Field validity of the STATIC-99 and MnSOST-R among sex offenders evaluated for civil commitment as Sexually Violent Predators

Article *in* Psychology Public Policy and Law · November 2009

DOI: 10.1037/a0017232

Daniel C Murrie
University of Virginia

Samuel W Hawes
Florida International University

Psychology, Public Policy, and Law
2009, Vol. 15, No. 4, 278–314

1076-8971/09/$12.00 DOI: 10.1037/a0017232

FIELD VALIDITY OF THE STATIC-99 AND MnSOST-R AMONG SEX OFFENDERS EVALUATED FOR CIVIL COMMITMENT AS SEXUALLY VIOLENT PREDATORS

Marcus T. Boccaccini
Sam Houston State University

Daniel C. Murrie
University of Virginia

Jennifer D. Caperton and Samuel W. Hawes
Sam Houston State University

Despite their widespread use in forensic and correctional practice, surprisingly little research investigates how well actuarial risk assessment instruments (ARAIs) for sexual offenders work within the contexts where they are routinely applied. We examined the predictive validity ($M = 4.77$ years follow-up) of the two most widely used ARAIs for sexual offenders, the STATIC-99 and Minnesota Sex Offender Sex Offender Screening Tool–Revised (MnSOST-R), as administered in routine practice among 1,928 offenders screened for possible civil commitment as sexually violent predators. Effect sizes for both ARAIs were lower than in most published research and meta-analytic reviews, although the STATIC-99 was a more consistent predictor of recidivism than the MnSOST-R. Recidivism rates for the STATIC-99 were much closer to those expected based on the 2009 norms than the 2003 norms. Offender characteristics (e.g., age at release, prior arrests, release type) were often as or more effective than ARAIs for predicting recidivism. This study, apparently the largest cross-validation study of popular ARAIs for sex offenders, suggests that the predictive validity of these measures in routine practice in the United States may be poorer than often assumed.

Keywords: risk assessment, sexually violent predator, sex offender civil commitment, STATIC-99, MnSOST-R

While sexual offenders comprise a sizable portion of U.S. prison inmates, they evoke an even greater portion of public concern. To protect potential victims from those sexual offenders perceived to be at highest risk for reoffense, states have enacted a number of policies unique to sexual offenders, including registries for convicted sexual offenders, community notification laws, and post-incarceration civil commitment for those determined to be sexually violent predators (SVPs). Of course, any policy that attempts to identify the offenders at highest risk for recidivism requires some form of risk assessment. Therefore, actuarial

---

Marcus T. Boccaccini, Jennifer D. Caperton, and Samuel W. Hawes, Department of Psychology, Sam Houston State University, Huntsville, Texas; Daniel C. Murrie, Institute of Law, Psychiatry, and Public Policy, University of Virginia, Charlottesville, Virginia.

Jennifer D. Caperton is now a psychologist at the Federal Correctional Institution, Seagoville, Texas.

This research is based, in part, on Jennifer D. Caperton's doctoral dissertation, completed at Sam Houston State University.

Correspondence concerning this article should be addressed to Marcus T. Boccaccini, Department of Psychology, Box 2447, Sam Houston State University, Huntsville, TX 77341. E-mail: boccaccini@shsu.edu

risk assessment instruments (or ARAIs; Janus & Prentky, 2003) have become popular in many forensic and correctional settings.

Broadly speaking, actuarial risk assessment involves explicit research-supported rules that specify which risk factors are examined, how those risk factors are scored, and how those scores are mathematically combined to yield an objective estimate of risk (Monahan, 2006). Researchers developed popular ARAIs, such as the STATIC-99 (Hanson & Thornton, 1999) and the Minnesota Sex Offender Sex Offender Screening Tool-Revised (MnSOST-R; Epperson, Kaul, & Hesselton, 1998), by studying samples of released offenders over a specific follow-up period, and documenting recidivism. Researchers identified risk factors—usually data easily retrieved from records such as age and prior offenses—that were statistically related with recidivism. They also documented recidivism rates among subgroups of the offenders who had specific numbers of risk factors (e.g., of offenders with X of the identified risk factors, Y% reoffended over Z years). Thus, the premise of ARAIs is that clinicians can observe the number of pre-defined risk factors that an offender demonstrates, and estimate the likelihood that an offender with a certain number of the pre-defined risk factors will recidivate, based on the observed recidivism rate in the risk measure's development sample.

In most respects, the movement to develop and adopt ARAIs has been positive. Relying on ARAIs as a core component of forensic risk assessments tends to make the risk assessment process more standardized and transparent across clinicians (Janus & Prentky, 2003). Generally, actuarial approaches tend to yield more accurate risk estimates than unstructured clinical judgments (Grove & Meehl, 1996; Grove, Zald, Lebow, Snitz, & Nelson, 2000). Regarding sex offender risk assessment in particular, the most recent and comprehensive meta-analysis of the subject (Hanson & Morton-Bourgon, 2009) revealed that ARAIs designed to predict sexual reoffense (d = .67) clearly outperformed unstructured professional judgment (d = .42). In other words, most research suggests that ARAIs result in evaluations that are more uniform and more accurate.

Perhaps for these reasons, most of the professional community has embraced ARAIs. National surveys reveal that the vast majority of states use ARAIs at some point in sex offender supervision (Interstate Commission for Adult Offender Supervision, 2007). At least thirty states reported using the STATIC-99 specifically. Another popular sex offender risk measure, the MnSOST-R, has been adopted by seven state systems (ICAOS, 2007) and more than 20% of sex offender treatment programs in the United States (McGrath, Cumming, & Bouchard, 2003). Indeed, some scholars conclude that "performing risk assessments without (ARAIs for sexual offenders) is unethical for mental health professionals and improper for courts" (Janus & Prentky, 2003 p. 4).

