A Longitudinal Study of Administrative Segregation

Maureen L. O'Keefe, MA, Kelli J. Klebe, PhD, Jeffrey Metzner, MD, Joel Dvoskin, PhD, Jamie Fellner, JD, and Alysha Stucker, BA

The use of administrative segregation for inmates with and without mental illness has generated considerable criticism. Segregated inmates are locked in single cells for 23 hours per day, are subjected to rigorous security procedures, and have restricted access to programs. In this study, we examined whether inmates in segregation would show greater deterioration over time on psychological symptoms than would comparison offenders. The subjects were male inmates, with and without mental illness, in administrative segregation, general population, or special-needs prison. Subjects completed the Brief Symptom Inventory at regular intervals for one year. Results showed differentiation between groups at the outset and statistically significant but small positive change over time across all groups. All groups showed the same change pattern such that there was not the hypothesized differential change of inmates within administrative segregation. This study advances the empirical research, but replication research is needed to make a better determination of whether and under what conditions harm may or may not occur to inmates in solitary confinement.

J Am Acad Psychiatry Law 41:49–60, 2013

Placement of offenders in long-term administrative segregation (AS; also called Ad Seg), particularly those with serious mental illness, has been subject to considerable criticism. AS generally involves locking an inmate in a cell for 23 hours per day, with out-of-cell time occurring with significant security restrictions (e.g., hands and ankles cuffed) and escort by two correctional officers. Critics have argued that the conditions of AS confinement exacerbate symptoms of mental illness and create mental illness in those who previously had no such disorders.[1,2] The use of AS across the country has persisted as a corrections management tool despite litigation, although in some states its use in inmates with mental illness is no longer permitted. Prior research has shed light on the problem, but because of methodological limitations, the core questions remain unresolved.[3] Researchers have been unable to settle the question of whether the high rates of mental illness found in AS are caused by this harsh environment 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.

In 1983, Grassian[4] described psychopathological features associated with rigidly imposed solitary confinement that he believed formed a clinical syndrome. He interviewed 14 plaintiffs in a conditions-of-confinement case and described clinical observations resulting from those interviews. He noted perceptual changes, affective disturbances, cognitive difficulties, disturbed thought content, and impulse-control problems that subsided after release from such confinement. In more recent research, Haney[5] found elevated symptoms of psychological trauma (e.g., anxiety, headaches, and impending nervous breakdown) and psychopathological features (e.g., ruminations, social withdrawal, and irrational anger) among 100 security housing unit (SHU) prisoners, compared with such symptoms in national adult population samples. This constellation of symptoms composes the primary features of what has

Ms. O'Keefe is Research Director, Office of Planning and Analysis, Colorado Department of Corrections, Colorado Springs, CO. Dr. Klebe is Associate Professor of Psychology, and Ms. Stucker is Professional Research Assistant, University of Colorado, Colorado Springs, CO. Dr. Metzner is Clinical Professor of Psychiatry, School of Medicine, University of Colorado, Denver, CO. Dr. Dvoskin is Assistant Clinical Professor, University of Arizona College of Medicine, Tucson, AZ. Mr. Fellner is Senior Advisor, U.S. Program, Human Rights Watch, New York, NY. This project was supported by Grant 2006-IJ-CS-0015 awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. Points of view in this article are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice. Address correspondence to: Maureen L. O'Keefe, Office of Planning and Analysis, Colorado Department of Corrections, 2862 South Circle Drive, Colorado Springs, CO 80906. E-mail: maureen.okeefe@state.co.us.

Disclosures of financial or other potential conflicts of interest: None.

been coined the SHU syndrome in the wake of *Madrid v. Gomez*,[6] a class-action suit that successfully challenged conditions of confinement in a California supermax prison.

Research on the effects of AS has been criticized for being deficient in quality designs that allow one to rule out plausible alternative explanations.[7–10] Because of the lack of a comparison group, some frequently cited studies are demonstrations of the potential impacts of AS.[4,11,12] Other researchers have used a variety of comparison groups including non-inmate populations and norms, inmate volunteers, general population prisoners, and inmates in different security levels.[5,13,14] Most, although not all, of these studies found that inmates in AS demonstrate higher levels of psychological distress. These cross-sectional studies lack the ability to attribute differences to the conditions of confinement because of the potential for pre-existing differences, including psychological impairments that may have existed before entry into AS.

There have been few longitudinal studies about the effects of segregation. In early studies, Gendreau and colleagues[15–19] used repeated-measures experimental designs over periods of up to 10 days. Few negative impacts of segregation were found over these brief periods. Although the use of a repeated-measures experimental paradigm improves over cross-sectional studies that may have selection bias, the short confinement periods are unrealistic for providing information on the effects of segregation as currently used in U.S. prisons.

In two studies, inmates were followed for longer periods after placement in segregation.[10,20] Andersen *et al.*[20] studied participants over a 4-month period, but most of the participants had data for less than a month. Zinger *et al.*[10] observed inmates over a 60-day period. Each study demonstrated that segregated populations had more psychological disorders at the start than did the comparison subjects, but had conflicting evidence on whether conditions worsened over time. Because these studies had high refusal and attrition rates, the conclusions must be interpreted cautiously. Further longitudinal studies are needed to sort out these discrepancies and understand the long-term impacts of segregation.

We hypothesized that inmates in segregation would develop psychological symptoms consistent with the SHU syndrome and that they would deteriorate over time relative to comparison offenders. In addition, we hypothesized that segregated offenders, with or without mental illness, would deteriorate over time, but the rate at which it occurred would be more rapid and more extreme in the mentally ill.

Method

Setting

The Colorado Department of Corrections managed 19,279 inmates in 25 state and 7 private prisons at the start of data collection. Colorado State Penitentiary (CSP) was one of four state prisons designed to hold AS-classified offenders. As a 756-bed male facility, CSP was the largest and only dedicated AS facility in the state. Therefore, any study participants who were classified as AS were waitlisted and placed in CSP.

