports, in fact, bore little or no relationship at all to psychiatric status.

Perhaps in an attempt to bolster the credibility of the inmate self-reports, the researchers had custody officers and mental health clinicians fill out brief forms regarding the mental health of the subject inmates. However, by their own admission at public forums,⁸ the authors acknowledged that these reports were of little value. They have no idea which staff members were selected to fill out the forms, or how the forms were completed. No specific instructions were provided, and over half the forms were never filled out at all. Similarly with the forms filled out by the clinicians, the authors gave no guidelines or requirements as to how the forms would be filled out. They had no information whatsoever to suggest that the clinicians did more than they would normally do in a screening interview, that is, attempt to speak to the inmate through the cell door, either by talking through the crack at the edge of the door or else by opening up the food slot and bending down in an uncomfortable position to speak through the slot. Given the daily burden of routine paper work, it would not be surprising to find that the staff put minimal or no effort at all into checking off the researchers' forms.

And indeed, the clinician forms found even less symptomatology than the forms completed by the inmates.

The Authors Chose to Ignore Critical Sources of Data

The most important comparison groups are the two groups of inmates with mental illness (MI) diagnosis referred for disciplinary hearing — one group was then housed in Ad Seg and the other group was then housed in GP. Since both groups have psychiatric

diagnoses, there are records of mental health contacts, including symptoms reflected in clinicians' notes, diagnoses, medications prescribed, and so forth. The Colorado researchers failed to review any of this available data, and therefore they cannot answer even a simple question such as "Did those in Ad Seg end up requiring more medication than those in GP?"

Indeed, at deposition, Ms. O'Keefe acknowledged that the study entirely failed to track the mental health history and records of the study inmates, including their medication history; for example, whether an inmate's need for medication increased during the study period. At an oral presentation of the report in Denver, it was pointed out by an ex-inmate that, as a result of the logistics of medication distribution, inmates actually receive prescribed medications much more consistently in Ad Seg than in GP. Ms. O'Keefe acknowledged that this issue, and the availability of mental health services in general, was not examined by the study group. However, she did acknowledge that the level of mental health services was greater at CSP than in GP, and that it was indeed possible that after transfer to CSP, inmates with mental illness required increased services and medication. That issue, however, was never examined.

In general, then, the study group chose to ignore major direct sources of information (mental health records, medication records, etc.) about how the inmates with mental illness fared during the study period.

The Authors Chose to Ignore DOC Data that Squarely Contradicted their

Conclusions

Colorado DOC files record incidents of emergency psychiatric contact (e.g. suicidal or self-destructive behavior) and emergence of psychotic symptoms. Among the

group of inmates with mental illness in Ad Seg (N=59) there were 37 such episodes during the course of the study (an average of .62 episodes per inmate — almost two for every three inmates). Among the group of inmates with mental illness in GP (N=33), on the other hand, there were only three (.09 per inmate — less than one for every 10 inmates). Could this have been random — i.e., not a reflection of some significant difference in the result? Statistically, the chance of that is entirely minute, approximately $p=.0002$; i.e., a chance of 1 in 5,000, an extremely small number. (In research, statistical significance requires only a probability of randomness of .05, i.e., as much as 1 in 20!) Thus, this objective data *squarely* contradicts the authors' conclusion that Ad Seg does not produce significantly more psychiatric difficulties than does GP housing. The authors simply declined to perform this straightforward statistical analysis of data they actually reported, even after the oversight in their early public reports was explicitly pointed out by Dr. Grassian.

Additionally, this data is critical as a proper means of assessing validity of the self-reports: If the self-reports *were* a valid measure of psychiatric distress, we should see each crisis episode reflected in the inmate's corresponding self-report. If, in filling out his self-report, the inmate responds that he is doing just fine, then the self-reports are worthless. They are in no way a measure of psychiatric distress. It would have been quite easy for the authors to review these cases, a total of 37 recorded instances that would require simply a review of the corresponding self-report rating by the inmate during the time period at issue. Dr. Grassian explicitly pointed this out to the authors prior to their public presentation of the data and prior to their submission of the report. Yet the authors declined to perform this crucial check on their data.

There is irrefutable evidence that the study group *knew* there was a major problem with the validity of the self-report data. In 2008, Ms. Stucker sent an e-mail to Ms. O'Keefe expressing concern that an inmate subject in the study had just committed suicide. She then reviewed his self-report. In his self-report, he had revealed no evidence at all of any distress. Thus, at an early stage, Ms. O'Keefe was entirely aware of a major question about the validity of the inmates' self-report ratings. Evidently, the study group chose to do nothing at all to address this concern even though it would have been entirely possible to do so.

In the end, though, the authors could not escape the inevitable conclusions to be drawn from this data. As we stated in the introductory portion of this critique, statistical evidence demonstrates a dramatically increased incidence of suicide among prisoners in segregation. In this study, we see the very same result: Psychiatric crises, whether of suicidality or psychotic or other symptomatology, were dramatically more prevalent among the prisoners with mental illness placed in Ad Seg compared with those with mental illness housed in GP. Again, this can mean only that more disturbed inmates are the ones most likely to be sentenced to Ad Seg, or that housing a group of psychiatrically impaired inmates in Ad Seg creates a much worse result than housing them in GP.

The Colorado researchers start by praising themselves for creating a comparison group (i.e., the only variable distinguishing the group with mental illness in Ad Seg from those with mental illness in GP is one variable: housing). Thus their report explicitly *excludes* the first possibility, that the mentally ill inmates sent to Ad Seg were a *different* group — a sicker group — than those housed in GP. In short, contrary to the researchers'

conclusions, the study clearly demonstrates the second possibility: That Ad Seg housing is psychiatrically toxic.⁹

Conclusion

When evaluating an inmate who has suffered some form of psychiatric deterioration during incarceration, there are several sources of data that can establish causation. Interview data, prison mental health records, and D.O.C. incident reports all provide important information about the circumstances surrounding the deterioration and the nature of the resulting psychiatric symptomatology. Over the course of years, we and others have described literally thousands of cases of individuals who decompensated in solitary confinement, recompensated when removed, and then decompensated when returned, in an endless revolving door.

The Colorado researchers elected not to talk to their subjects, nor to review records. They did paper and pencil tests but no clinical interview or even a researcher-conducted interview. Unfortunately, the results of this kind of stand-alone testing are such that the researchers can claim no harm from supermax confinement merely because the data is a scramble of numbers that mean almost anything to anyone who wants to interpret them.

There are a number of other methodological difficulties with the Colorado research report, but in the end, much of the 163-page final report consists of long and endless statistical dissections of the self-report data. Yet these minute dissections are entirely confounding and erroneous because the data they dissect does not in any meaningful manner reflect the psychiatric pathology they are supposed to be studying.

The Colorado research team did not find an absence of harm. Far from it. They found, not surprisingly, that many of the inmates who faced disciplinary sanctions for disruptive behavior were very damaged people with serious mental illness diagnoses and with very serious psychiatric problems. Their data also demonstrated emphatically that among those inmates with preexisting serious psychiatric problems, those who were placed in administrative segregation suffered far more psychiatric crises during the study period than those not placed in administrative segregation. The authors chose to ignore this glaring reality. Instead, relying only upon their very flawed methodology, they claim their study demonstrates that there was no change, or even some early improvement, in the psychological status of these inmates. In the process they ignored objective data that squarely contradicted their self-report data. This is entirely unacceptable.

Returning to the stunning statistic that, on average, 50% of completed suicides in corrections occur among the 2% to 8% of prisoners in any system who are in isolated confinement,¹⁰ there are only two plausible explanations for this fact: Either administrative segregation causes psychiatric harm; and/or (*and* we believe "and" is the applicable word) the sentencing to Ad Seg is very often a tragic, punitive response to irrational and self-destructive behavior on the part of severely mentally ill inmates – just when a therapeutic response is urgently needed.

We need to think carefully about this. There are many very damaged people in the Colorado supermax under study. Quite a few suffer acute incidents of psychosis and commit suicide during the course of the study while others experience many disturbing symptoms, but because they are in isolated confinement, they are not being given the

intense treatment their conditions require. Does it make any sense at all to conclude that supermax confinement does no harm?

Endnotes

1. Dr. Grassian is a Board-certified psychiatrist and was on the teaching staff of the Harvard Medical School continually from 1974 until 2002. He has had extensive experience evaluating the psychiatric effects of stringent conditions of confinement, and has served as an expert in a number of both individual and class-action lawsuits addressing this issue. His observations and conclusions regarding the psychiatric effects of such confinement have been cited in a number of federal court decisions. Much of this work is described in “Psychiatric Effects of Solitary Confinement”, published in the *Washington University Journal of Law and Policy*, 2006, vol. 22, pp. 325-383.
2. Dr. Kupers is Institute Professor at The Wright Institute and practices psychiatry in Oakland. He provides expert testimony as well as consultation and staff training regarding the psychological effects of prison conditions including isolated confinement in supermaximum security units, the quality of correctional mental health care, and the effects of sexual abuse in correctional settings. He is the author of *Prison Madness: The Mental Health Crisis Behind Bars and What We Must Do About It* (1999), a co-editor of *Prison Masculinities* (2002), and Contributing Editor of *Correctional Mental Health Report*.
3. Scharff-Smith, P. (2006). The effects of solitary confinement on prison inmates: A brief history and review of the literature. In M. Tonry (Ed.), *Crime and justice* (Vol. 34, pp. 441-528). Chicago: University of Chicago Press. See also Amicus

Brief to the Supreme Court of the United States. (2005). *Brief of professors and practitioners of psychology and psychiatry as amicus curiae in support of respondents*. Supreme Court of the United States, No. 04-495.