ARAIs and SVP Proceedings

ARAIs are a common element in many sex offender assessment and management procedures. But they are particularly common in SVP proceedings. SVP laws allow states to identify sexual offenders perceived to pose a high risk for repeated sexual offenses, and civilly commit them after their incarceration, in order to protect potential victims and provide treatment to the offender (for detailed descriptions of forensic evaluations in SVP proceedings, see Campbell, 2007; Doren, 2002; Jackson & Richards, 2008; Miller, Amenta, & Conroy, 2005; Witt & Conroy, 2009). Because most SVP commitment laws follow criteria that the Supreme Court set forth in *Kansas v. Hendricks* (1997), they require four elements to civilly commit an offender as an SVP: (a) a history of sexual offending, (b) a mental abnormality (sometimes defined as a mental disorder or personality disorder), (c) a volitional impairment rendering him less able to control his sexual behavior, and (d) significant risk for future sexual offending (Miller et al., 2005).

ARAIs are certainly not the *only* evidence in SVP trials; criminal history and clinical diagnosis are typically the evidence most relevant to the first few SVP criteria. But ARAIs are the form of evidence typically introduced to address this fourth SVP criterion, i.e., an offender's risk of future offending. Indeed, one state statute (Virginia Code Ann. § 37.2–903) delineating procedures related to SVP proceedings specifically *requires* the state Department of Corrections to administer to all sexual offenders a particular ARAI (currently the STATIC-99) and refer for a subsequent clinical evaluation any inmates who score above a certain threshold. In typical clinical evaluations for SVP proceedings, nearly all clinicians choose to administer ARAIs (Jackson & Hess, 2007), and the ARAI scores that clinicians' assign are strongly associated with their opinions about which offenders meet criteria for commitment (Levenson, 2004). Likewise, guidelines from professional groups (Association for the Treatment of Sexual Abusers, 2001) and professional texts (Doren, 2002; Jackson & Richards, 2008) direct evaluators to use ARAIs in SVP evaluations. Finally, ARAIs are routinely admitted as evidence during SVP trials (Janus & Prentky, 2003).

Actuarial Risk Assessment Instruments for Sexual Offenders: Development Research

One clear strength of ARAIs is their empirical basis. ARAIs such as the STATIC-99 and the MnSOST-R were developed by following samples of released sexual offenders and documenting risk factors and recidivism rates. One potential limitation of ARAIs relates to their generalizability (Monahan, 2003). Any empirically based instrument-development process takes place with a specific sample, within a particular context, at a particular time in history. As with any measure, it is important to understand the details and context of the instrument-development process to gauge how appropriate an instrument might be for use in other contexts. Although a measure developed in one context may work as well in other contexts, this is not always the case.

The STATIC-99 developers (Hanson & Thornton, 1999, 2000) tested its performance within three developmental samples, as well as a fourth sample (for cross validation). The first three samples were Canadian (two secure psychiatric facilities and a provincial prison), whereas the fourth comprised all sexual offenders ($n > 500$) released in 1979 from Her Majesty's Prison, England and Wales. The STATIC-99 showed reasonably good predictive validity (area under the receiver operating characteristic curve [AUC] = .71) with respect to sexual reoffense, and the instrument authors also provided a table of observed recidivism rates for both sexual and violent reoffense, at 5, 10, and 15-year intervals. The recidivism rates were further subdivided by scores on the STATIC-99, which range from 0 to 12 (although 6–12 are grouped together because of small sample sizes and similar rates of recidivism corresponding to these higher scores). The table is a resource that has become widely used in practice (see Phenix, Hanson, & Thornton, 2000; Harris et al., 2003) because it allows an evaluator to identify observed recidivism rates for offenders in the STATIC-99 development sample who scored similarly to the offender being evaluated.

The STATIC-99 developers recently made available a set of 2009 norms (Helmus, Hanson, & Thornton, 2009), which comprise data from larger and more geographically diverse samples than the originally published norms. The 2009 norms present recidivism data somewhat differently (e.g., based on logistic regression), and in greater detail (e.g., subdividing scores in the 6–12 range), but like the 2003 norms, provide evaluators with tables to estimate recidivism rates for offenders with a particular score.

The process of developing the MnSOST-R was similar to the STATIC-99, though smaller in scale (see Epperson et al., 2003). The MnSOST-R development sample included 256 felony sexual offenders: 107 released in 1988, 108 released in 1990, and "41 offenders readmitted to the Minnesota Department of Corrections during the time the sample was being put together regardless of release year" (p. 10). Epperson and colleagues then assembled a cross-validation sample "of comparable sex offenders released in 1992 for whom complete data were available ($N = 220$)" (p. 11). Next, they examined sexual reoffense rates for a 6-year follow-up period. Finally, they developed a scoring procedure in which risk factors are not only scored as present or absent, but a point is added for each 5% increase in recidivism (over the base rate) associated with the presence of that factor.

Epperson and colleagues (2003) reported that this development process resulted in an instrument that predicted sexual reoffense in their cross-validation sample (AUC = .73). They also reported "presumptive risk levels"—the descriptors low, moderate, and high—for particular score ranges. Finally, much like the developers of the STATIC-99, they provided a table of observed recidivism rates corresponding to each risk level. As with the STATIC-99, the table allows an evaluator using the MnSOST-R to identify observed recidivism rates among offenders in the validation sample who scored similarly to the offender he or she evaluated.

Cross-Validation of ARAIs

Despite an empirically sound development process, ARAIs, like any test or prediction measure, require cross-validation (Gottfresson & Moriarty, 2006; Wollert, 2002). Cross-validation refers to the process of replicating results in new samples, an essential component of prediction model development (American Educational Research Association, American Psychological Association, & National Council of Measurement in Education, 1999; Cohen, Cohen, West, & Aiken, 2003; Pedhazur, 1997). Standard 3.10 of the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999) states explicitly, "Test developers should conduct cross-validation studies when items are selected primarily on the basis of empirical relationships . . ." (p. 44). Cross-validation is especially important when measures are going to be used within a group that was not represented in the development sample. Before using an empirically derived measure, clinicians should ensure that supportive data is available from a sample "sufficiently large and representative of the population for which the test is intended" (Standard 3.13, American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999, p. 46).