AS is the most secure and restrictive of five security classification levels in Colorado. Placement is determined through an administrative action (during a hearing) that is separate and distinct from both the usual classification system and the disciplinary system. Classification to the other four levels is determined through a scored instrument, and the disciplinary process is a punitive response to a finding of guilt for an institutional rule violation that may result in punitive segregation for up to 60 days. AS is of longer duration and is used for management purposes; Colorado did not place protective-custody inmates or new prison inmates into AS at the time of the study. In addition, prehearing segregation may occur immediately after a serious incident, for safety and security reasons. Thus, in the time leading up to and during their AS hearing, inmates have typically been in segregation. All segregation cells in Colorado are single occupancy, and inmates may only leave their cells with a two-person escort while in full restraints.

Offenders reclassified to AS remain in a punitive segregation bed until an AS bed becomes available. Once transferred to CSP, inmates have increased access to services compared with punitive segregation, such as library, education courses, and treatment programs. Most services are provided cellside, including meals, medications, library, and even programs. Each cell is equipped with an intercom system for on-demand communication between the inmate and the unit's control center. Officers also make rounds every 30 minutes to perform a visual check. Inmates are permitted to leave their cells for at least

one hour of recreation five times per week and to shower for 15 minutes three times per week. CSP provides incentive-based behavior modification and cognitive programs. The incentive-based programming consists of three quality-of-life levels, bringing more privileges with each level earned. Once inmates progress from level one (usually after the first seven days), they are permitted televisions in their cells. To progress from CSP, every offender must successfully complete three televised cognitive classes, each lasting three months. A variety of mental health services are available within the facility, including monthly cell-front rounds, individual counseling sessions, psychiatry, and crisis management.

Colorado has a dedicated 255-bed special needs (SN) prison for inmates with acute psychiatric symptoms who cannot be managed in the general prison population. When inmates are admitted to the SN prison, they are held at a highly restricted level while undergoing intake and assessment, and they quickly progress to less restrictive environments unless their behavior prohibits progression. When CSP and the SN prison were excluded because of their unique missions, 26 male general population (GP) prisons remained. GP inmates have access to significant out-of-cell time (e.g., >10 hours/day), jobs, and programming in contrast to AS inmates.

Subjects

Subjects included male inmates placed in AS and comparison inmates drawn from 10 GP facilities housing higher security inmates. Placement into AS or GP conditions occurred as a function of routine prison operations in the context of an inmate's being charged with a prison rule infraction. Following an AS hearing, inmates were waitlisted for CSP if prison officials determined that AS placement was warranted or were returned to GP if not reclassified to AS. GP comparison subjects also included disruptive inmates at high risk of AS placement who were transferred to a diversionary program; the program discontinued shortly after the study commenced, so only 17 percent of GP subjects were identified this way.

Inmates in both study conditions (AS, GP) were classified into two groups, those with mental illness (MI) and those with no mental illness (NMI), giving four study groups. These groups were based on the prison system's existing mental health classification system, which takes into account clinical diagnosis, acuity of symptoms, and consumption of resources. The primary diagnoses that met criteria for elevated mental health ratings are bipolar mood disorders, major depressive disorder, depressive disorder not otherwise specified, dysthymia, schizophrenia and other psychotic disorders, and posttraumatic stress disorder. Inmates with serious mental illnesses placed in the SN prison comprised a fifth study group (labeled SN MI). SN MI inmates were included only if they had institutional histories of disciplinary violations. The AS NMI group's primary comparison group was the GP NMI group, whereas the AS MI group was compared with both the GP MI and the SN MI groups.

Figure 1 illustrates the eligibility and selection of inmates for participation in the five study groups. Because the focus was long-term segregation, 226 inmates were excluded for having less than 15 months remaining on their sentences. Eighteen were excluded because of illiteracy or language barriers. Before contact by the researcher, inmates with mental illness were reviewed by clinicians, and two SN MI inmates were determined to be unable to comprehend the consent form. Subjects were selected from the remaining inmates by using nonprobability sampling according to inmates' proximity by timing or location to others who could be included in the study.

A total of 302 male inmates were approached to participate in the study. After complete description of the study to the subjects, written informed consent was obtained from 270 inmates. Thirty refused to participate, two were removed for inappropriate behavior toward the researcher, and 23 later withdrew their consent (data provided up to withdrawal were included). The participation of subjects within each group at each testing interval is shown in Figure 2.

Subjects' ages ranged from 17 to 59 (mean (M) = 31.8; standard deviation (SD) = 9.1). The racial/ethnic breakdown was 40 percent white, 36 percent Hispanic, 18 percent African American, 4 percent Native American, and 1 percent Asian. There were few or no significant group differences in the following contrasts: AS subjects versus eligible pool, those who refused versus those who participated, and subjects who completed every testing session versus those who did not (analyses can be found in the report for the funding agency[21]).
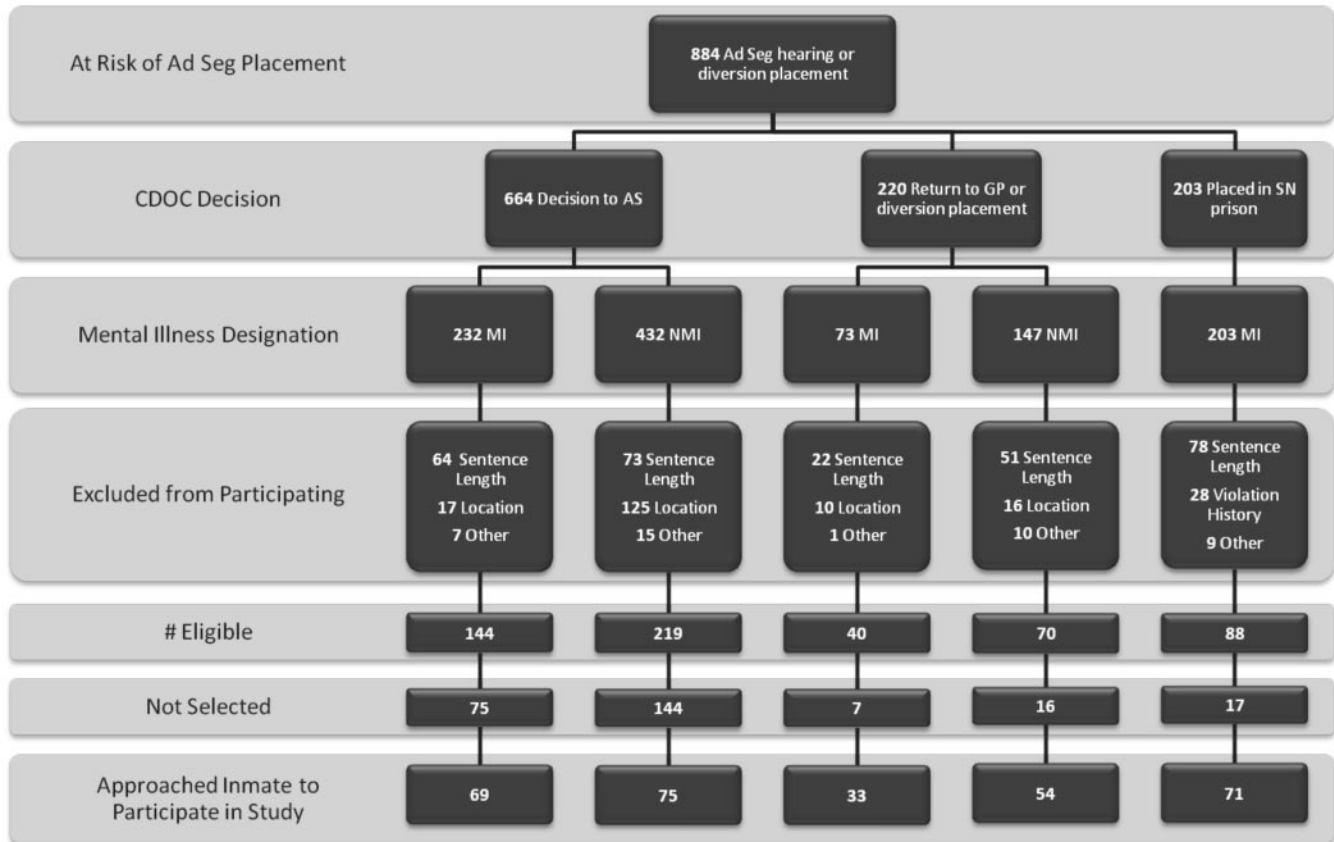
Figure 1. Eligibility and selection of subjects.