4. Mears, D.P. & Watson, J. (2006). Towards a fair and balanced assessment of supermax prisons. *Justice Quarterly*, 23(2),232-270; Way, B., Miraglia, R., Sawyer, D., Beer, R., & Eddy, J. (2005). Factors related to suicide in New York state prisons. *International Journal of Law and Psychiatry*, 28(3), 207-221; Patterson, R.F. & Hughes, K. (2008). Review of Completed Suicides in the California Department of Corrections and Rehabilitation, 1999 to 2004, *Psychiatric Services*, 59(6), 676-682.
5. The authors point out that to be a controlled study, the study must randomly assign subjects to each group, whereas in this study, the subject inmates were assigned to Ad Seg or to GP by the Correctional staff, not by the study group.
6. **Dunlap v. Zavaras**, USDistCt, Colorado, Civ. No. 09-CV-01196-CMA-MEH. Deposition of Maureen O'Keefe, October 5, 2010, pages 136-137.
7. O'Keefe deposition, p. 137
8. E.g., "Longitudinal Study of the Psychological Effects of Administrative Segregation", presentation at 2010 Annual Meeting of American Psychological Association, San Diego, August 14, 2010.
9. Ironically, in deposition, when Ms. O'Keefe was confronted with this massive discrepancy in crises among prisoners with mental illness in Ad Seg versus those in GP, she offered that maybe those in Ad Seg were a sicker group. Under the pressure of deposition, she seems to have momentarily forgotten that her study

was founded upon a notion that there was a "control group," i.e. that the two groups were psychiatrically comparable! *O'Keefe deposition pp. 196-198.*

10. Op. cit. #4, Mears....

EXHIBIT 20

**Zinger et al.,
"The Psychological Effects of
60 Days in Administrative
Segregation"**

SEGREGATION

SÉGRÉGATION

The psychological effects of 60 days in administrative segregation

Ivan Zinger

*Correctional Service of Canada and Carleton University
Ottawa, Ontario*

Cherami Wichmann

*Carleton University
Ottawa, Ontario*

and

D.A. Andrews

*Carleton University
Ottawa, Ontario*

Les participants dans cette étude sont 60 détenus canadiens des pénitenciers de Kingston, de Collins Bay et de Millhaven qui (a) avaient été placés en ségrégation à leur propre demande ou involontairement et y avaient passés 60 jours, ou (b) avaient été choisis au hasard dans la population carcérale générale et y étaient demeurés pour au moins 60 jours. Les participants ont dû compléter des tests psychologiques et être interviewés pour déterminer leur santé mentale et leur fonctionnement psychologique. Ce processus d'évaluation a été répété après 30 jours et après 60 jours. Les détenus en ségrégation avaient les mêmes niveaux académiques, les mêmes histoires criminelles et les mêmes besoins que les détenus de la population carcérale générale. Les détenus en ségrégation, toutefois, avaient des personnalités différentes et étaient considérés plus dangereux que les détenus de la population carcérale générale. De façon générale, ils ont une santé mentale plus détériorée et ils fonctionnaient moins bien. Par ailleurs, il n'y avait pas d'indication à l'effet que la période en ségrégation avait détérioré de façon marquante leur santé mentale ou leur fonctionnement psychologique.

Participants in this longitudinal study included 60 Canadian inmates from Kingston, Collins Bay, and Millhaven Penitentiaries who had either been (a) voluntarily or involuntarily placed in administrative segregation and remained in segregation for 60 days, or (b) randomly selected from the general inmate population and remained in the general inmate population for 60 days. Participants initially completed written

Canadian Journal of Criminology
Revue canadienne de criminologie

47 to 83
January/janvier 2001

psychological tests and took part in a structured interview that assessed their overall mental health and psychological functioning. The same procedure was undertaken 30 days later, and again 60 days later. Segregated prisoners had similar education levels, offence histories and criminogenic needs than non-segregated prisoners. Segregated prisoners had distinct personalities, however, and were higher risk cases than non-segregated prisoners. Overall, segregated prisoners had poorer mental health and psychological functioning. There was no evidence, however, that, over a period of 60 days, the mental health and psychological functioning of segregated prisoners significantly deteriorated.

In Canada, the percentage of segregated prisoners has more than doubled in the last ten years, now representing approximately 5.5% of federally sentenced prisoners (Pierson 1988; Kane 1997). However, little research has been conducted on these prisoners. Moreover, many scholars have assessed the literature on penal segregation as sparse, conflicting, rife with speculations, and based upon far-fetched extrapolations and generalizations (Barak-Glantz 1983; Brodsky and Scogin 1988; Suedfeld, Ramirez, Deaton, and Baker-Brown 1982; Wormith, Tellier, and Gendreau 1988).

Notwithstanding the inadequacy of the existing research, two conflicting perspectives on the effects of segregation on prisoners have emerged. Some researchers describe segregation as "cruel and unusual punishment" and psychologically damaging (Benjamin and Lux 1975; 1977; Grassian 1983; Immarigeon 1992; Jackson 1983; Korn 1988; Luise 1989; Martel 1999), whereas others provide evidence that segregation has little, if any, negative psychological effect on prisoners (Bonta and Gendreau 1995; Ecclestone, Gendreau, and Knox 1974; Gendreau, Freedman, Wilde, and Scott 1972; Gendreau and Bonta 1984; Suedfeld *et al.* 1982).

Resolving the question of the impact of segregation carries important policy implications for areas such as: (a) the level and frequency of monitoring and assessment required for prisoners in segregation (mandatory vs. upon request); (b) programming to reduce mental health deterioration (need for, and type of, intervention programs); and (c) the adequacy of current assessment strategies (what aspects of psychosocial functioning are important to assess, and which are less affected by segregation).

This article contains two sections. First, a review of methodological issues highlights the current unsatisfactory state of the literature on the effects of segregation. This review shows that supporters of one view often fail to appreciate the findings of the opposing view, as well as to recognize the limitations of their own findings when drawing their conclusions. The ability to generalize the results of these studies is affected to varying degrees by improper attention to methodological shortcomings of the research conducted. Second, the findings of a research project which addressed the shortcomings of the existing literature is presented.

Part I: Evaluation of existing research on segregation: A review of methodological shortcomings

1. Reliance on qualitative data (casual observations, interviews and anecdotes)

Many authors use anecdotal evidence to support their claims (Benjamin and Lux 1975, 1977; Brodsky and Scogin 1988; Grassian 1983; Jackson 1983; Korn 1988; Martel 1999). These authors often take selected but powerful excerpts from interviews of segregated prisoners or mental health professionals with experience with segregated prisoners to provide general evidence of the harmful effects of segregation. Some authors rely on testimony on the use of isolation in the 19th century to produce corroborative evidence of the harmful effects of segregation in today's correctional context (Grassian 1983; Immarigeon 1992; Luise 1989). Others cite human rights violation litigation to depict the general conditions of confinement and treatment of segregated prisoners, as well as the psychological and physical harm that ensues (Benjamin and Lux 1977; Birkinshaw 1981; Jackson 1983; Luise 1989).

The evidence of the damaging effects of segregation on prisoners adduced by these authors is very disturbing, and cannot be ignored. Because of the nature of the methodology, it is often unclear whether the pathologies displayed by some segregated prisoners were directly attributable to the conditions of confinement in segregation or whether these prisoners

displayed similar pathologies in the general prisoner population or in the community, prior to being segregated (Gendreau and Bonta 1984).

In addition, Suedfeld *et al.* (1982) found that some authors inappropriately use findings from case studies of persons who experienced severe abuse and sensory deprivation to illustrate the damaging effects of segregation. Testimony of tortured political and war prisoners who were denied food, clothing, medical assistance and procedural fairness are at times relied upon to establish the damaging effects of segregation in contemporary North American correctional settings (Benjamin and Lux 1975; Korn 1988). The generalization of the findings of these case studies has been questioned (Gendreau and Bonta 1984; Suedfeld *et al.* 1982). Isolation in a political or war camp is not comparable to the highly regulated and formalized procedures for imposing segregation on prisoners in North American penitentiaries. Conditions of confinement, procedural safeguards, and the level of safety and security provided to the prisoners differ to such an extent that a comparison may well be inappropriate (Gendreau and Bonta 1984; Suedfeld *et al.* 1982).

2. Conditions of confinement

One of the problems with segregation research stems from difficulties in defining the constructs being evaluated. Many terms, such as administrative segregation, dissociation, isolation, seclusion, protective custody, and solitary confinement are used, often interchangeably, to describe various restrictive environments. These terms encompass a wide range of conditions of confinement in which the number of restrictions on freedom of association and freedom of movement may vary, and in which levels of perceptual deprivation, sensory deprivation, and social isolation may also vary. There is such a diversity in the nature of conditions of confinement used in segregation research that aggregating all studies under the same "solitary confinement" label has been described by some as inappropriate (Suedfeld *et al.* 1982).

Many authors recognize the importance of the environment with respect to its impact upon the segregation experience and the difficulty associated with generalizing results (Grassian 1983). Conditions of confinement and daily routine vary so greatly

among correctional institutions (Kane 1997; Vantour 1975) that results derived from one institutional setting may not be applicable to others. For example, the frequency and quality of interactions with staff or other prisoners, the physical layout of segregation cells (e.g., solid doors, cell size, etc.), the size of the exercise yard, the availability of recreational equipment and hobby items, and the access to personal effects, programs, and services, may all affect the segregation experience. As a result, the majority of studies describe, at great length, the conditions of confinement and the daily routine of segregated prisoners being studied.

Many authors have reviewed the proliferation of control units in the United States and abroad in an attempt to determine their effects on prisoners' mental and physical health (Birkinshaw 1981; Coyle 1987; Dowker and Good 1993; Korn 1988; Immarigeon 1992). The establishment of control units in the United States originated in 1963 when a penitentiary in Marion (Illinois) was built to replace Alcatraz (Coyle 1987). Since then, more than 33 States have comparable Marion-like facilities (Immarigeon 1992). Control units provide a good illustration of the difficulty in defining the constructs being evaluated. Although control units are not formally recognized by correctional authorities as segregation units, and although they sometimes impose fewer restrictions on prisoners than in traditional segregation units, they often impose many similar conditions of confinement (Coyle 1987; Dowker and Good 1993; Immarigeon 1992; Korn 1988). For example, Dowker and Good (1993) describe some of the defining features of these institutions. Prisoners are confined in small cells for 22 or 23 hours per day. The cells are often equipped with solid steel doors, which prevent any communication between prisoners. Further, these institutions are often equipped with remote electronic sliding doors, which minimize, if not eliminate, most contact with correctional staff. There are no communal dining, exercise, or religious services, and few, if any, work opportunities.