Cross-validation is essential for ARAIs like the STATIC-99 and MnSOST-R because they are routinely administered in contexts quite different from the contexts in which they were developed. For example, different regions (and even different jurisdictions in the same region) are likely to differ in terms of law enforcement investigation and prosecution practices, sex offender supervision policies, and sex offender treatment resources. Each of these differences might influence the true rates of sex offender recidivism, the recidivism rates detected by law enforcement, and perhaps the extent to which some well-recognized risk factors correspond to recidivism.[1] Jurisdictions might also differ in the extent to which criminal justice records are detailed and easily available and the extent to which evaluators are trained or skilled in the use of ARAIs. These differences, too, might influence the accuracy of ARAIs observed across jurisdictions. Finally, there may be differences in the base rate of recidivism between the development sample(s) and the jurisdiction in which the ARAI is being used. If the base rate of recidivism is lower in the new jurisdiction, recidivism rates for specific scores will be lower than they were in the development sample (Mossman, 2006).

Fortunately, many researchers have recognized the importance of cross-validation research for ARAIs. They have examined popular ARAIs in contexts beyond those in which they were developed and reported results that tend to suggest ARAIs are moderate predictors across many contexts (see Hanson & Morton-Bourgon, 2009). However, a careful review of this replication research suggests that we still know little about one context in which ARAIs are applied often, and with considerable influence: SVP proceedings in the United States.

Cross-Validation of the STATIC-99

The most comprehensive and current review of STATIC-99 cross-validation research (Hanson & Morton-Bourgon, 2009) reported moderate to strong predictive validity (d = .67, 95% confidence interval [CI] = .62 to .72) for STATIC-99 scores with respect to sexual recidivism, across more than 60 studies and 24,000 offenders. Unstructured professional judgment (d = .42) and actuarial measures designed to predict general violence (d = .39) were considerably less accurate. These findings suggest the STATIC-99 appears to be a well cross-validated

[1] To take a simple example, age bears a well-documented relation to criminal and sexual reoffense. But, the observed relation between age and reoffense might be weaker in jurisdictions that tend assign more supervision, or more stringent release conditions, for younger offenders. Similarly, treatment participation may bear no relationship to recidivism in a context where quality treatment is unavailable; yet a treatment participation variable might prove related to recidivism in contexts where rigorous treatment is available.

measure. However, the STATIC-99 (like other predictors examined in the meta-analysis) yielded stronger effect sizes in the United Kingdom (*d* = .90) than the United States (*d* = .60). Although some unpublished reports and raw data are available, we could identify only two published, peer-reviewed studies that examined the STATIC-99 in U.S. samples. In one (Bartosh, Garby, Lewis, & Gray, 2003), STATIC-99 scores were only modestly (AUC = .64; *d* = .50) predictive of sexual reoffense among a cohort of 186 Arizona sexual offenders followed over 5 years.[2] In the other published study, Sreenivasan and colleagues (2007) reported a modest AUC of .62 (*d* = .43) with respect to sexual recidivism among a select sample (*N* = 137) of high-risk California sexual offenders, among whom the base rate of sexual reoffense was unusually high (31% over 5 years). Effect-sizes from both of these studies fell below the lower bound of the 95% CI for the STATIC-99 in the Hanson and Morton-Bourgon (2009) meta-analysis.[3] These findings suggest that the STATIC-99 can attain moderate levels of predictive accuracy in U.S. samples, but these predictive effects in the United States tend to be smaller than the effects typically reported in the research literature.

Cross-Validation of the MnSOST-R

Compared with the STATIC-99, the MnSOST-R has been the subject of fewer cross-validation studies and more critiques (e.g., Vrieze & Grove, 2008; Wollert, 2002; 2003). Nevertheless, Hanson and Morton-Bourgon (2009) identified twelve studies that included the MnSOST-R and calculated a mean *d* of .76 (95% CI = .65 to .87) across 12 studies (total *N* = 4,672). In the only peer-reviewed, published study investigating a U.S. sample (Bartosh et al., 2003), predictive accuracy (*d* = .32) was well below the lower bound of the CI for the meta-analytic effect.

Cross-Validation and "Local Data"

Cross-validation in new samples is especially important for instruments such as the STATIC-99 and MnSOST-R because users sometimes convey scores in terms that are directly linked to findings from the original test development sample. That is, evaluators often report the expected rates of reoffense associated with a particular score *in the original instrument development sample* (see Mossman, 2006). Indeed, the 2003 scoring manual for the STATIC-99 provided templates to help evaluators convey results in this manner (Harris et al., 2003). But, as Mossman (2008) explained,

> An ever-present (but under-appreciated) potential problem with using [ARAIs for sex offenders] stems from the practice of believing that published numerical results (e.g., rates of recidivism) found for specific populations in specific social

[2] To facilitate comparisons between findings from existing studies, the Hanson and Morton-Bourgon (2009) meta-analysis, and the current study, we used the conversion table provided by Rice and Harris (2005) to estimate Cohen's *d* values for existing studies that reported only AUC values.

[3] Effect sizes for the STATIC-99 in U.S. samples have also been relatively small in non-peer-reviewed reports. For example, in a combined sample of offenders from Alaska and Iowa (*N* = 205), the AUC for sexual recidivism was .56 (*d* = .21) and the AUC for violent recidivism was .61 (*d* = .40; Hanson, Harris, Scott, & Helmus, 2007).

> contexts (say, sex offenders released from Canadian prisons in the 1980s) will be correct for different populations in different social contexts (say, sex offenders released in South Carolina in the year 2010). (p. 287)

Indeed, scholars who study risk assessment in criminal justice contexts (e.g., Gottfresson & Moriarty, 2006) and among sexual offenders, in particular (e.g., Campbell, 2007; Mossman, 2006; Wollert, 2006; Woodworth & Kadane, 2004), warn that we need to be very cautious about taking ARAIs from one context and applying them to another context. The STATIC-99 authors now recommend that evaluators should avoid associating an offender's score with a *single* recidivism estimate from the normative sample, because of differences in the predictive properties of specific STATIC-99 scores across samples (Helmus et al., 2009). Rather, they encourage evaluators to report a *range* of recidivism rate estimates (high and low) based on findings from multiple samples. They also encourage evaluators to use professional judgment to determine where an offender is likely to fall within the range of rates (Helmus et al., 2009).