Measures

The Brief Symptom Inventory[22] (BSI) was selected to measure the variety of psychiatric constructs hypothesized to be affected by AS, with specific reference to the SHU syndrome. Subjects were administered a battery of 12 psychological and cognitive tests as part of a larger study; however, the BSI is reported here because it covers a broad range of psychological symptoms (measuring constructs that overlapped with the other instruments) and yielded the same results as the other tests.[21] As a standardized paper-and-pencil test, the BSI provides objective self-report data. The measure was also selected for its demonstrated reliability and validity, testing length, and ease of administration within the prison setting (e.g., no specialized equipment, no contact, reading level).

The BSI 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; and positive symptom distress index[23]). Respondents are asked to rate the degree of distress experienced over the past week, using a five-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 than on the nature of it.[23] A minimum sixth-grade reading ability is needed to complete this measure, and it generally takes 10 minutes to complete.

The BSI demonstrated adequate reliability across forensic populations with internal consistency reliabilities of 0.52 to 0.86[10,24] and item-total correlations of 0.73 to 0.91[23]; two-week test-retest reliability was 0.90 for the GSI.[22] Convergent validity estimates of the BSI ranged from 0.30 to 0.72 compared with clusters on the MMPI[23,25] and from 0.49
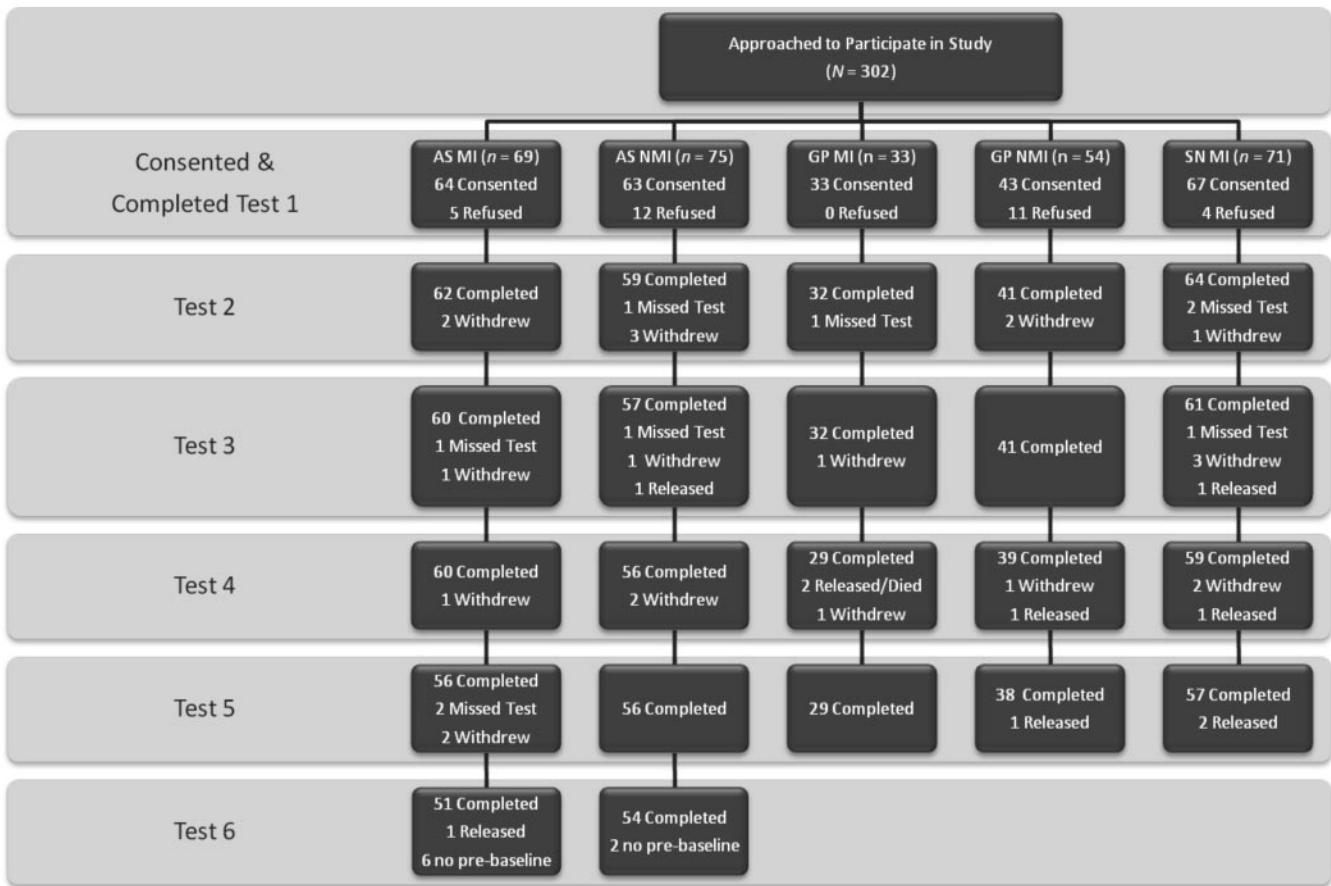
Figure 2. Flow of subjects through study.