3. Relevance of field and laboratory experiments on sensory deprivation

Most of the experimental studies on the effects of segregation come from the field of sensory deprivation. Gendreau and his

colleagues have generated and evaluated many theories and hypotheses on sensory deprivation in the correctional context. For example, Gendreau and colleagues examined whether: isolated prisoners show higher arousal potential because of a lower arousal level induced by solitary confinement (Gendreau *et al.* 1972); segregation enhances learning (Gendreau, McLean, Parsons, Drake, and Ecclestone 1970); isolated subjects desire a lower level of stimulation (visual and auditory sensory input) after a deprivation experience (Gendreau, Freedman, Wilde, and Scott 1968); and stress levels, as indicated by adrenocortical activity, can detect whether solitary confinement is harmful (Ecclestone *et al.* 1974).

Others have commented upon or tested theories and hypotheses of sensory deprivation in the correctional context as well. For example, Benjamin and Lux (1977) argue that segregation is harmful because it dramatically reduces levels of needed stimulation. Dowker and Good (1993) believe that prisoners who are segregated for long periods of time may be deprived of necessary meaningful human contacts, and, as a result, these prisoners have difficulties in coping with normal social situations again.

Suedfeld *et al.* (1982) argue that the comparison between field or laboratory experiments on isolation and stimulus reduction and today's typical North American segregation environment is inappropriate. They contend that it is highly questionable whether the typical segregation unit in fact imposes much reduction in stimulus input. They state that most segregated prisoners can communicate with guards and other prisoners and have access to reading material, mail, lawyers, other visitors, and frequently possess radios and television sets. Gendreau and Bonta (1984) argue that the conditions of confinement in many of the sensory deprivation and isolation experiments are more severe than those found in today's segregation units. They argue that, since these field and laboratory experiments show little support for the position that sensory deprivation and isolation are psychologically damaging, the conclusions drawn from these studies are especially informative and relevant.

4. Selection of subjects

4.1 Use of Volunteers. Experimental studies on segregation rely primarily on volunteers who agree to be segregated for a fixed period of time. Some authors have been reluctant to accept results of studies which have relied on volunteers (Arbour 1996; Jackson 1983; Vantour 1975). Walters, Callagan, and Newman (1963) believe that the problem with using volunteers is that they are apparently not too frightened by the prospect of facing a few days of isolation, and they may have personality characteristics and past experiences which enable them to cope with, and remain unaffected by, segregation.

In Canada, approximately half of all prisoners placed in segregation are confined against their will (Kane 1997). In addition, it is questionable whether “voluntary” segregation is truly voluntary. Arguably, most prisoners would prefer to remain in the general prisoner population if the threat to their personal safety was to be removed. Nonetheless, some authors claim that, based on their “clinical experience”, prisoners who initially strongly object to being placed in segregation appear to adapt as well as prisoners who voluntarily request isolation (Ecclestone *et al.* 1974; Gendreau *et al.* 1972).

In addition to the issue of using volunteers, the use of alternative populations may also lead to limited generalization of findings. For example, the use of university students who, in general, exhibit good adjustment, stable personality, and higher levels of intelligence, education, and socioeconomic status may not lead to accurate comparisons with the segregated prisoner population. Suedfeld *et al.* (1982) argue that attempting to use findings from these sources as an indication of what one can expect from prisoners in segregation is inappropriate because it is not relevant to the phenomenon being evaluated. The high prevalence of severe mental disorders among segregated prisoners (Hodgins and Cote 1991) makes any comparison with university student samples somewhat questionable.

4.2 Use of prisoners involved in human rights violation litigation. Some studies on the negative effects of segregation have relied on segregated prisoners who were involved in lawsuits alleging violations of their constitutional rights (Brodsky and

Scogin 1988; Grassian 1983). Subjects involved in human rights violation litigation may have a special interest in demonstrating that their conditions of confinement have negative psychological and physiological effects. Therefore, the results of studies which rely on such prisoners will always remain questionable. Further, Suedfeld *et al.* (1982) suggest that prisoners engaged in litigation are perhaps not representative of average prisoners; their reactions to segregation may not be the norm. Similarly, Gendreau and Bonta (1984) question the reliability of information of case studies performed by Jackson (1983). They suggest that many of Jackson's interviewed prisoners were notorious, far from representative, and had filed an inordinately large number of grievances, legitimate or otherwise, against the prison system.

4.3 Screening out subjects with psychiatric history. A significant proportion of segregated prisoners have a psychiatric history (Hodgins and Cote 1991; Motiuk and Blanchette 1997). Some studies purporting to examine the impact of segregation have screened-out such subjects (Ecclestone *et al.* 1974; Gendreau *et al.* 1972). As a result, findings from these studies may be difficult to apply to the population of segregated prisoners.

Hodgins and Cote (1991) report that, in their sample of 32 long term segregated prisoners, 31% suffered from some kind of severe lifetime mental disorder (25% schizophrenia, 3.1% major depression, and 3.1% bipolar disorder). The rate of schizophrenia among this sample was more than three times the rate of the disorder among non-segregated prisoners. The rate of major depression in their sample was lower than the rate in the general prisoner population. This suggests that non-disruptive mentally-ill prisoners may remain in the general prisoner population, whereas prisoners who are "disturbed and disruptive" are isolated from the general prisoner population.

Wormith *et al.* (1988) evaluated the attributes of prisoners in protective custody (PC) in a provincial institution. PC prisoners typically can associate among themselves but do not have access to the same level of programs, services, and privileges offered to the general prisoner population. They found that PC prisoners were more likely to have a history of psychiatric problems. They suggest that PC prisoners' psychological weaknesses and idiosyncratic behaviours may not be well tolerated by the general

prisoner population, and it appears that inappropriate behaviours are often punished regardless of the underlying basis for the conduct (Carriere 1989; Gendreau, Tellier, and Wormith 1985; Rold 1992). Consequently, numerous prisoners with mental disorders are segregated (Gendreau *et al.* 1985; Rold 1992).

Little research has focused on the effects of segregation on prisoners with psychiatric conditions. Many authors argue that segregation can exacerbate some existing psychiatric conditions (Haney 1993; Hodgins and Cote 1991; Wadeson and Carpenter 1976). For example, Wadeson and Carpenter (1976) concluded that segregation stimulates hallucinatory activity and provokes paranoia among some mental health patients.

The existence of psychiatric problems may very well be a defining characteristic of the population of segregated prisoners. Moreover, findings from the studies reviewed above underline the importance of not restricting research samples to those without a history of psychiatric disorders.

5. Reasons for segregation

Prisoners may voluntarily request segregation or be involuntarily segregated for a multitude of reasons (Kane 1997; Gendreau *et al.* 1985; Wormith *et al.* 1988). The most common reasons given by prisoners for seeking various forms of PC and segregation include: conflicts in the general population (e.g., gambling and drug debts); the nature of the prisoner's offense; whether the prisoner is suspected of being an informant; the existence of personality problems; the presence of phobias (including fear of gays); being the target of sexual aggression; and escaping the crowded and often violent atmosphere of maximum security (Gendreau *et al.* 1985).

As noted above, approximately 50% of segregation placements are involuntary in nature (Kane 1997). Research on segregation thus far has failed to assess the effects of long term segregation on these prisoners. Such an omission has rendered generalization of findings even more difficult. For example, the underlying reasons for segregating prisoners may influence their abilities to cope with the experience (Weinberg 1967). Whether they view their placement in segregation as a result of their own behaviour

or as the result of being an innocent victim of circumstances beyond their control may influence their ability to cope with the more restrictive regime of segregation.

6. Attrition

Some segregation studies reported attrition among subjects participating in the experimental condition (i.e., segregation), and provided little, if any, explanations (Ecclestone *et al.* 1974; Walters *et al.* 1963; Weinberg 1967; Zubek, Bayer, and Shephard 1969). For example, Ecclestone *et al.* (1974) reported a 32% attrition rate, and Weinberg (1967) reported a 68% attrition rate. Even when more than adequate monetary incentives are provided, attrition has been reported (Bexton, Heron, and Scott 1954; Zubek *et al.* 1969).

Attrition is a major drawback to psychological research in general. The problem with attrition is especially relevant to the evaluation of the psychological effects of segregation. Subjects who decide no longer to participate in the experiment may be the same individuals who would not cope well with the conditions of segregation and would be negatively affected by them.

7. Reliance on cross-sectional research

Cross-sectional research is inadequate for evaluating the effects of segregation. The results of cross-sectional segregation research are limited to the identification of differences between groups (segregated and non-segregated). The results of this type of research do not permit inferences concerning the causes of these differences (Suedfeld *et al.* 1982). Nevertheless, after conducting a cross-sectional study and observing poorer mental and physical health among segregated prisoners than among non-segregated prisoners, some authors have attributed the cause of such poorer health to segregation (Brodsky and Scogin 1988; Miller and Young 1997). The possibility that segregated prisoners already were of poorer mental and physical health prior to their segregation must at least be considered as an alternative explanation in cross-sectional studies.

8. Duration and indeterminate nature of stay

Another problem with current experimental studies on segregation surrounds the issue of the length and indeterminate

nature of the stay (Jackson 1983; Suedfeld *et al.* 1982). In previous experimental research, the length of stay is limited to ten days or less (e.g., 2 days: Gendreau *et al.* 1970; 4 days: Walters *et al.* 1963; 5 days: Weinberg 1967; 7 days: Gendreau *et al.* 1972; Gendreau *et al.* 1968; Gendreau, Horton, Hooper, Freedman, and Scott 1968; Zubek *et al.* 1969; 10 days: Ecclestone *et al.* 1974). Moreover, volunteers for these studies know exactly when the experiment will end, and that they can end their participation at will.