Ideally, evaluators who take ARAIs developed in one context, and use them with different populations in different contexts, could consider an ARAI score with respect to "local data" rather than data from offenders released in another country in another era. To continue with Mossman's example, a psychologist conducting a risk assessment with a sex offender released in South Carolina in 2010 will probably perform a better (and more defensible) risk assessment if she considers local data from the population of South Carolina offenders in recent years, rather than from Canadian or United Kingdom offenders decades earlier. Local data would also be the most appropriate source of data for making the types of recidivism rate judgments that the STATIC-99 authors (Helmus et al., 2009) now recommend.

Of course, appropriate local data is not always easy to amass or access, and small-scale cross-validation studies are typically not sufficient. Meehl (1954) explained that we need base rate data from populations that are defined narrowly enough to be useful, but from samples that are large enough that the obtained data are stable and meaningful. The large combined normative sample that has recently become available for the STATIC-99 may provide data that is stable and meaningful (see www.static99.org). Nevertheless, the recidivism rates most relevant to a particular offender are likely the rates observed in the offender's immediate context (see Conroy & Murrie, 2007 for discussion of "local base rates" in risk assessment). Ideally, large jurisdictions, such as state correctional systems, could interpret ARAI results from offenders in their system by considering base rate data from their own system. For example, when evaluating offenders released from a Texas prison, it would be preferable to consider recidivism in a large sample of Texas offenders, rather than to draw inferences from the large STATIC-99 dataset that includes relatively few U.S. samples, and no data specific to Texas prison inmates.

Current Study

Because ARAIs have more empirical support than any other methodology for assessing sex offender recidivism risk (Hanson & Morton-Bourgon, 2009), it is reasonable that state correctional systems and independent evaluators use ARAIs to evaluate offenders for SVP civil commitment in the United States. Nevertheless, we know little about the effectiveness of ARAIs for predicting recidivism among SVP candidates in the United States. Indeed, after reviewing all available STATIC-99 research, Hanson and Morton-Bourgon (2009) suggested, "the observation of jurisdictional differences in predictive accuracy suggests that all countries should conduct local validity studies prior to routine implementation of measures developed in other jurisdictions" (p. 16).

In this study, we examine the use and predictive validity of the STATIC-99 and MnSOST-R in a statewide process that screened offenders for possible commitment as SVPs. First, we consider whether scores on these measures are associated with offenders' progress through the SVP evaluation process, by examining whether higher scores are associated with referral for additional risk evaluations and commitment as an SVP. Although existing research suggests that ARAI scores are associated with clinicians' *opinions* about who should be committed (Levenson, 2004; Levenson & Morin, 2008), our data allows for an examination of the relation between ARAI scores and actual commitment decisions.

Second, we examine the predictive validity of the STATIC-99 and MnSOST-R with respect to recidivism, among non-committed offenders who had at least 2.25 years of opportunity to reoffend ($M = 4.77$ year follow up). To help gauge the utility of these ARAIs, we also compared their predictive validity to that of easily obtained offender characteristics (e.g., age at release, prior offenses). To our knowledge, this is the first study examining predictive validity of ARAIs among offenders evaluated for commitment as SVPs in the United States. For the STATIC-99, this is one of few studies examining predictive validity in the United States and represents, by far, the largest cross-validation study in the United States. The large sample allows us to examine recidivism rates by scores and consider the extent to which risk probability values reported by the instrument authors apply in this sample. Finally, we present this study to illustrate the importance of the "local validity studies" recommended by Hanson and Morton-Bourgon (2009, p. 16) and the ways in which data from individual jurisdictions might better inform decisions in those jurisdictions than can large-scale aggregated normative data.

Method

Study Sample and the Texas SVP Evaluation Process

In September of 1999, Texas implemented a systematic risk assessment process to assess sexual offenders who are nearing release and who meet specific statutory criteria (see Texas Health & Safety Code §841.021, which specifies offenders who are serving time for sexually violent offenses and who may be repeat sexual offenders). The risk assessment process was developed after careful reviews of research literature and national practice, to identify the best-supported, or most promising sex offender risk measures as of 1999. Therefore, the Texas Department of Criminal Justice (TDCJ) began administering the STATIC-99 and MnSOST-R, no later than 16 months before the end of each potentially qualifying offender's sentence. Both measures are scored by either parole officers or mental health staff (e.g., MA-level clinicians). Scores from these measures and other offender records are then reviewed by a Multidisciplinary Team (MDT), comprised of members of TDCJ, the state's sex offender council, law enforcement agencies, and the state's department of mental health. If the MDT determines that the inmate has two qualifying sexual offenses they may then refer the offender for a commissioned "assessment for behavioral abnormality," to help determine whether he meets statutory criteria for civil commitment as an SVP. Of the offenders considered to have a behavioral abnormality, the state pursues only a small subset for civil commitment (10–15 per year during the period of this study).

In this study, we examine the risk scores from the first 1,983 male sex offenders in Texas prisons who were evaluated by TDCJ staff as part of this SVP screening process (between September 1999 and September 2004). For this study, TDCJ allowed the research team to access an existing database of offender characteristics and risk scores that included total scores for the STATIC-99 and MnSOST-R. The database also provided information about each offender's date of release, date of birth, sex, and race.

We removed twelve offenders from the study because we were not able to obtain their arrest records in order to examine recidivism (process described below). For the remaining 1,971 offenders, age at the time of evaluation ranged from 18.18 to 88.19 years (M = 42.94, SD = 12.13). Almost half of the offenders were identified by the TDCJ database as White (n = 938, 47.6%), while others were identified in the TDCJ database as Hispanic (n = 602, 30.5%), Black (n = 420, 21.3%), or other (n = 11, 0.6%). About one in five offenders were referred for a behavioral abnormality evaluation (n = 395, 20.0%), and 43 (2.2%) of the offenders were eventually committed as SVPs by a judge or jury.