to 0.69 in comparison to scales on the Brief Psychiatric Rating Scale.[26] In the present study, internal consistency estimates for scores on the BSI subscales ranged between 0.71 and 0.91 (M = 0.85), and test-retest reliability estimates ranged between 0.53 and 0.79 (M = 0.72). When the data from the complete study were included, scores on the BSI subscales showed reasonable convergent validity, as correlations with other self-report measures of the same constructs ranged between 0.15 and 0.89 (M = 0.56), but there were lower validity estimates with staff reports with correlations ranging between −0.01 and 0.43 (M = 0.23).

Procedures

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 pending AS hearings by prison staff when the offender received notice and of SN prison placements before transfers. Research staff screened electronic inmate files for eligibility, including mental health status, time remaining on sentence, and literacy.

The BSI was administered by a female field researcher. 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 entailed a noncontact booth for inmates in segregation. 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 compensated $10 per testing session (subject to a $3 fee for restitution payment plus a $5 fee if an inmate had a negative bank balance) for a maximum of $60 for those who completed six sessions. They were advised that the purpose was to learn about prison adjustment and that inmates across the state were participating in the study. At the time of consent, the initial test battery was administered.

Subjects were added to the study at the time of their AS hearings, usually while in segregation, or SN prison placement (baseline assessment). The GP and SN groups were tested at approximately three-month intervals for five testing sessions. Because of the long waitlist for the AS facility, AS subjects had their second test after placement and approximately every three months thereafter, for a total of six testing sessions. The median between first and second testing intervals was 89 days (range, 41–190).

Statistical Analyses

Multilevel modeling for repeated-measures data was used to determine the underlying function of change and to determine whether groups changed in different ways over time. The linear mixed model command in SPSS 20 was used for all analyses. BSI scores were positively skewed with a significant number of outliers so that scores were transformed by using a square root transformation. This transformation reduced the number of outliers in the data and each distribution was less skewed. Because time between assessments varied for participants, time was coded as the number of months from baseline, with the baseline coded as time 0. BSI scores were centered at the mean of the baseline period using all participants. These two parameterizations allowed the intercept to be interpreted as an estimate of the score at the initial assessment interval. We followed the basic procedure and steps for testing multilevel models suggested by West[27] and Heck *et al.*[28] Multiple models were fit to the data to determine if a linear, quadratic, cubic, or logarithmic function best explained the underlying change over time. This method gave 12 models to fit. To determine the best fitting model, we used Akaike's Information Criterion (AIC) and assessed nested models with the chi-square test for differences between log-likelihood ratio values (-2 restricted log-likelihood), with maximum-likelihood estimation. Additional parameter estimates were examined with the Wald z test to determine statistical significance. Each model parameter was also assessed to determine if it should be treated as a random or fixed effect by examining whether the variance and covariance elements were significantly different from zero, as well as by comparing model fit when the elements were treated as random versus fixed. The repeated-measures error covariance matrix was fit by using an autoregressive structure, following comparison of other possible structures, and the random

coefficients covariance matrix was estimated by using an unstructured form.

Once the best fitting change function was determined, multiple models were estimated to test the hypotheses that the AS groups changed in different ways than the comparison groups did. Thus, we have two levels, with the first level estimating the intraindividual change over time and the second level estimating interindividual differences in function parameters. Three different sets of models were assessed to test the hypotheses. For the first two sets, all study groups were used, and the first five time periods were used. One model was coded so that the AS MI parameter estimates could be compared with the other groups' estimates, and the other model was coded so that comparisons of the AS NMI parameter estimates could be made with the other groups' estimates. These two models are equivalent except for the tests of each group with the other groups. A third set of models used only the AS groups and all six time assessments to compare the change over time for the AS groups using all data (the three other groups were not assessed at Time 6). All consenting participants were used regardless of the number of assessments that were completed ($n = 270$).

Results

Table 1 provides the descriptive statistics for the original BSI subscale scores and the centered square-root-transformed scores, along with normative means. The original scores are provided so that comparisons could be made with normative BSI data, and the transformed centered data are used for the mixed-level analysis. Initially, all BSI subscales were used; however, the results were the same for all subscales; that is, the conclusions about change and group differences at initial value and change over time were the same. Thus, for ease of interpretation and understanding, only the BSI global score index (GSI) is reported (full results are available from the authors). Figure 3 provides a graphic representation of the best fitting function for each BSI subscale and GSI score.