The reality of segregation is that the length of stay is always unknown, and more than 80% of prisoners spend more than 10 days in segregation at any one time (Kane 1997). Suedfeld *et al.* (1982) argue that making general statements on the effects of segregation without qualifying the length of stay is inappropriate. Bonta and Gendreau (1995) specifically state that their conclusion that segregation is not detrimental only applies to periods of segregation of 10 days or less. As these studies confirm, generalizing the results of experimental studies beyond 10 days is questionable.

9. Lack of comparison group

Some studies utilizing structured and non-structured interviews with segregated prisoners have failed to include a comparison group of non-segregated prisoners (Brodsky and Scogin 1988; Grassian 1983; Martel 1999). Brodsky and Scogin (1988) interviewed 45 segregated prisoners about their confinement in solitary confinement but did not include a control group. Although they reported disturbing negative psychological and physiological effects, since no comparison group was included, the results are of little value because it remains undetermined whether prisoners in the general prisoner population would have reported similar effects about their confinement in the general prisoner population.

10. Prisoner/staff interaction and the punitive reality of segregation

Several authors have suggested that the relationship between staff and prisoners is an important factor which may affect how prisoners cope with segregation (Benjamin and Lux 1977; Bonta

and Gendreau 1995; Carriere 1989; Gendreau and Bonta 1984; Korn 1988; Suedfeld *et al.* 1982; Vantour 1975; Wormith *et al.* 1988). Prisoners may be more affected by the way they are treated by correctional staff than by the conditions of confinement typically found in North American segregation units (Bonta and Gendreau 1995; Gendreau and Bonta 1984; Vantour 1975).

Bonta and Gendreau (1995) argue that there is some evidence that, when prisoners are treated capriciously by management or correctional staff, psychological stress can result even in the most humane of prison environments. Harassment, physical violence, enforcement and non-enforcement of rules, and unpredictable withholding of privileges may play a greater role than complaints about physical conditions, the social isolation and the sensory deprivation associated with segregation (Suedfeld *et al.* 1982).

Many authors have found that contrary to legislative and policy provisions, the management of administrative segregation is based on a punitive philosophy, and that segregated prisoners have fewer rights, privileges, and access to programs and services than prisoners in the general prisoner population (Arbour 1996; Carriere 1989; Gendreau *et al.* 1985; Kane 1997; Tellier, Wormith, and Gendreau 1989; Vantour 1975). For example, Arbour (1996: xiii) concluded that the Correctional Service of Canada's management of administrative segregation was not in accordance with the law and its policies, and demonstrated a systemic "prison culture which did not value individual rights".

Wormith, Tellier, and Gendreau (1988) reported that correctional employees often have negative views towards, and discriminate against, segregated prisoners. They found that PC prisoners complained about the attitudes of correctional staff towards them and the adverse psychological effects of being in PC, whereas prisoners in the general population were more likely to complain about institutional living conditions, rules and regulations. Similarly, Carriere (1989) states that PC prisoners are often treated in a demeaning manner by correctional staff. Further, he contends that segregated prisoners are treated as maximum security prisoners regardless of the security risk they pose.

For generalization purposes, the evaluation of the effects of segregation must include real interactions between staff and prisoners, and should not be limited to courteous interactions typically found in laboratory experiments.

11. Personality

The prisoners' personality or temperament may play a role in how they will be affected by segregation. Some personality characteristics may reduce tolerance for segregation, while other characteristics may enhance it (Suedfeld *et al.* 1982; Walters *et al.* 1963). Little, if any, research on the effects of segregation has focused on personality. Assessment of personality must be included in segregation research in order to identify prisoners' abilities and predispositions to cope with segregation.

12. Other factors

Segregation may have a detrimental impact on prisoners' chances of parole, reduce their chances of being admitted to a half-way house, and affect their security classification (Carriere 1989; Gendreau *et al.* 1985; Tellier, Wormith, and Gendreau 1989). Knowledge of these consequences may negatively affect how prisoners adapt to segregation. Further, a prisoner who was housed in a single cell prior to segregation may be reassigned to a "double-bunked" cell after a placement in segregation. This future loss of privacy may also affect how prisoners cope with the experience of segregation.

Complaints about other issues such as cold food and delayed response to requests for assistance (e.g., medication, telephone calls, counselors, reading material, etc.) may also influence the segregation experience (Suedfeld *et al.* 1982). In the segregation environment, these complaints cannot be viewed as trivial because they are often the only distractions available to break the monotony of the segregation experience.

Part II: Longitudinal study on the psychological effects of 60 days in administrative segregation

Method

Design and procedures

Participants

Participants included prisoners from Kingston, Collins Bay, and Millhaven Penitentiaries who had either been (a) placed in administrative segregation and remained in segregation for 60 days (quasi-experimental group), or (b) randomly selected from the general prisoner population and remained in the general prisoner population for 60 days (comparison group). Data were collected over an eight month period beginning in October 1997.

Testing and procedures

Senior psychologists at the selected institutions supervised the data collection. The psychologists selected and trained/oriented three research assistants (RA's) concerning institutional security protocols and the use of the psychological testing instruments. The RA's were graduates or students of psychology (one 4th year student, one M.A. candidate, and one M.A.).

Prisoners who were just placed (voluntarily and involuntarily) in administrative segregation and provided their informed consent were asked to complete written psychological tests and take part in a structured interview. After each session, participants were debriefed. The same procedure was undertaken 30 days later and again 60 days later if the prisoners remained segregated. Non-segregated prisoners were selected at random and underwent the same testing procedures at the same intervals.

Measures

The initial testing session (session one) lasted approximately two hours. In addition to the battery of tests which were utilized at each session, the initial session included a general measure of intelligence and a short personality inventory. Because

performance on these additional instruments was not expected to fluctuate over 60 days, these measures were administered only once. The follow-up assessments conducted at 30 days (session two) and 60 days (session three) were therefore shorter, each lasting approximately one hour. The comparison group underwent the same testing procedure as the segregated group.

The measures which were selected for use in this study were chosen based on several criteria. Measures were selected which possessed acceptable psychometric properties, had a short administration time, and had been previously used with prisoner samples. Consideration was also given to measures which had been used in previous segregation research. Table 1 illustrates the list of measures which were selected for use in this study and their respective alphas (reliability estimates which measure internal consistency).

Additional data collection.

Physical conditions. Research assistants gathered information on the physical layout of the segregation units.

Prisoner intake assessment. All prisoners sentenced to penitentiaries (i.e., for prison terms exceeding two years) must complete the Offender Intake Assessment (OIA) prior to their penitentiary placement. In most instances, the OIA lasts eight weeks, and allows Correctional Service Canada (CSC) to render informed decisions with respect to placement, classification, and programming. During the OIA, information on prisoners is collected and stored on the computerized Offender Management System (OMS).

OIA information was retrieved on prisoners' current and past criminal history and the seven need domains (Employment, Marital/Family, Associates, Substance Abuse, Community Functioning, Personal/Emotional, and Attitude). Prisoners' scores on the Statistical Information on Recidivism (Nuffield 1982; SIR Scale) were also retrieved. The SIR score provides an estimate of the probability that an individual will re-offend within three years after release. Each prisoner's total score on the SIR Scale can range from -30 (very poor risk) to + 27 (very good risk).

Table 1
Measures used and respective Alphas

Measures (initial assessment only)	Alpha*
Interview assessment	n/a
NEO Personality Inventory (short form)	
Neuroticism	.84
Extraversion	.70
Openness	.61
Agreeableness	.71
Conscientiousness	.80
Shipley	n/a
Measures (all three sessions)	Alpha*
Aggression questionnaire	.89
Balanced Inventory of Desirable Responding (short form)	.69**
Beck Depression (abbreviated)	.89
Brief Symptom Inventory	
Somatization	.86
Obsessive-compulsive	.83
Interpersonal	.79
Depression	.84
Anxiety	.83
Hostility	.85
Phobic anxiety	.80
Paranoid ideation	.80
Psychoticism	.66
Holden Psychological Screening Inventory	.84
Hopelessness scale	.89
Interview assessment	n/a
State-trait Anxiety Inventory (State-short form)	.83
WAIS Sub-test: Digit Span	n/a
WAIS Sub-test: Digit Symbol	n/a

Note: * Reliability estimates which measure internal consistency; ** Items 4, 7 and 9 were removed to improve psychometric properties.

Results

Descriptive findings

Conditions of confinement

Information on the conditions of confinement of segregation units at Collins Bay, Kingston, and Millhaven penitentiaries was collected. Table 2 describes the physical conditions at each

penitentiary. Conditions of confinement at Kingston Penitentiary are divided into two sections because one of the segregation units is noticeably different from the others.

Table 2
Conditions of confinement in segregation at Collins Bay,
Kingston, and Millhaven Penitentiaries

Characteristic	Institutions			
	Collins Bay	Millhaven	K.P. 1	K.P. 2
Cell size (sq.ft)	80.6	57.2	56	46
Ceiling height (ft./in.)	7'10"	11'8"	9'	11'5"
Number of cells per range	19	16	20	37
Solid door	Yes	Yes	Yes	No
Yard size (sq.ft.)	750	1200	1500	1500
Concrete wall around yard	Yes	Yes	Yes	Yes
Yard covered overhead with wired fence	Yes	Yes	Yes	Yes

Participation and attrition

The refusal rate for participating in this study was 44% for segregated and 40% for non-segregated prisoners. Table 3 illustrates the number of completed sessions broken down by Group (i.e., segregated vs. non-segregated) and institutions. It shows that 83 segregated prisoners and 53 non-segregated prisoners participated in this study. Complete data for all three sessions (60 days) were only available for 23 segregated and 37 non-segregated prisoners.