Although SVP commitment in Texas is technically an outpatient program, we did not include SVP offenders in the recidivism analyses for two reasons. First, offenders committed as SVPs are managed closely in the community, even as compared with other supervised sex offenders. For example, they are monitored via real-time satellite tracking, are mandated to participate in treatment, and can be required to submit to polygraphs, penile plethysmographs, and drug testing (see Council on Sex Offender Treatment, 2007, for additional information about the Texas outpatient program). Second, because Texas is the only state with an outpatient SVP commitment program, combining committed offenders and noncommitted offenders into the same sample would make it difficult to compare results from Texas to those from other states, which rely on inpatient commitment. Although we could have examined the relation between ARAIs and recidivism for civilly committed offenders without combining them with other offenders, at the time of this study, no offender who has been committed as an SVP has recidivated with a sexually violent offense (though some have returned to prison for other violations of their supervision terms).

For the 1,928 offenders who were not civilly committed, the length of follow-up ranged from 2.25 to 7.50 years (M = 4.77, SD = 1.52). TDCJ listed the type of release for these offenders as either "mandatory supervision required" (n = 959, 49.5%) or "discharge" (n = 977, 50.4%). There was no release type specified for three offenders. Offenders released on "discharge" status receive no further monitoring or supervision. Offenders released under mandatory supervision are subject to restrictions that are typical of most parole arrangements (see http://www.tdcj.state.tx.us/bpp/what_is_parole/mandsupv.htm for a description of the similarities between mandatory supervision and parole). Offenders can be released under mandatory supervision when their actual time served combined with earned "good time" credit equals their total sentence length. Thus, the length of time an offender spends under mandatory supervision varies from offender to offender, depending on the amount of good time credit he received and the original sentence length. The data available for this study did not allow us to determine when offenders' mandatory supervision terms expired. In other words, some offenders may have been on mandatory supervision throughout the follow-up period, while others may have completed their mandatory supervision requirements sometime during the follow-up period.

Measures

The TDCJ risk score database was the main source of data used for the study (STATIC-99 score, MnSOST-R score, offender background characteristics, release dates). The TDCJ database did not provide information about pre- or post-release offenses. We obtained this information from arrest records provided by the Texas Department of Public Safety (DPS). Figure 1 summarizes the

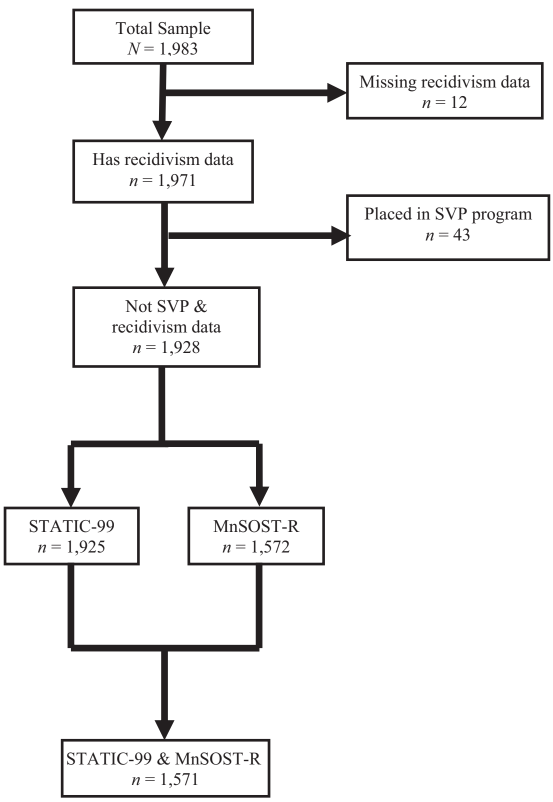

Figure 1. Flowchart summarizing the availability of actuarial risk assessment instrument (ARAI) scores and recidivism data for offenders.

measures used in the primary study analyses and the number of offenders who had each type of information.

MnSOST-R. The MnSOST-R is a 16-item ARAI designed to predict sexual recidivism among offenders who have committed sexual offenses other than incest (Epperson et al., 1998).[4] Twelve of the MnSOST-R items assess historical or static predictors of recidivism, such the number of sex offenses, offending in a public place, and use of force or threat during the offense (see Epperson et al., 1998). Four items assess institutional or dynamic predictors, such as receiving treatment while incarcerated and age at release. Item scores are based on weights that vary from item to item. For example, disciplinary history while incarcerated is scored as 1 (*present*) and 0 (*absent*), while length of sexual offending history is scored −1 (less than 1 year), 3 (1 to 6 years), or 0 (more than 6 years). Scores on the MnSOST-R can range from −14 to 31. Epperson et al. (2003) described scores of 3 and below as indicating a low risk level (12% recidivism likelihood within 6 years), scores from 4 to 7 as indicating a moderate risk level (25% recidivism likelihood), and scores of 8 or above as indicating a high risk level (57% recidivism likelihood).

The TDCJ database contained MnSOST-R scores for 1,610 (81.7%) of the 1,971 offenders in the sample. Excluding 38 offenders with MnSOST-R scores who were pursued for civil commitment, scores were available for 1,572 (81.5%) of the remaining 1,928 offenders. The mean MnSOST-R score among offenders was 5.52 (SD = 5.38).

Of course, relying on archival data from a correctional system's everyday practice does not allow us to present the kind of formal reliability check that researchers typically conduct when beginning a formal study of risk assessment instruments. Simply put, we do not know the interrater reliability among staff who scored the MnSOST-R, a limitation common to archival studies and routine "real world" practice. We can, however, present some related data that may allow for an indirect, and imperfect, estimate of reliability in this context. In a small sample of 29 committed offenders who were scored on the MnSOST-R by two different state evaluators who testified during depositions or trials (i.e., not the correctional staff who scored MnSOST-R for the present study), the single evaluator intraclass correlation coefficient (ICC) for absolute agreement (ICC A,1) for MnSOST-R total score was .68 (95% CI = .42 to .83).[5] Thus, there appears to be a moderate level of agreement (though somewhat lower than in more controlled MnSOST-R studies; Barbaree, Seto, Langton, & Peacock, 2001; Epperson et al., 2003) when

[4] In our sample, staff administered the MnSOST-R regardless of whether an offender's sexual offenses involved incest or not. Because sexual offenses against family members are not specifically designated in the offense record (i.e., they are not differentiated from other sexual offenses against children), we are unable to identify and exclude these incest offenders from our sample. However, reviewing a subsample of case files, and consulting with the relevant correctional staff, suggested that it was an extremely small minority of our sample whose charges involved instances of incest.