Table 2 provides the fit statistics for the 12 estimated models used to estimate the best fitting function for transformed GSI scores (the Level 1 models) for all groups, when using the first five assessments. Models differed in the underlying mathematical function (e.g. linear, quadratic) and whether parameters were treated as fixed or random. Random coef-

Table 1 Summary Statistics for BSI Subscales for Raw Scores and Transformed Centered Scores

| | Assessment | | | | | | | | | | | | Nonpatient Norms | Outpatient Norms |
| | 0 | | 1 | | 2 | | 3 | | 4 | | 5 | | | |
BSI Scale	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M (SD)	M (SD)
Original Scale														
Anxiety	1.05	0.99	0.86	0.96	0.87	0.98	0.82	0.94	0.75	0.89	0.62	0.82	0.26 (0.31)	1.51 (0.95)
Depression	1.34	1.07	1.12	1.06	1.12	1.10	1.06	1.05	1.02	1.05	0.84	0.91	0.21 (0.33)	1.65 (1.11)
Hostility	1.03	0.99	0.84	0.89	0.91	0.95	0.94	1.06	0.88	0.94	0.78	0.94	0.34 (0.40)	1.07 (0.90)
Interpersonal sensitivity	1.20	1.05	1.00	0.96	0.99	1.05	0.95	1.02	0.95	1.01	0.78	0.94	0.24 (0.38)	1.48 (1.06)
Obsessive-compulsive	1.34	1.01	1.19	1.03	1.18	1.07	1.15	1.05	1.07	0.99	0.92	0.92	0.37 (0.41)	1.53 (0.98)
Paranoid ideation	1.52	1.01	1.28	0.98	1.29	1.02	1.26	1.02	1.21	1.02	1.17	0.99	0.33 (0.41)	1.06 (0.93)
Phobic anxiety	0.77	0.95	0.62	0.88	0.66	0.92	0.59	0.89	0.56	0.79	0.42	0.80	0.11 (0.25)	0.79 (0.84)
Psychoticism	1.28	1.02	1.07	0.95	1.04	0.99	0.97	0.93	0.96	0.93	0.83	0.86	0.15 (0.27)	1.12 (0.84)
Somatization	0.82	0.87	0.65	0.75	0.69	0.82	0.62	0.80	0.61	0.79	0.45	0.67	0.23 (0.32)	0.67 (0.71)
GSI global score	1.15	0.82	0.96	0.80	0.97	0.86	0.92	0.83	0.88	0.79	0.75	0.72	0.25 (0.24)	1.20 (0.70)
Transformed														
Anxiety	0.00	0.55	−0.15	0.58	−0.13	0.58	−0.16	0.57	−0.21	0.56	−0.29	0.54		
Depression	0.00	0.55	−0.14	0.59	−0.16	0.62	−0.18	0.60	−0.21	0.60	−0.31	0.58		
Hostility	0.00	0.54	−0.11	0.53	−0.08	0.55	−0.09	0.59	−0.10	0.55	−0.17	0.56		
Interpersonal sensitivity	0.00	0.58	−0.11	0.57	−0.16	0.62	−0.16	0.61	−0.17	0.61	−0.27	0.59		
Obsessive-compulsive	0.00	0.54	−0.10	0.58	−0.11	0.59	−0.12	0.59	−0.16	0.57	−0.24	0.55		
Paranoid ideation	0.00	0.48	−0.13	0.52	−0.13	0.54	−0.15	0.54	−0.19	0.57	−0.21	0.56		
Phobic anxiety	0.00	0.59	−0.10	0.57	−0.10	0.60	−0.14	0.58	−0.14	0.55	−0.27	0.53		
Psychoticism	0.00	0.53	−0.13	0.56	−0.16	0.57	−0.18	0.56	−0.20	0.57	−0.27	0.54		
Somatization	0.00	0.52	−0.13	0.52	−0.10	0.53	−0.16	0.53	−0.17	0.53	−0.26	0.47		
GSI global score	0.00	0.41	−0.11	0.44	−0.12	0.46	−0.14	0.45	−0.16	0.46	−0.24	0.44		
Time (mo)	0	NA	3.1	.95	6.0	.92	9.1	.92	12.1	.97	15.4	1.37		

BSI subscale norms were taken from the BSI manual using male nonpatients and psychiatric outpatients (Ref. 22, p 35). Transformed data were centered at first assessment for entire sample with a square-root transformation to normalize the positively skewed distributions. Time is the mean number of months (with standard deviation) taken from the baseline assessment.

ficients imply that individuals differ on specific values for parameters and allow for intraindividual change. Fixed coefficients provide information about the mean function that fits the data. Comparisons between nested models demonstrated that a nonlinear function was most appropriate for the change over time and that coefficients should be random. Based on the AIC statistics, the logarithmic function was selected as the change model that best fit the data. A logarithmic model implies initial fast change

Figure 3. Logarithmic change functions for BSI subscales. Time is the average number of months between assessments.

with change slowing over time. The intercept estimate was −0.01 (SE = .02; p = .74), which was not statistically significant (i.e., not different from zero), as expected, because the data were centered for initial scores for the entire sample. The change parameter was statistically significant (b = −0.06; SE = 0.01, p < .001) and negative, indicating that the scores decreased significantly (i.e., showing improvement on BSI scores) over time (as can be seen in Fig. 3). There was statistically significant variability in the intercept (σ = 0.13; SE = 0.01; Wald z = 9.29; p < .001) and in the change parameter (σ = 0.01; SE = 0.002; Wald z = 4.98; p < .001), but there was not a statistically significant relationship between intercept and change parameters (σ = −0.001; SE = 0.004; Wald z = −0.34; p = .74). Because there was significant variability in the random coefficients, the second-level models were estimated to determine whether this variability could be explained by group membership.
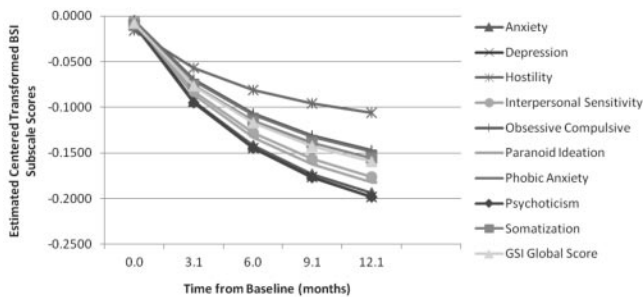
Table 3 provides the results for the second-level models with logarithmic change for GSI global

Table 2 Level 1 Model Statistics for BSI GSI Global Scores to Determine Best Fitting Change Function

Model Estimated	−2LLR	AIC
Cubic polynomial: Int, T1, T2, T3 random*	387.87	417.87
Cubic polynomial: Int, T1, T2 random	387.70	411.70
Cubic polynomial: Int, T1 random	407.92	425.92
Cubic polynomial: Int random	442.20	456.20
Quadratic polynomial: Int, T1, T2 random	391.71	413.71
Quadratic polynomial: Int, T1 random	411.52	427.52
Quadratic polynomial: Int random	445.38	457.38
Logarithmic Model: Int, TLN random	387.68	401.68
Logarithmic Model: Int random	488.72	498.72
Linear: Int, T1 random	425.94	439.94
Linear: Int random	458.39	468.39
Intercept-only model (no change)	521.46	529.46