The loss of participants from the segregated group was primarily due to releases to the general prisoner population or transitional units (i.e., protective custody), or transfers to other institutions. True attrition, the refusal to participate in a subsequent session, occurred in nine cases (10.8%). It should be noted that true attrition included cases in which prisoners expressed their intent to participate in the study but their conduct jeopardized the personal safety of the RA's (e.g., threats, and one incident in which a prisoner attempted to grab an RA). The average elapsed time after placement in segregation for session one, two and three was 3.6, 29.8 and 57.8 days respectively.

For non-segregated prisoners, the loss of participants was mainly due to transfers to other institutions and placement in segregation. True attrition occurred in only two cases (3.8%).

Table 3
Number of completed sessions (S.1, S.2, and S.3) broken down by group and institutions

Institution	Segregated (n=83)			Non-Segregated (n=53)		
	S.1	S.2	S.3	S.1	S.2	S.3
Collins Bay	31	8	7	19	16	16
Kingston	19	12	11	20	17	14
Millhaven	33	12	5	14	13	7
Total	83	32	23	53	46	37

Initially (i.e., upon placement), 39% ($n = 32$) of segregated prisoners were voluntary cases whereas 61% ($n = 51$) were involuntary cases. Moreover, voluntary cases were all seeking protection, whereas the majority of involuntary cases (71%, $n = 36$) were legally placed in administrative segregation for jeopardizing the safety of other prisoners, staff, or the security of the institution. After 60 days the percentage of voluntary cases increased to 57% ($n = 13$) and the percentage of involuntary cases decreased to 43% ($n = 10$).

Demographics

Age. Using t-tests, segregated prisoners were found to be younger ($M = 28.9$) than non-segregated prisoners ($M = 32.20$, $t(134) = 2.66$, $p < .01$). The age of segregated and non-segregated prisoners ranged from 20 to 54 years old. Bivariate correlation analyses showed that Age was not significantly correlated with any measure (i.e., dependent variable).

Race. The Offender Intake Assessment (OIA) was used to provide background information on the prisoners. OIA information on race was available on 119 prisoners: 66% of prisoners were Caucasian, 25% black, 7% aboriginal, and 3% from other visible minority groups. Among segregated prisoners ($n = 73$), the percentage of Caucasians, blacks, aboriginal, and other visible minority groups was 64%, 27%, 7%, and 1% respectively.

Education. OIA information was used to assess the educational background of prisoners. Using chi-square analyses, segregated and non-segregated offenders did not significantly differ in educational background.

Criminal history. OIA data also provided criminal history information for 131 prisoners. No significant difference between segregated and non-segregated prisoners on past and current offence history was found. Similar non-significant findings were obtained on history of disciplinary infractions ($\chi^2 (1, N = 84) = 1.93, p = 0.17$), escape/UAL ($\chi^2 (1, N = 91) = 0.95, p = 0.33$), and failure on conditional releases ($\chi^2 (1, N = 91) = 0.95, p = 0.39$).

With respect to session one segregated prisoners ($n = 83$), nine were serving life sentences (11%). The average sentence length (excluding life sentences; $n = 74$) imposed by the courts for their index offence was 6.97 years.

Of the 53 session one non-segregated prisoners, 13 prisoners were serving life sentences (25%). The average sentence length (excluding life sentences; $n = 40$) imposed by the courts for their index offence was 5.98 years.

Case needs. Segregated and non-segregated prisoners did not significantly differ on any of the OIA need domains (Employment, Marital/Family, Associates, Substance Abuse, Community Functioning, Personal/Emotional, and Attitude).

SIR Scale. Using t-tests, segregated prisoners were found to be higher risk of recidivism ($M = -8.26$) than non-segregated prisoners ($M = -1.07, t (110) = 4.70, p < .001$).

I.Q. Estimates of I.Q. from the Shipley Institute of Living Scale-Revised (Shipley 1940) were compared for segregated and non-segregated prisoners. Non-segregated prisoners ($M = 97.25$) possessed higher estimates of I.Q. than segregated prisoners ($M = 89.70, t (131) = 2.93, p < .01$).

Personality. Table 4 shows the differences between segregated and non-segregated prisoners on the "big five" personality constructs as assessed by the NEO Personality Inventory (NEO-FFI; Costa and McCrae 1992). Segregated prisoners scored higher

on Neuroticism ($t(125) = 3.73, p < .001$), and lower on Extraversion ($t(129) = 2.26, p < .05$), Openness ($t(127) = 3.09, p < .01$), Agreeableness ($t(121) = 2.99, p < .01$) and Conscientiousness ($t(127) = 3.54, p < .001$) than non-segregated prisoners.

Table 4
NEO-FFI sub-scores for segregated and non-segregated prisoners

	Seg. (n=83)	Non-seg. (n=53)	$t(df)$
Factors (T-Scores)	<i>M</i>	<i>M</i>	
Neuroticism***	54.9	48.4	3.73 (125)
Extraversion*	45.9	50.0	2.26 (129)
Openness**	49.0	53.5	3.09 (127)
Agreeableness**	41.6	47.4	2.99 (121)
Conscientiousness***	45.2	51.2	3.54 (127)

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

Social desirability. A between/within-subject repeated measures univariate analysis was performed on the sample of 60 prisoners who completed all three sessions (segregated ($n = 23$) and non-segregated prisoners ($n = 37$)) using the Balanced Inventory of Desirable Responding (BIDR, short form; Paulhus 1984) as a dependent variable (DV). Non-segregated prisoners showed significant higher scores of impression management and self-deception than segregated prisoners ($F(1, 58) = 11.36, p < .001, \eta^2 = .16$).

Segregated and non-segregated prisoners displayed significantly increased scores on impression management and self-deception across time ($F(2, 116) = 3.68, p < .05, \eta^2 = .06$). The ANOVA, however, revealed no significant Time (i.e., session 1, 2 & 3) by Group interaction.

Mental health and psychological functioning: Prisoners who completed three sessions

Between/within-subject repeated measures univariate analyses were performed on the sample of 60 prisoners who completed all three sessions (segregated ($n = 23$) and non-

segregated prisoners ($n = 37$) using the eight measures as dependent variables (DV's). Table 5 presents the means and respective standard deviations for each of the eight measures for all three sessions.

Table 5
Session means of measures for segregated ($n = 23$) and non-segregated ($n = 37$) prisoners who completed all three sessions (S.1, S.2, and S.3)

DV	Group	Means (SD)		
		S.1	S.2	S.3
Aggression Questionnaire	Seg.	77.6 (14.2)	71.5 (17.7)	72.1 (21.2)
	Non-seg.	68.2 (14.2)	66.9 (18.6)	65.5 (20.0)
Beck Depression Inventory	Seg.	8.8 (7.3)	6.3 (6.0)	6.6 (5.5)
	Non-seg.	5.5 (5.6)	4.6 (5.8)	3.9 (5.0)
Brief Symptom Inventory	Seg.	0.92 (0.18)	0.62 (0.39)	0.62 (0.44)
	Non-seg.	0.58 (0.46)	0.44 (0.42)	0.38 (0.40)
HPSI	Seg.	52.4 (15.6)	46.7 (11.6)	49.3 (11.2)
	Non-seg.	40.9 (12.1)	39.2 (10.5)	37.8 (12.6)
Hopelessness Scale	Seg.	5.3 (4.6)	3.6 (3.6)	4.3 (4.3)
	Non-seg.	4.3 (4.3)	3.1 (4.4)	2.8 (4.3)
State-trait Anxiety Inventory	Seg.	13.4 (4.4)	12.1 (4.0)	13.1 (4.5)
	Non-seg.	12.0 (3.4)	9.8 (3.5)	9.6 (3.3)
WAIS Digit Span ¹	Seg.	8.7 (2.9)	9.5 (2.8)	9.5 (3.1)
	Non-seg.	9.6 (2.5)	9.5 (2.6)	10.1 (2.3)
WAIS Digit Symbol ¹	Seg.	7.8 (2.4)	9.0 (3.1)	9.4 (3.3)
	Non-seg.	8.7 (2.3)	9.9 (3.0)	10.5 (2.9)

Note. ¹ Indicates scaled scores.

Aggression Questionnaire (AQ). The Analysis of Variance (ANOVA) on the AQ (Buss and Perry 1992) revealed no significant main effect or interaction.

Beck Depression Inventory - Short Form (BDI). Both segregated and non-segregated prisoners reported significantly fewer depressive symptoms across time ($F(2, 116) = 8.3, p < .001$,

$\eta^2 = .13$) on the BDI (Beck and Beck 1972). The ANOVA revealed no significant interaction (Time by Group).

Brief Symptom Inventory (BSI). Segregated prisoners reported significantly more depressive symptoms than non-segregated prisoners ($F(1,58) = 5.67, p < .05, \eta^2 = .09$) on the BSI (Derogatis 1992). In addition, both segregated and non-segregated prisoners reported significantly fewer depressive symptoms across time ($F(2,116) = 19.57, p < .001, \eta^2 = .25$). The ANOVA revealed no significant interaction.

Holden Psychological Screening Inventory (HPSI). Segregated prisoners reported significantly more problems in psychosocial adjustment than non-segregated prisoners ($F(1,58) = 11.40, p < .001, \eta^2 = .16$) on the HPSI (Holden, Mendonca, Mazmanian, and Reddon 1992). As well, segregated and non-segregated prisoners reported significantly fewer problems in psychosocial adjustment across time ($F(2,116) = 6.27, p < .01, \eta^2 = .10$). The ANOVA revealed no significant interaction.

Hopelessness Scale (HS). Segregated and non-segregated prisoners did not significantly differ on the HS (Beck and Steer 1988). Segregated and non-segregated prisoners indicated significantly less hopelessness across time ($F(2,116) = 10.19, p < .001, \eta^2 = .15$). The ANOVA revealed no significant interaction.