[5] The sample used to calculate the rater agreement coefficient for the MnSOST-R (and STATIC-99, below) partially overlaps with the sample Murrie et al. (2009) used to calculate a similar rater agreement coefficient. The data reported here include 12 new cases for the MnSOST-R (and 15 new cases for the STATIC-99).

MnSOST-R scores come from two evaluators retained by the same side in Texas SVP cases. Rater agreement is lower when the two MnSOST-R scores come from opposing (petitioner and respondent) evaluators (ICC A,1 = .38 to .48; see Murrie et al., 2009). Although these rater agreement data do not come from TDCJ staff evaluators, they are for offenders from the same population who are evaluated (by PhD or MD level evaluators) for the same purpose.

STATIC-99. The STATIC-99 is a 10-item ARAI designed to predict sexual recidivism. As the name of the measure indicates, all of the predictors assessed by the STATIC-99 are static in nature (see Hanson & Thornton, 1999, 2000). Examples include the following: young age at release, prior convictions for nonsexual violence, number of prior sentencing dates, history of offending against unrelated victims, history of offending against male victims, and number of prior offenses. Each of the items except number of prior offenses is scored as 1 (*present*) or 0 (*absent*). Scores for the number prior offenses item range from 0 (*none*) to 3 (*4 or more convictions, or 6 or more charges*). Scores on the STATIC-99 can range from 0 to 12, with scores of one or below indicating low risk, scores of two or three indicating moderate-low risk, scores of four or five indicating moderate-high risk, and scores of six or above indicating high risk (Harris et al., 2003).

The TDCJ database contained STATIC-99 scores for 1,968 (99.8%) of the 1,971 offenders in the sample. Excluding those 43 pursued for civil commitment, scores were available for 1,925 (99.8%) of the remaining 1,928 offenders. The mean STATIC-99 score among offenders was 2.94 (*SD* = 1.86).

Again, this archival study did not allow for a formal reliability check. Therefore, we can only consider reliability data from committed offenders with two different state-evaluator STATIC-99 scores presented in depositions and trials. For the 30 offenders with two STATIC-99 total scores, the single evaluator intraclass correlation coefficient for absolute agreement (ICC A,1) was .61 (95% CI = .33 to .79). Thus, there appears to be a moderate level of agreement when STATIC-99 scores come from two evaluators retained by the same side in Texas SVP cases. Rater agreement is also in the moderate range when the two STATIC-99 scores come from opposing (petitioner and respondent) evaluators in Texas SVP cases (ICC A,1 = .58 to .64, see Murrie et al., 2009). These rater agreement values are somewhat lower than those typically reported in STATIC-99 studies. For example, Hanson and Morton-Bourgon (2009) reported a median rater agreement coefficient of .90 across 12 samples, with coefficients ranging from .63 to .97. Although these rater agreement data do not come from offenders included in the current study, they do come from offenders evaluated with the STATIC-99 as part of the same statewide evaluation process.

Recidivism. The Texas Department of Public Safety (DPS) provided a list of arrest charges in the state of Texas for all but 12 offenders. For each arrest, DPS provided a National Crime Information Center (NCIC) code number and arrest date. DPS provided information about all arrests prior to July 1, 2007.

Each NCIC code number corresponds to a specific type of criminal offense, which allowed us to group offenses into five main (nonoverlapping) categories: violent sexual (e.g., contact offenses such as rape, sexual assault, attempted rape, kidnapping of minor to sexually assault), noncontact sexual (e.g., exhibitionism, possession of child pornography, solicitation), violent nonsexual (e.g., murder, assault, robbery), nonviolent and nonsexual (e.g., weapons possession, substance related charges, driving violations, probation violation, obstruction), and sex offender registry violations (e.g., failing to register, failing to report every 90 days). Recidivism for noncontact sexual arrests was so uncommon (14 offenders, 0.7%) that we did not examine predictive effects for this type of offending. Because the Texas statute that guides the SVP evaluation process (Texas Health & Safety Code § 841.023 2000) specifies that individuals eligible for SVP commitment must be at risk for future acts of sexual violence, we focused on outcome measures related to sexual violence. However, we do report effects for a combined category of contact and non-contact sexual offenses to examine the ability of ARAIs to predict any type of sexual recidivism.

We defined recidivism as a dichotomous variable for the study analyses, with even a single arrest charge qualifying an offender as having recidivated. For the 1,928 offenders who were not committed as SVPs, the base rates of recidivism were: 2.6% ($n = 51$) for violent sexual recidivism, 3.2% ($n = 62$) for any sexual recidivism, 6.1% ($n = 117$) for violent nonsexual recidivism, 18.3% ($n = 353$) for nonviolent nonsexual recidivism, and 11.8% ($n = 227$) for sex offender registry violations. We also created a recidivism category that identified offenders who had *either* a violent sexual or violent nonsexual re-arrest (base rate = 8.3%, $n = 160$). The rationale for focusing on this combined violent recidivism category is that many violent arrest charges from sex offenders result from crimes that were sexually motivated (see Rice, Harris, & Lang, 2006). Although some violent offenses from sex offenders are truly nonsexual in nature, Rice et al. (2006) argue that the combination of violent and sexually violent arrests is a more accurate indicator of sexually motivated offending than sexually-violent arrests alone. Furthermore the combination of violent or sexually violent offenses is likely to be of the most concern from a public safety perspective, and the STATIC-99 authors now report recidivism rates for this combined category of offending in addition to reporting recidivism rates specific to sexual offending (see Helmus et al., 2009).