−2LLR, log likelihood ratio statistic; AIC, Akaike's information criterion: smaller values are better. Int, Intercept coded so it is an estimate of BSI GSI at first assessment period. T1, Time coded for linear function. T2, Time squared coded for quadratic function. T3, Time cubed coded for cubic function. TLN, Time coded for logarithmic function. Time is number of months since initial assessment.
* Hessian matrix was positive definite although all convergence criteria were met.

scores with random intercept and change parameters as well as the statistics to compare parameter estimates of each AS group with each other group. Statistically significant differences in intercepts indicate that the two groups are not the same at the initial assessment. A significant difference between the change parameters indicates that the two groups changed in different ways. The change was tested with two sets of parameters based on AS group so that each AS group could be compared with each of the other groups. The overall fit and variance estimates were the same for either coding; only the parameter estimates changed with the different codes. The intercept and change parameters continued to demonstrate statistically significant variability, indicating that group membership did not account for all the intraindividual variability. Group membership accounted for approximately 30 percent of the variability in intercept scores, as demonstrated by a reduction in the intercept variance estimate (0.09 vs. 0.13); however, group membership did not account for any of the variance in the change parameter. There is not a statistically significant relationship between the intercept and change parameters, indicating that the subjects' initial scores were not related to how they changed over time. The change parameter is statistically significant, indicating that there was significant change over time, with scores declining (i.e., showing improvement on BSI scores). Neither AS group showed any statistically significant differ-

ences in change parameters in comparison with the other groups, demonstrating a lack of support for the hypothesis of differential change over time. Figure 4 demonstrates the underlying function of change over time for each group.

Each AS group had a statistically significant intercept, indicating that the initial group mean was different from the total sample mean and demonstrated differences from individual comparison groups. The AS MI group demonstrated significantly higher scores (i.e., more psychological distress) than did the total sample and scored significantly higher than the two non-mentally ill groups but were not different from the two other mentally ill comparison groups at the initial assessment. In contrast, the AS NMI group had significantly lower scores than did the total sample, scoring significantly lower than the three mentally ill groups and significantly higher than the GP NMI group.

Because the AS groups had six assessment periods, a second set of analyses was completed to make a direct comparison of change over time for the two AS groups. All models given in Table 2 were assessed, but the logarithmic model with random coefficients showed the best fit. The estimates for that model are provided in Table 4. Results are similar to the above results with the AS NMI group demonstrating lower overall initial values than the entire sample and statistically significant improvements over time. The AS MI group had significantly higher initial scores than the AS NMI group. There was not differential change over time, as indicated by the nonsignificant difference in the change parameters. Figure 5 provides a graph of each group's estimated change functions over time.

Discussion

The results of this study were inconsistent with the hypothesis that inmates, with or without mental illness, experience significant psychological decline in AS. Intercept comparisons showed that baseline differences were largely related to mental health status. Segregated inmates with mental illness displayed more symptoms than did inmates without mental illness. Mentally ill inmates in segregation were fairly similar to their comparison groups, but, from the beginning of the study, non-mentally ill segregated inmates had more symptoms than their GP comparison group had. It should be noted, however, that all offenders, regardless of their mental health status,

Table 3 Logarithmic Level 1 Models Testing Group and Group by Time Interactions for All Groups Using 5 Time Periods

Model Statistic	Estimates
Overall model fit	
Fit statistic (−2RLL)	290.92
AIC statistic	320.92
Repeated measures error structure	
Autoregressive variance estimate	0.04 (SE = 0.002, $p < .001$)
Autoregressive covariance estimate	0.64 (SE = 0.24, $p = .007$)
Random coefficients variance matrix	
Intercept random variance	0.09 (SE = 0.01, $p < .001$)
Intercept-change random covariance	−0.002 (SE = 0.003, $p = .51$)
Change parameter random variance	0.01 (SE = 0.002, $p < .001$)
Random Coefficients (Level 1 Parameters) With AS MI Coding	
AS MI intercept estimate	0.12 (SE = 0.04, $p = .007$)
AS MI change parameter estimate	−0.07 (SE = 0.02, $p < .001$)
Comparison of each group from AS MI Intercept	
AS NMI	−0.29 (SE = .08, $p < .001$)
SN MI	0.10 (SE = .06, $p = .09$)
GP NMI	−0.48 (SE = .07, $p < .001$)
GP MI	−0.08 (SE = .08, $p = .26$)
Comparisons of AS MI with each group on difference in change parameter	
AS NMI	0.01 (SE = 0.02, $p = .60$)
SN MI	0.02 (SE = 0.02, $p = .34$)
GP NMI	0.002 (SE = 0.03, $p = .94$)
GP MI	0.02 (SE = 0.03, $p = .45$)
Random Coefficients (Level 1 Parameters) With AS NMI Coding	
AS NMI intercept estimate	−0.16 (SE = 0.04, $p < .001$)
AS NMI change parameter estimate	−0.06 (SE = 0.02, $p = .001$)
Comparison of each group from AS NMI intercept	
AS MI	0.29 (SE = 0.06, $p < .001$)
SN MI	0.39 (SE = 0.06, $p < .001$)
GP NMI	−0.20 (SE = 0.07, $p = .005$)
GP MI	0.20 (SE = 0.08, $p = .008$)
Comparisons of AS NMI with each group on difference in change parameter	
AS NMI	−0.01 (SE = 0.02, $p = .60$)
SN MI	0.01 (SE = 0.02, $p = .68$)
GP NMI	−0.01 (SE = 0.03, $p = .69$)
GP MI	0.01 (SE = 0.03, $p = .75$)

Estimating logarithmic model with five time points using all groups. Models are coded in two ways so that each AS group is compared with all other groups on intercept and change parameters. Model fit statistics and covariance structures are the same for each model. Only the random and fixed parameter estimates change for the two models.

reported symptoms that were significantly elevated over normative community samples. Although the initial values showed group differences, the change function indicated significant change in psychological symptoms over time with early fast improvements slowing to stability. In contrast to the hypotheses, this pattern of change was similar in all five study groups.