State-Trait Anxiety Inventory (STAI). Segregated prisoner displayed significantly more state anxiety than non-segregated prisoners ($F(1,58) = 8.09, p < .01, \eta^2 = .12$) on the STAI (Spielberger 1983). Further, segregated and non-segregated prisoners displayed significantly less state anxiety across time ($F(2,116) = 7.63, p < .001, \eta^2 = .11$). The ANOVA revealed no significant interaction.

WAIS Digit Span and Digit Symbol. Performance on the Digit Symbol improved significantly across time ($F(2,116) = 5.44, p < .01, \eta^2 = .09$). The ANOVA revealed no significant interaction. Performance on the Digit Symbol improved significantly across time ($F(2,116) = 22.56, p < .001, \eta^2 = .28$). The ANOVA revealed no significant interaction.

Segregated prisoners who completed three sessions versus segregated prisoners who completed one or two sessions

It was hypothesized that prisoners who remain in segregation for longer periods of time would display more mental health and psychological functioning problems than those who are more quickly reintegrated into the general prisoner population. Therefore, using the eight measures as DV's, t-tests were performed to evaluate whether prisoners who stayed in segregation for all three sessions ($n = 23$) differed from segregated prisoners who were released or transferred after session one or two ($n = 51$). True attrition cases ($n = 9$) were removed from the analyses because they could have been part of the group of prisoners who stayed in segregation for all three sessions.

No significant difference was found between prisoners who stayed in segregation for all three sessions ($n = 23$) and segregated prisoners who were released or transferred after session one or two ($n = 51$).

Voluntary versus involuntary cases. Using the eight measures as DV's, t-tests were completed to evaluate whether voluntary ($n = 32$) and involuntary ($n = 51$) cases differed in mental health and psychological functioning. Again, no significant difference was found between voluntary and involuntary cases.

Interview assessment

Suicide ideation. Prisoners were asked questions on suicide ideation. Prisoners who completed all three sessions ($n = 60$) were asked if they ever *thought* of committing suicide. At session one, 40% ($n = 9$) of segregated and 33% ($n = 12$) of non-segregated prisoners responded "yes". When asked if they had ever *attempted* suicide, 22% ($n = 5$) of segregated and 29% ($n = 11$) of non-segregated prisoners said "yes".

At each session, prisoners were asked if they had thought of committing suicide within the last week: 17% ($n = 4$) of segregated prisoners answered "yes" at session one, 4% ($n = 1$) at session two, and 4% ($n = 1$) at session three. As for non-segregated prisoners, 14% ($n = 5$), 11% ($n = 4$) and 3% ($n = 1$) answered "yes" respectively.

Segregation experience. Prisoners who completed all three sessions ($n = 60$) were asked if they have ever been placed in segregation in the past. The vast majority of segregated (96%, $n = 22$) and non-segregated prisoners (87%, $n = 32$) reported having been in segregation before. When asked how many times they have been placed in segregation, segregated prisoners ($M = 11.5$) reported almost twice as many times than non-segregated prisoners ($M = 6.3$).

Discussion

Generalization issues

This study represents the most comprehensive empirical review of the psychological effects of administrative segregation in today's Canadian federal correctional context. To begin with, this study applied the rigour of an experimental longitudinal design to a "real" segregation environment. Participants were actual inmates and not volunteers who agreed to be segregated for a fixed period of time. As such, the sample included actual inmates (some with existing psychiatric conditions and others who feared for their personal safety) who were voluntarily or involuntarily placed for periods up to 60 days in administrative segregation pursuant to the current Canadian federal administrative segregation process. They were segregated under "real" conditions of confinement, which included partial isolation and sensory deprivation. In addition, the participants were confronted with all of the uncertainties surrounding their segregation, such as (a) when the segregation period would end, (b) whether they would be transferred to another institution or returned to the general inmate population, and (c) whether their stay in segregation would affect their security classification, chances for parole, or cell assignment. As well, some prisoners may have been confronted with correctional employees who may have had a punitive approach to managing segregated inmates. All these factors potentially affect the experience of segregation and were not considered by studies using students or inmates who voluntarily agreed to be segregated for a fixed period of time. This study, therefore, examined the psychological effects of today's administrative segregation in Canadian federal corrections, and its results cannot be construed as unrealistic extrapolations of scenarios which are too remote from the "real" experience.

Surprisingly, although various forms of administrative segregation have been used for decades (if not centuries), there has been only one longitudinal study previously conducted using an approximation of the empirical approach used in this study. Weinberg (1967) assessed the effects of segregation on 20 inmates who were involuntarily placed in administrative segregation. The study was, however, limited to a segregation period of only five days, and reported a 68% attrition rate among the experimental group. Again, no other longitudinal study has been completed using prisoners in a “real” segregation context.

The fact that the current longitudinal study was conducted with prisoners who were subjected to “real” segregation conditions of confinement clearly enhance its ability to be generalized. Other factors should also be considered when assessing the issue of generalization of findings. First, this study was conducted at several sites, thereby enhancing the degree of external validity.

Second, the penitentiaries selected have historically been perceived as some of the toughest in Canada. These institutions have some of the largest segregation units and heavily rely on administrative segregation to manage their inmate populations. It was therefore expected that segregated prisoners in those penitentiaries would be more likely to be affected by the harsher realities of some of Canada’s toughest penitentiaries.

Third, the participation rate in this study was comparable to studies that employed inmates for subjects and which do not offer any incentive for participation (e.g., money). The true attrition rate among the segregated group was also relatively low (10.8%) for a longitudinal study. It is important to note that none of the attrition was attributable to prisoners being incapable of participating in the study because of episodes of delusion or hallucination or suicide attempts. Although always a concern, the rate and nature of the attrition in this study does not significantly undermine its ability to be generalized.

Fourth, this study relied on multiple assessments of mental health and psychological functioning of prisoners (i.e., externalizing/aggression, internalizing/interpersonal distress, psychiatric symptomatology, and cognitive ability). This approach provided a more comprehensive assessment of potential

psychological effects of administrative segregation, and is consistent with preferred contemporary psychological and psychiatric assessment practices (DSM-IV 1994).

Finally, non-segregated prisoners scored significantly higher on a measure of impression management than segregated prisoners. Arguably, since segregated prisoners did not show significant signs of mental health and psychological deterioration and were more accurate in their responses than non-segregated prisoners, the results of this study are more convincing.

The above factors enhance the level of confidence in the results of this research. There are, however, clear limitations to this study which may reduce the generalizability of the findings. First, a large number of prisoners in both segregated (96%) and non-segregated (87%) groups had previously experienced segregation. Arguably, a previous stay in segregation may have already negatively affected the mental health and psychological functioning of the prisoners in the comparison group. General inmate population prisoners who have never been segregated before in the three selected penitentiaries are rare. Moreover, these prisoners may arguably possess special coping abilities and attributes and be atypical of the general inmate population, making them a poor choice for comparison purposes. Second, the findings are limited to 60 days in administrative segregation, and any extrapolation to lengthier stays would be inappropriate. It is important to note, however, that statistical data collected by the Correctional Service of Canada (Laplante 1998) indicate that during the period of June 1997 to May 1998, 93% of involuntary cases and 69% of voluntary cases were released prior to the 60 day regional review. This fact suggests that a majority of prisoners are segregated for periods of less than 60 days; therefore, the findings of this study are very relevant to the Canadian federal context.

Third, as stated above, the three penitentiaries selected in this study are among the toughest medium and maximum-security institutions in the country. These penitentiaries rely heavily on administrative segregation to control their inmate populations, which are composed of high-risk and high-need federally sentenced prisoners. The applicability of the results from this study should be limited to such inmate populations.

Further, the findings of this study may be less applicable to other jurisdictions, such as the United States, in which segregated prisoners typically remain in administrative or disciplinary segregation for much longer periods of time, and often under harsher conditions of confinement (Coyle 1987; Dowker and Good 1993; Immarigeon 1992; Korn 1988). Finally, it would also be inappropriate to extend the findings of this study to aboriginal (Bertrand 1996) and women prisoners (Korn 1988; Martel 1999). The realities and experiences of women and aboriginal prisoners may affect their ability to adapt and cope with segregation.

Personality

It was expected that certain personality types would react to the segregation experience differently. Although no deterioration was found, differences in personality between segregated and non-segregated prisoners were found. These differences have been suggested in the PC and segregation literature, but have seldom been assessed using standardized measures, such as the NEO Personality Inventory (Gendreau *et al.* 1985; Hodgins and Cote 1991; Rold 1992).

The NEO was developed to operationalize the five-factor model of personality, a representation of the structure of traits developed over the last forty years (Digman 1990). Costa and McCrae (1990) found that since 1985, research using the NEO has demonstrated that the five factors can account for the major dimensions in personality questionnaires designed to measure, *inter alia*, the DSM-III-R personality disorders. Segregated prisoners were found to score higher on Neuroticism (N) than non-segregated prisoners. Costa and McCrae (1992: 14) explained that “the general tendency to experience negative affects such as fear, sadness, embarrassment, anger, guilt, and disgust is the core of the N domain”. They also suggest that neurotic individuals tend to cope more poorly with stress than others. Although segregated prisoners were found to score higher on Neuroticism and may, therefore, be ill equipped to cope with the stress associated with segregation, the findings of this study suggest that they nonetheless adapted and coped well with the segregation experience.

Segregated prisoners scored significantly lower on Extraversion (i.e., less sociable, likely to prefer large groups,

assertive, active, and less talkative), Openness (i.e., less active imagination, sensitivity, attentiveness to inner feelings, intellectual curiosity, and independence of judgment), Agreeableness (i.e., less altruistic and sympathetic to others and eager to help them, and more egocentric, skeptical of others' intentions, and competitive rather than cooperative), and Conscientiousness (i.e., less strong-willed and determined) than non-segregated prisoners. Arguably, these trait patterns depict individuals which have personalities that may bring them at odds with non-segregated prisoners as well as correctional staff. The general inmate population may not tolerate prisoners with such personality patterns. Due to their lack of assertiveness, general tendency to experience negative affects, and overall poorer mental health and psychological functioning, segregated prisoners may be more easily victimized or less apt at adapting and coping with prison life.