Prior arrests. DPS also provided information about offenders' prior arrests in Texas for 1,924 of the 1,928 offenders. We considered any prerelease arrest to be a prior arrest, so arrests related to index offenses are included in the prior arrests measure. Because all offenders had prior arrests, we analyzed prior arrests as a continuous (count) variable. We grouped prior arrests into the same arrest categories used for post-release arrests. The mean number of prior arrests was 1.76 ($SD = 1.36$) for violent sexual offenses, .09 ($SD = .42$) for nonviolent sexual offenses, .46 ($SD = .99$) for violent nonsexual offenses, 2.22 ($SD = 1.69$) for violent sexual or violent nonsexual offenses, 2.56 ($SD = 3.44$) for nonviolent nonsexual offenses, and .04 ($SD = .23$) for sex offender registry violations. The total number of prior arrests ranged from 1 to 55, with a mean of 4.95 ($SD = 4.21$).

We also grouped offenders into those whose prior arrests indicated that they had been arrested for a sexual offense against a child ($n = 1{,}105$, 57.3%) and those who had no arrests for a sexual offense against a child ($n = 819$, 42.5%).

The analyses reported in this study focus on the total number of prior arrests and history of offending against children as measures of offense history.[6]

Results

Risk Scores and Offender Progress Through the SVP Evaluation System

Recall that the Texas SVP evaluation system follows a process that identifies increasingly narrow subgroups of offenders. Sex offenders who are potentially eligible for civil commitment are administered the STATIC-99, and MnSOST-R at the earliest stage of the screening process. Those with two or more qualifying sexual offenses may then be referred to an outside evaluator for an assessment of "behavioral abnormality" (i.e., whether the offender has a requisite disorder that might qualify for civil commitment). Of those evaluated for a "behavioral abnormality," a subset (i.e., those with a potentially qualifying requisite disorder) are then referred to a prosecution unit, who then selects an even smaller subset (historically, around fifteen per year) to pursue for civil commitment.

Table 1 lists mean risk scores for: those offenders initially screened for commitment but not referred for a "behavioral abnormality assessment"; those offenders screened and referred for a "behavioral abnormality assessment" but not finally pursued for civil commitment; and those who progressed through both evaluation stages and were pursued for SVP commitment.

The findings in Table 1 suggest offenders with higher ARAI scores were more likely than other offenders to progress further through the SVP evaluation process. For the STATIC-99, offenders who underwent behavioral abnormality evaluations had higher scores than offenders who did not proceed past the initial screening stage, regardless of whether their cases did (d = 1.85) or did not (d = 1.00) proceed to trial. Moreover, offenders whose cases proceeded to trial had significantly higher STATIC-99 scores than offenders who underwent behavioral abnormality evaluations, but were not further pursued for commitment (d = .92). Likewise for the MnSOST-R, offenders who underwent behavioral abnormality evaluations had higher scores than offenders who did not proceed past the initial screening stage, regardless of whether their cases did (d = 1.18) or did not (d = .86) proceed to trial. For those who underwent behavioral abnormality evaluations, MnSOST-R scores tended to be somewhat higher for those pursued for commitment (d = .33), although this difference was not large enough to reach statistical significance.

Progress through the SVP evaluation process was not associated with age at release or total number of prior arrests (see Table 1). However, progress through the process was associated with having a history of sexual offending against children, $\chi^2(2, N = 1967) = 39.92, p < .01$, and offender race, $\chi^2(4, N = 1960) = 52.33, p < .01$. Those with a history of sexual offenses against children were more likely than other offenders to be referred for a behavioral abnormality evaluation (whether eventually pursued for commitment or not) compared with other offenders (odds ratio

[6] We examined predictive effects for specific types of prior arrests (e.g., sexually violent, violent, nonviolent) and found that total number of prior arrests performed similarly to or outperformed these more specific prior offense variables. The only exception to this pattern was prior sexual offending against children, which is included in the analyses reported in Results.

Table 1
Risk Scores and Background Characteristics for Offenders Based on How Far They Progressed Through the Sexually Violent Predator (SVP) Evaluation and Commitment Process

Study variable	Final stage of SVP evaluation process[a]			Comparison
	Screened only	Behavioral abnormality evaluation	SVP trial	
STATIC-99				
Mean	2.64_a	4.25_b	5.91_c	$F(2, 1965) = 181.99^{**}$
SD	1.77	1.67	1.64	partial $\eta^2 = .16$
n	1,573	352	43	
MnSOST-R				
Mean	4.71_a	9.09_b	10.74_b	$F(2, 1607) = 106.61^{**}$
SD	5.11	5.09	4.35	partial $\eta^2 = .12$
n	1,283	289	38	
Age				
Mean	42.98	42.60	44.24	$F(2, 1968) = 0.39$
SD	12.36	11.24	10.71	partial $\eta^2 < .001$
n	1,576	352	43	
Prior arrests				
Mean	4.88	5.26	5.95	$F(2, 1964) = 2.36$
SD	4.07	4.78	3.73	partial $\eta^2 = .002$
n	1,572	352	43	

Note. MnSOST-R = Minnesota Sex Offender Sex Offender Screening Tool–Revised. Means in the same row with different subscripts differ significantly at the $p < .01$ level.
[a] Each offender who proceeded through multiple stages of the SVP evaluation process contributed a score to only one evaluation stage group (the final stage for that offender).
$^{**} p < .01$.

[OR] = 2.03). Regarding race, a smaller proportion of Hispanic offenders (10.8%) were referred for behavioral abnormality evaluations (whether eventually pursued for commitment or not) than black (27.1%) or white (22.7%) offenders.

Type of Discharge, Risk, and Offense Characteristics

Offenders may leave prison under conditions of mandatory supervision, or through a discharge with no supervision conditions. Those who were discharged and those released under mandatory supervision differed on most risk, background, and offense characteristics (see Table 2). Most importantly, reoffense rates were lower for offenders released under mandatory supervision. ORs for these differences indicated that the odds that discharged offenders would reoffend were anywhere from 1.75 (violent nonsexual recidivism) to 2.95 (sexually violent recidivism) times the odds that those released under mandatory supervision would reoffend.