The longitudinal design allowed assessment of whether change was occurring and in which direction. The presence of comparison groups avoids an attribution error; because findings were typically similar for people in segregation and in the general population, the findings cannot be attributed to segregation. These conclusions replicate those drawn by Zinger and colleagues,[10] although that study was criticized for high refusal rates, high attrition rates, small sample sizes, and short durations. Furthermore, the use of a reliable and valid standardized measure in the present study enabled objective assessment of psychological functioning.

A review of the findings warrants a discussion of plausible alternative explanations for inmates' responses to the questionnaire that might account for the results. Improvements may be due to reactivity; participants knew they were in a study and responded in a particular way. Perhaps they had a need to respond in a way that put them in the most favorable light (e.g., the ability to handle demands of confinement); however, comparisons to normative data
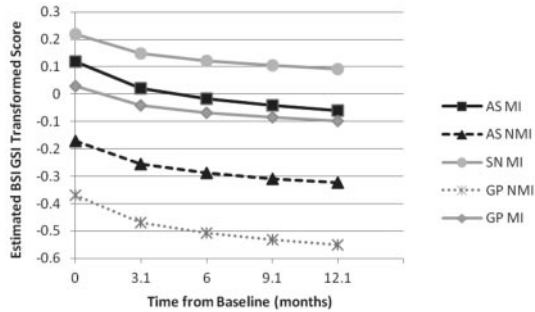
Figure 4. Logarithmic change functions for BSI GSI scores for each study group for five assessment periods. Although these lines are not statistically parallel, the interaction term was not statistically significant, indicating that the groups changed at the same basic rate over time. GP, general prison; AS, administrative segregation; MI, mental illness needs; NMI, no identified mental illness needs; SN, special needs prison. Time is the average number of months between assessments.
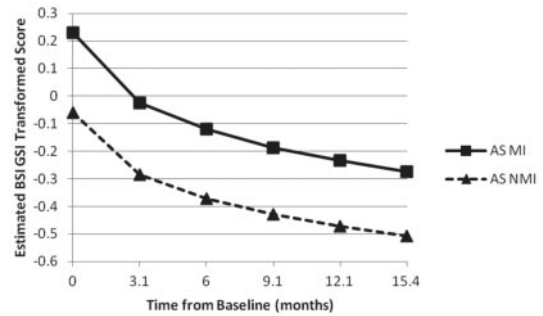

Figure 5. Logarithmic change functions for BSI GSI scores for AS groups for six assessment periods. Although these lines are not statistically parallel, the interaction term was not statistically significant, indicating that the groups changed at the same basic rate over time. AS, administrative segregation; MI, mental illness needs; NMI, no identified mental illness needs. Time is the average number of months between assessments.

indicate that the participants on average did not have good psychological functioning. 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 findings of improvement.[29] It is also possible that there are demand characteristics introduced by the field researcher that cues participants on how to respond; that seems unlikely as participants would be expected to respond in the hypothesized direction. Although a

Table 4 Logarithmic Level 1 Model Testing Group and Group by Time Interactions for AS Groups Using Only Six Time Periods

Model Statistic	Estimate
Fit statistic (−2RLL)	221.21
AIC statistic	239.21
Repeated-measures error structure	
Autoregressive variance estimate	0.04 (SE = .003, $p < .001$)
Autoregressive covariance estimate	0.77 (SE = .27, $p = .004$)
Random coefficients variance matrix	
Intercept random variance	0.11 (SE = .02, $p < .001$)
Intercept-change random covariance	−0.009 (SE = .006, $p = .12$)
Change parameter random variance	0.01 (SE = .003, $p < .001$)
Random coefficients (Level 1 parameters)	
AS NMI intercept estimate	−0.16 (SE = .05, $p = .001$)
AS NMI change parameter estimate	−0.06 (SE = .02, $p = .001$)
Group intercepts effect (overall)	
AS MI difference from AS NMI intercept	0.29 (SE = .07, $p < .001$)
Group by time interaction	
AS MI difference from AS NMI change	−0.02 (SE = .02, $p = .40$)

All models (as in Table 2) were re-estimated for the AS groups with six time periods; the logarithmic model was still the best fitting model.

testing or practice effect might explain improvements on cognitive measures, we were unable to find evidence that psychological measures should be influenced by testing effects. Study demands may lead to positive ratings, but it seems unlikely that response biases would overshadow the negative impacts of AS if they existed. However, there is not enough information in the data collected to account for the positive change. The most likely explanation is that all subjects were included in the study when in the midst of a crisis and, with time, the crisis dissipated and they adapted to their environment, a finding that is consistent with the research of Zamble and Porporino[30,31] on adaptation to prison and that is delineated by the logarithmic change function.

Although this study incorporated several design features that improved on the capacity of previous research to draw conclusions about the effects of AS, there are several limitations that affect its generalizability to other settings. First, it included literate adult male offenders and should therefore not be generalized to female offenders, illiterate offenders, or juveniles. Second, because we studied behaviorally disruptive inmates, they may have experienced punitive or administrative segregation previously, and thus we are not assessing persons with no prior experience in isolation conditions. Third, segregation conditions vary from state to state on a host of variables, including average duration of AS, double-bunking, televisions, exercise, selection criteria for AS, and quality and quantity of mental health and medical services. Thus, the results of the study can be generalized only to other prison systems to the extent

that their conditions of AS confinement are similar to Colorado's.

The duration of the study was limited to one year because it was postulated in earlier research that the effects of segregation would be quickly evident.[5,32–34] Kupers 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 " (Ref. 2, p 1006). Therefore, we expected that deleterious effects would become evident within a year, but it is possible that they do not appear until after longer periods of segregation.

This study was not designed to address the question of whether segregation is an appropriate confinement option for offenders, including those with serious and persistent mental illness. We are unaware of any treatment guideline that suggests that long-term confinement in an AS environment would be clinically helpful. We examined both intraindividual differences in change and intergroup differences. Although the data suggest that there is variability in change over time, it is not the study conditions that explain these differences. We used smooth mathematic functions to study change over time; 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 the data because testing occurred at three-month intervals.