Psychological effects

Overall, *both* segregated and non-segregated prisoners reported better mental health and psychological functioning over time. This finding is common in studies which rely on repeated-measures designs and has been primarily attributed to practice effects (Pedhazur 1982). Participants lose interest in answering repeatedly to identical questions and tend to report less problems overtime.

It was hypothesized that as a group, segregated prisoners overall would report greater mental health and psychological functioning problems than non-segregated prisoners. This hypothesis was supported by the fact that segregated prisoners indicated significantly more internalized problems, interpersonal distress and psychiatric symptoms than non-segregated prisoners. Segregated prisoners also displayed significantly more depressive symptoms, problems in psychosocial adjustment, and transient anxiety than non-segregated prisoners. These results are consistent with many cross-sectional and qualitative studies (Brodsky and Scogin 1988; Grassian 1983; Hodgins and Cote 1991; Wormith *et al.* 1988; Rold 1992). It is important to reaffirm that these between group differences may not be attributed to placement in administrative segregation.

The most important questions raised in this study were whether the poor mental health of segregated prisoners was attributable to segregation or whether segregated inmates already were of poorer mental health prior to their segregation. The hypothesis that the mental health and psychological functioning of segregated inmates would deteriorate over a period of 60 days in segregation was not supported. The ANOVA's performed on each of the eight measures did not reveal any deterioration. These results can be interpreted in two ways: (a) segregated prisoners generally adapted and coped well with the conditions of today's Canadian federal administrative segregation; or (b) the segregated inmates did not perceive the conditions of their confinement as threatening or stressful and therefore were not affected by them.

On one hand, there is no shortage of researchers, ourselves included, who have observed or reacted strongly to the difficult conditions of confinement placed upon segregated prisoners (e.g., 23 out of 24 hours of cell confinement, small yard size, lack of programs and services, constant state of idleness, etc.). On the other hand, other researchers have commented on all the distractions, programs, and services that are available in segregation units in Canadian penitentiaries (e.g., TV, radios, books, computers, exercise period often with the company of other prisoners; Suedfeld *et al.* 1982). Moreover, the Task Force Reviewing Administrative Segregation (Kane 1997) was confronted with many correctional staff who thought segregation units were "too comfortable" for prisoners. They often suggested, contrary to current legal and policy provisions, that the conditions of confinement should be made more harsh in order to discourage prisoners from requesting segregation and to provide an "incentive" for segregated inmates to reintegrate into the general inmate population.

Another explanation to account for these results could be that the environment that prisoners were in before segregation was such that it was viewed more negatively than the conditions of confinement in segregation. If that is the case, it suggests that the correctional authorities must take further steps to ensure that the general inmate population is safe and secure.

Nonetheless, regardless of the possible explanations to account for the lack of deterioration, this study is somewhat

encouraging because it provides evidence that segregation for 60 days as currently administered in Canadian penitentiaries does not negatively affect prisoners' mental health and psychological functioning.

Policy issues

It was anticipated that this research would have important policy implications in areas such as: (a) the level and frequency of monitoring and assessment required for inmates in segregation (mandatory *vs.* upon request); (b) programming to reduce mental health deterioration (the need for, and type of, intervention programs); and (c) the adequacy of current assessment strategies (what aspects of psychosocial functioning are important to assess, and which are less affected by segregation). Since detrimental effects were not found, the policy implications are somewhat less significant than anticipated.

First, with regard to monitoring and assessment of Canadian federally sentenced segregated prisoners, psychologists are required by policy to assess segregated prisoners every 30 days, and health care workers and wardens are required by law to make daily visits to segregation units. Although this study revealed no evidence of detrimental effects, the 30-day requirement should be preserved, as well as the daily visits by health care workers and wardens. Arguably, reducing the few contacts segregated prisoners currently enjoy could have negative consequences. It could be that regular contact itself is an important factor reducing the likelihood of deterioration. Moreover, this research only suggests that the possibility of negative effects is likely to be an exception rather than the norm. Since the findings of this study do not preclude in any way the possibility that some prisoners may in fact be negatively affected by segregation, close monitoring should continue.

Due to their overall poorer mental health and psychological functioning, it may be appropriate as a "best practice" for psychologists to meet with all prisoners upon their placement in segregation. This could serve to establish a baseline for subsequent evaluations of mental health and psychological deterioration and to provide support for segregated inmates at times of crisis. In addition, since segregated prisoners were found

to have poorer mental health and psychological functioning, employees working with segregated prisoners may benefit from special training on mental health issues.

Second, it was expected that this research would provide specific areas of mental health and psychological functioning which needed particular attention when conducting assessment and monitoring of segregated prisoners. But again, since this study did not detect detrimental effects, little can be said in the way of policy on what aspects of mental health and psychological functioning should be carefully scrutinized. Some general comments can be made regarding psychological assessments, however.

Currently, psychologists utilize a standard form which highlights general mental health issues (e.g., risk of suicide or self-injury, depression, anxiety, aggression, psychosis, mania) when completing their 30-day assessments. How to assess each component is left to the psychologist's discretion. Typically, psychologists conduct a brief semi-structured interview with the segregated prisoner. It may be appropriate as a "best practice" to conduct more elaborate assessment procedures to ensure that minor or perhaps less obvious deterioration can be detected and documented.

Finally, the findings of this study have programming implications for segregated prisoners. Since segregated prisoners were found to be higher risk cases than non-segregated prisoners, programs delivered to segregated prisoners should be intensive to maximize success (Andrews, Zinger, Hoge, Bonta, Gendreau, and Cullen 1990). Although the primary concern with providing programs to segregated prisoners is to facilitate their reintegration into the general inmate population, if such reintegration cannot occur quickly, intensive treatment programs that target variables that are known to be linked to criminal conduct should be introduced (Andrews and Bonta 1994). The distinct personality patterns of segregated prisoners may be important to consider and assess when delivering treatment programs to them (Andrews *et al.* 1990).

It is obvious that research evaluating the effects of segregation beyond 60 days is needed. Once again, it would be ill advised to

attempt to extrapolate the findings of this study (a) beyond 60 days of administrative segregation, and (2) to other jurisdictions. For example, the findings of this study are somewhat irrelevant to current segregation practices in the United States where prisoners can sometimes be segregated for years for disciplinary infractions with virtually no distractions, human contacts, services, or programs.

The difference between the personality of segregated and non-segregated prisoners is an important finding. Although many have suggested that segregated prisoners' psychological weaknesses and idiosyncratic behaviours were not well tolerated by the general inmate population (Carriere 1989; Gendreau *et al.* 1985; Rold 1992; Wormith *et al.* 1988), the personality of segregated prisoners had seldom been assessed. Whether a distinct personality profile may increase a prisoner's risk of being placed in administrative segregation should be further examined using more comprehensive measures of personality.

Although this research revealed no evidence that administrative segregation for periods of up to 60 days was damaging, the findings of this study should not be used to legitimize the practice of administrative segregation. Administrative segregation remains a management tool which is grossly overused in Canadian penitentiaries. Regardless of whether prisoners adapt and cope well with the segregation experience, it is not healthy for anyone to idle aimlessly in a cell for 23 out of 24 hours a day; it simply is not a constructive way of serving a sentence; and, it is likely to impede attempts to rehabilitate and safely reintegrate prisoners into society.

Although it will always remain a legitimate management tool to deal effectively with problematic situations and individuals, its current use is perhaps symptomatic of correctional authorities' inability to reduce tensions and resolve conflicts in the prison context. Administrative segregation has clearly become the number one way of managing inmates and "doing business". For example, the Correctional Service of Canada (1999) reported that during fiscal year 1998/99, out of an inmate population that averaged 13,131 federally sentenced prisoners, 7,942 placements in administrative segregation took place. Such high reliance on the use of segregation needs to be carefully examined. Moreover,

the costs associated with processing these prisoners in accordance with due process requirement are extraordinary (i.e., paperwork, enhanced security and staffing, and reviews by wardens, Segregation Review Boards and Regional Headquarters, etc.). Clearly, it is time to rethink conflict resolution in Canada's penitentiaries.

Implementing alternative/appropriate dispute resolution processes on a large scale is the most promising initiative to reduce the disproportionate number of segregation cells and units in Canada's federal correctional system. Providing the tools to resolve conflicts and fostering a correctional environment respectful of human rights is the only way to break down this over-reliance on administrative segregation for managing prisoners.

References

- Andrews, D.A., and J. Bonta
1994 The Psychology of Criminal Conduct. Cincinnati: Anderson Publishing.
- Andrews, D.A., I. Zinger, R.D. Hoge, J. Bonta, P. Gendreau, and F.T. Cullen
1990 Does correctional treatment work? A psychologically informed meta-analysis. *Criminology* 28: 369-404.
- Arbour, L. (Commissioner)
1996 Commission of Inquiry into Certain Events at Prison for Women in Kingston. Ottawa: Queen's Printer.
- Barak-Glantz, I.L.
1983 Who's in the "hole"? *Criminal Justice Review* 8: 29-37.
- Beck, A.T. and R.W. Beck
1972 Screening depressed patients in family practice: A rapid technique. *Postgraduate Medicine* 52: 81-85.
- Beck, A.T. and R.A. Steer
1988 Manual for the Beck Hopelessness Scale. New York: Psychological Corporation.
- Benjamin, T.B. and K. Lux
1977 Solitary confinement as punishment. *California Western Law Review* 13: 265-296.
- Benjamin, T.B. and K. Lux
1975 Constitutional and psychological implications of the use of solitary confinement: Experience at the Maine State Prison. *Clearinghouse Review* 9: 83-90.

Bertrand, M.-A.

- 1996 Women in prison, a comparative study. *Caribbean Journal of Criminology and Social Psychology* 1(1): 38-58.