Replication is needed in other prisons to determine whether these findings hold true when conditions of confinement vary. Further research is needed to understand how increased services, privileges, staff, and out-of-cell time may ameliorate the unintended consequences of AS, and research should inform prison officials about the standards and practices necessary to protect inmates in segregation from potentially harmful psychological effects. It is also important to note that there may be other negative consequences of AS that we did not study, and research has yet to demonstrate the efficacy of AS in improving inmate behavior and conditions for the rest of the system. Thus, we make no empirical or value judgments about whether and to what degree the use of AS balances the purported benefits (e.g., a safer prison system) with costs (e.g., significant reductions in freedom).

We do not claim, nor believe, that these data definitively answer the question of whether long-term segregation causes psychological harm. We used one rigorous methodology to study, for one year, inmates in one state prison system that may or may not be similar to other prisons. Frankly, having seen individuals in psychological crisis in segregation, we were surprised that such effects did not appear in these data. We believe that this study moves us forward, but that future research will shed additional light on this crucial question.

References

1. Haney C: A culture of harm: taming the dynamics of cruelty in supermax prisons. Crim Justice Behav 35:956–84, 2008
2. Kupers T: What to do with the survivors?—coping with the long-term effects of isolated confinement. Crim Justice Behav 35:1005–16, 2008
3. Mears DP, Watson J: Towards a fair and balanced assessment of supermax prisons. Justice Q 23:232–70, 2006
4. Grassian S: Psychopathological effects of solitary confinement. Am J Psychiatry 140:1450–4, 1983
5. Haney C: Mental health issues in long-term solitary and "supermax" confinement. Crime Delinq 49:124–56, 2003
6. Madrid v. Gomez, 889 F. Supp. 1146 (N.D. Cal. 1995)
7. Arrigo BA, Bullock JL: The psychological effects of solitary confinement on prisoners in supermax units: reviewing what we know and recommending what should change. Int J Offender Ther Comp Criminol 52:622–41, 2008
8. Metzner JL, Dvoskin JA: An overview of correctional psychiatry. Psychiatr Clin North Am 29:761–72, 2006
9. Pizarro JM, Narag RE: Supermax prisons: what we know, what we do not know, and where we are going. Prison J 88:23–42, 2008
10. Zinger I, Wichman C, Andrews DA: The psychological effects of 60 days in administrative segregation. Can J Criminol 43:47–88, 2001
11. Brodsky S, Scogin F: Inmates in protective custody: first data on emotional effects. Forensic Rep 1:267–80, 1988
12. Haney C: Infamous punishment: the psychological effects of isolation. Natl Prison Project J 8:3–21, 1993
13. Hodgins S, Côté G: The mental health of penitentiary inmates in isolation. Can J Criminol 33:175–82, 1991
14. Suedfeld P, Ramirez C, Deaton J, et al: Reactions and attributes of prisoners in solitary confinement. Crim Justice Behav 9:303–40, 1982
15. Ecclestone CEJ, Gendreau P, Knox C: Solitary confinement of prisoners: an assessment of its effects on prisoners' personal constructs and adrenocortical activity. Can J Behav Sci 6:178–91, 1974
16. Gendreau P, Bonta J: Solitary confinement is not cruel and unusual punishment: people sometimes are! Can J Criminol 26:467–78, 1984
17. Gendreau PE, Freedman N, Wilde GJS, et al: Stimulation seeking after seven days of perceptual deprivation. Percept Mot Skills 26:547–50, 1968
18. Gendreau P, McLean R, Parsons T, et al: Effect of two days' monotonous confinement on conditioned eyelid frequency and topography. Percept Mot Skills 31:291–3, 1970
19. Gendreau PE, Freedman, N, Wilde GJS, et al: Changes in EEG alpha frequency and evoked response latency during solitary confinement. Abnorm Psychol 79:54–9, 1972
20. Andersen HS, Sestoft D, Lillebaek T, et al: A longitudinal study of prisoners on remand: psychiatric prevalence, incidence, and psy-

chopathology in solitary vs. non-solitary confinement. Acta Psychiatr Scand 102:19–25, 2000

21. O'Keefe ML, Klebe KJ, Stucker A, *et al*: One year longitudinal study of the psychological effects of administrative segregation. Technical Report. CO Department of Corrections 2010. Available at http://www.doc.state.co.us/sites/default/files/opa/AdSeg Report_2010.pdf. Accessed January 5, 2011
22. Derogatis L: Brief Symptom Inventory: Administration, Scoring, and Procedures Manual. Minneapolis, MN: NCS Pearson Inc, 1993
23. Boulet J, Boss MW: Reliability and validity of the Brief Symptom Inventory. Psychol Assess 3:433–7, 1991
24. Kellett S, Beail N, Newman DW, *et al*: Utility of the Brief Symptom Inventory in the assessment of psychological distress. J Appl Res Intellect 16:127–34, 2003
25. Cundick BP: Review of the Brief Symptom Inventory. In: Conoley JC, Kramer JJ, eds. The Tenth Mental Measurement Yearbook. Lincoln, NE: University of Nebraska Press, 1975, p 111
26. Morlan KK, Tan S: Comparison of the Brief Psychiatric Rating Scale and the Brief Symptom Inventory. J Clin Psychol 54:885–94, 1998
27. West BT: Analyzing longitudinal data with the linear mixed models procedure in SPSS. Eval Health Prof 32:207–28, 2009
28. Heck RH, Thomas SL, Tabata LN: Multilevel and Longitudinal Modeling With IBM SPSS. New York: Routledge, 2010
29. Gottfredson GD: The Hawthorne misunderstanding (and how to get the Hawthorne effect in action research). J Res Crime Delinq 33:28–48, 1996
30. Zamble E: Behavior and adaptation in long-term prison inmates: descriptive longitudinal results. Crim Just Behav 19:409–25, 1992
31. Zamble E, Porporino FJ: Coping, imprisonment, and rehabilitation: some data and their implications. Crim Just Behav 17:53–70, 1990
32. Grassian S, Friedman N: Effects of sensory deprivation in psychiatric seclusion and solitary confinement. Int J Law Psychiatry 8:49–65, 1986
33. Lovell D, Johnson LC, Cain KC: Recidivism of supermax prisoners in Washington state. Crime Delinq 53:633–56, 2007
34. Toch H: Living in Prison: The Ecology of Survival. Washington, DC: American Psychological Association, 1992