Bexton, W.H., W. Heron, and T.H. Scott

- 1954 Effects of decreased variation in the sensory environment. *Canadian Journal of Psychology* 8: 70-76.

Birkinshaw, P.

- 1981 The control unit regime: Law and order in prison. *Howard Journal* 20: 69-80.

Bonta, J. and P. Gendreau

- 1995 Reexamining the cruel and unusual punishment of prison life. In T.J. Flanagan (ed.), *Long-term Imprisonment: Policy, Science, and Correctional Practice*. Thousand Oaks, CA: Sage.

Brodsky, S.L. and F.R. Scogin

- 1988 Prisoners in protective custody: First data on emotional effects. *Forensic Reports* 1: 267-280.

Buss, A. H. and M. Perry

- 1992 The aggression questionnaire. *Journal of Personality and Social Psychology* 63: 452-459.

Carriere, K.D.

- 1989 Protective custody in Canada: A review of research and policy responses. *Canadian Criminology Forum* 10: 17-25.

Correctional Service of Canada

- 1999 Performance Report: Estimates. Ottawa: Correctional Service of Canada.

Costa, P.T., Jr. and R.R. McCrae

- 1990 Personality disorders and the five-factor model of personality. *Journal of Personality Disorders* 4: 362-371.

Costa, P.T., Jr. and R.R. McCrae

- 1992 Revised NEO Personality Inventory and NEO Five-factor Inventory Professional Manual. Odessa, FL: Psychological Assessment Resources.

Coyle, A.G.

- 1987 The management of dangerous and difficult prisoners. *Howard Journal* 26: 139-152.

Derogatis, L.R.

- 1992 The Brief Symptom Inventory (BSI): Administration, Scoring, and Procedures Manual II. Towson, MD: Clinical Psychometrics Research.

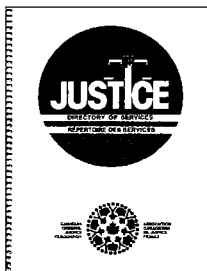
Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)

- 1994 Washington: American Psychiatric Association.

- Digman, J.M.
1990 Personality structure: Emergence of the five-factor model. *Annual Review of Psychology* 41: 417-440.
- Dowker, F. and G. Good
1993 The proliferation of control unit prisons in the United States. *Journal of Prisoners on Prisons* 4: 95-110.
- Ecclestone, C.E.J., P. Gendreau, and C. Knox
1974 Solitary confinement of prisoners: An assessment of its effects on prisoners' personal constructs and adrenocortical activity. *Canadian Journal of Behavioral Science* 6: 178-191.
- Gendreau, P. and J. Bonta
1984 Solitary confinement is not cruel and unusual punishment: People sometimes are! *Canadian Journal of Criminology* 26: 467-478.
- Gendreau, P.E., N. Freedman, G.J.S. Wilde, and G.D. Scott
1968 Stimulation seeking after seven days of perceptual deprivation. *Perceptual and Motor Skills* 26: 547-550.
- Gendreau, P.E., N. Freedman, G.J.S. Wilde, and G.D. Scott
1972 Changes in EEG alpha frequency and evoked response latency during solitary confinement. *Abnormal Psychology* 79: 54-59.
- Gendreau, P.E., J.G. Horton, D.G. Hooper, G.J.S. Freedman, and G.D. Scott
1968 Perceptual deprivation and perceptual skills: Some methodological considerations. *Perceptual and Motor Skills* 27: 57-58.
- Gendreau, P., R. McLean, T. Parsons, R. Drake, and J. Ecclestone
1970 Effect of two days' monotonous confinement on conditioned eyelid frequency and topography. *Perceptual and Motor Skills* 31: 291-293.
- Gendreau, P., M.-C. Tellier, and J.S. Wormith
1985 Protective custody: The emerging crisis within our prisons. *Federal Probation* 44: 55-63.
- Grassian, S.
1983 Psychopathological effects of solitary confinement. *American Journal of Psychiatry* 140: 1450-1454.
- Haney, C.
1993 "Infamous punishment": The psychological consequences of isolation. *National Prison Project Journal* 8: 3-7, 21.
- Hodgins, S. and G. Cote
1991 The mental health of penitentiary prisoners in isolation. *Canadian Journal of Criminology* 33: 175-182.

- Holden, R. R., J.D. Mendonca, D. Mazmanian, and J.R. Reddon
1992 Clinical construct validity of the Holden Psychological Screening Inventory (HPSI). *Journal of Clinical Psychology* 48: 627-633.
- Immarigeon, R.
1992 The marionization of American prisons. *National Prison Project Journal* 7: 1-5.
- Jackson, M.
1983 *Prisoners of Isolation: Solitary Confinement in Canada*. Toronto: University of Toronto Press.
- Kane, D. (Chair)
1997 *Commitment to Legal Compliance, Fair Decisions and Effective Results*. Ottawa: Correctional Service Canada.
- Korn, Richard
1988 The effects of confinement in the high security unit at Lexington. *Social Justice* 15(1): 1-20.
- Laplante, J.
1998 Paper presented at the Conference entitled *The Reasonable Alternative*. Kingston: Correctional Service of Canada.
- Luise, Maria A.
1989 Solitary confinement: Legal and psychological considerations. *New England Journal on Criminal and Civil Confinement* 15: 301-324.
- Martel, J.
1999 *Solititudes and Cold Storage: Women's Journeys of Endurance in Segregation*. Edmonton: Elizabeth Fry Society of Edmonton.
- Miller, H.A. and G.R. Young
1997 Prison segregation: Administrative detention remedy or mental health problem. *Criminal Behaviour and Mental Health* 7: 85-94.
- Motiuk, L.L. and K. Blanchette
1997 *Case Characteristics of Segregated Prisoners in Federal Corrections*. Ottawa: Correctional Service of Canada.
- Nuffield, J.
1982 *Parole Decision-making in Canada: Research towards Decisions Guidelines*. Ottawa: Ministry of Supply and Services Canada.
- Paulhus, D.L.
1984 Two-component models of socially desirable responding. *Journal of Personality and Social Psychology* 46: 598-609.
- Pedhazur, E.J.
1982 *Multiple Regression in Behavioral Research*. New York: Harcourt Brace College Publishers.

- Pierson, T.A.
1988 Use of protective custody: How different systems respond. *Corrections Today* 50: 150, 152, 154.
- Rold, W.J.
1992 Consideration of mental health factors in prisoner discipline. *Journal of Prison and Jail Health* 11: 41-49.
- Shipley, W.C.
1940 A self-administering scale for measuring intellectual impairment and deterioration. *Journal of Psychology* 9: 371-377.
- Spielberger, C.D.
1983 *Manual for the State-Trait Anxiety Inventory STAI (Form Y)*. Palo Alto, CA: Consulting Psychologist Press.
- Suedfeld, P., C. Ramirez, J. Deaton, and G. Baker-Brown
1982 Reactions and attributes of prisoners in solitary confinement. *Criminal Justice and Behavior* 9(3): 303-340.
- Tellier, C., S. Wormith, and P. Gendreau
1989 *Protective Custody: The Emerging Crisis Within the Prison System (Working Paper)*. Ottawa: Solicitor General Canada.
- Vantour, J.A.
1975 *Report of the Study Group on Dissociation*. Ottawa: Solicitor General Canada.
- Wadeson, H. and W.T. Carpenter
1976 Impact of seclusion room experience. *Journal of Nervous Mental Disease* 163: 318-328.
- Walters, R.H., J.E. Callagan, and A.F. Newman
1963 Effect of solitary confinement on prisoners. *American Journal of Psychiatry* 119: 771-773.
- Weinberg, M.M.
1967 *Effects of Partial Sensory Deprivation on Involuntary Subjects*. Unpublished doctoral dissertation, Michigan State University.
- Wormith, J.S., M.-C. Tellier, and P. Gendreau
1988 Characteristics of protective custody prisoners in a provincial correctional centre. *Canadian Journal of Criminology* 30: 39-58.
- Zubek, J.P., L. Bayer, and J.M. Shephard
1969 Relative effects of prolonged social isolation and confinement: Behavioral and EEG changes. *Journal of Abnormal Psychology* 74: 625-631.



The whole justice system in a book! There is no other book like it in Canada. Complete addresses, telephone and fax numbers, e-mail and web pages, and names of persons responsible for all justice services in Canada: police commissions, RCMP divisions, legal aid services, crime compensation boards, ombudsmen, courts, parole boards, correctional services for juveniles and adults, prisoner aid agencies, courtworkers, voluntary organisations, etc. (ISSN: 0225-4115) CCJA member: \$20.00 + GST = \$21.40; CCJA member resident of Ontario: \$20.00 + GST + OST = \$23.00; non-member: \$25.00 + GST = \$26.75; non-member resident of Ontario: \$25.00 + GST + OST = \$28.75; Foreign: \$25.00. **Canadian Criminal Justice Association**, 383 Parkdale Avenue, #304, Ottawa, Ontario K1Y 4R4. Telephone: (613) 725-3715, Fax: (613) 725-3720, E-mail: ccja@istar.ca.

PLEASE ADD \$3.00 to the above prices for postage and handling.
All orders must be prepaid

PLEASE SEND ME () COPY(IES) OF THE DIRECTORY.

I HAVE ENCLOSED THE APPROPRIATE AMOUNT (_____ \$) BASED ON APPLICATION RATES LISTED ABOVE.

with cheque to: **Canadian Criminal Justice Association**
 383 Parkdale Avenue, #304, Ottawa, Ontario, K1Y 4R4
 Telephone: (613) 725-3715 – Fax: (613) 725-3720 – E-mail: ccja@istar.ca

or with credit card: VISA MASTERCARD DINERS CLUB/enROUTE
 AMERICAN EXPRESS

_____ (Card No.) _____ (Expiry Date)

_____ (Name on the card, please print) _____ (Signature)

Name: _____

Address: _____

City: _____ Prov.: _____

Postal Code: _____ Phone: _____