Offenders released under mandatory supervision tended to have more prior arrests ($d = .20$) and higher risk scores (STATIC-99 $d = .33$; MnSOST-R $d = .36$) than discharged offenders, although these effects were relatively small in size. A greater proportion of White (55.7%) and Black (49.4%) offenders were released under mandatory supervision than Hispanic (40.6%) offenders, $\chi^2(2, N =$

Table 2
Risk Scores, Background Characteristics, and Recidivism Rates by Release Type

Study variable	Release type: Mandatory supervision	Release type: Discharge	Comparison	Effect size
STATIC-99				
Mean (*SD*)	3.24 (1.91)	2.64 (1.76)	$t(1920) = 7.18^{**}$	$d = .33$
MnSOST-R				
Mean (*SD*)	6.53 (5.24)	4.60 (5.36)	$t(1568) = 7.22^{**}$	$d = .36$
Age at release				
Mean (*SD*)	45.42 (10.89)	40.48 (12.82)	$t(1923) = 9.09^{**}$	$d = -.41$
Prior arrests				
Mean (*SD*)	5.38 (4.36)	4.53 (3.98)	$t(1919) = 4.48^{**}$	$d = .20$
Prior sex offense against child	51.8%	62.9%	$\chi^2(1) = 24.49^{**}$	OR = 1.58
Recidivism				
Sexually violent	1.4%	3.9%	$\chi^2(1) = 12.14^{**}$	OR = 2.95
Any sexual	2.1%	4.3%	$\chi^2(1) = 7.67^{**}$	OR = 2.11
Violent non-sexual	4.5%	7.6%	$\chi^2(1) = 8.17^{**}$	OR = 1.75
Violent or sexually violent	5.8%	10.8%	$\chi^2(1) = 16.09^{**}$	OR = 1.98
Registration violation	7.6%	15.9%	$\chi^2(1) = 31.17^{**}$	OR = 2.27
Non-violent non-sexual	12.4%	24.2%	$\chi^2(1) = 44.99^{**}$	OR = 2.26

Note. d = Cohen's d; MnSOST-R = Minnesota Sex Offender Screening Tool–Revised; OR = odds ratio, with values greater than 1.00 indicating higher likelihood for arrest in the Discharge group. Sample size (n) values for Mandatory supervision and Discharge group were, respectively: STATIC-99 (n = 953, 969), MnSOST-R (n = 747, 823), age at release and all recidivism variables (n = 954, 971), prior arrests and child offender (n = 952, 969).
** $p < .01$.

1914) = 32.74, $p < .01$. Discharged offenders were younger than mandatory supervision offenders ($d = -.41$) and were more likely to have prior sexual offenses against children (OR = 1.58).

Univariate Predictors of Recidivism

Because there were significant differences in risk scores, age at release, prior offending, and recidivism rates between the two release groups (i.e., mandatory supervision vs. discharge), we calculated predictive validity coefficients separately for each of these groups and for the sample as a whole. We calculated predictive coefficients for the ARAIs (i.e., the STATIC-99 and the MnSOST-R), but also for several easily recorded offender background characteristics (e.g., age at release, prior arrests, prior offending against children). This allows us to gauge the predictive value of specialized measures, as compared with basic background characteristics. We calculated two univariate effects for each predictor: AUC and Cohen's d (see Table 3). In the context of recidivism research, an AUC value indicates the likelihood that a randomly selected recidivist will score higher on the measure than a randomly selected non-recidivist. AUC values of .50 are equiv-

Table 3
Risk Scores, Age at Release, and Number of Prior Arrests as Predictors of Postrelease Arrests

	Mandatory supervision (n = 954)		
Recidivism type/predictor	AUC	*SE*	*d*
Sexually violent		(base rate = 1.4%)	
STATIC-99	.55	.08	.24
MnSOST-R	.54	.10	−.03
Age at release	.59	.09	−.23
Number of prior arrests	.53	.07	.04
Any sexual		(base rate = 2.1%)	
STATIC-99	.57	.06	.32
MnSOST-R	.59	.08	.19
Age at release	.63*	.07	−.40
Number of prior arrests	.48	.07	−.09
Violent nonsexual		(base rate = 4.5%)	
STATIC-99	$.49_a$	.04	−.05
MnSOST-R	.45	.05	−.17
Age at release	.72**	.04	−.73
Number of prior arrests	.59*	.05	.36
Violent or sexually violent		(base rate = 5.8%)	
STATIC-99	$.50_a$	.03	.02
MnSOST-R	.46	.04	−.16
Age at release	.69**	.04	−.62
Number of prior arrests	.58*	.04	.30
Sex offender registry violation		(base rate = 7.7%)	
STATIC-99	$.50_a$	.03	−.04
MnSOST-R	.58	.04	.24
Age at release	.66**	.03	−.53
Number of prior arrests	.60**	.03	.41
Nonviolent nonsexual		(base rate = 12.4%)	
STATIC-99	$.51_a$	.03	.04
MnSOST-R	.57*	.03	.21
Age at release	.68**	.02	−.63
Number of prior arrests	.64**	.03	.48

Note. AUC = Area under the receiver operatic characteristic curve; MnSOST-R = Minnesota Sex Offender Screening Tool–Revised. Cohen's *d* is an effect size for comparisons of scale scores for those who were and were not rearrested. AUC values in the same row with different subscripts are different from one another at the $p < .05$ level. Asterisks indicate whether the AUC value is significantly different from chance ($^{*}p < .05$, $^{**}p < .01$). Sample size values for mandatory supervision, discharge, and all offender samples, respectively: STATIC-99 (n = 953, 969, 1,925), MnSOST-R (n = 747, 823, 1,572), age at release (n = 954, 971, 1,927), number of prior arrests (n = 952, 969, 1,924).

alent to chance prediction, with values greater than .50 indicating that higher risk scores are associated with recidivism and values lower than .50 indicating that lower risk scores are associated with recidivism (which would be unexpected for the ARAI measures in this study). In other words, a value below .50 for an ARAI would indicate that a randomly selected recidivist would be more likely to have a lower score on the ARAI that a randomly selected nonrecidivist. Cohen's *d* is an effect size for comparing mean scores, reflecting the number of pooled standard deviation units between means for the two groups being compared (e.g.,

Table 3 (*continued*)

Discharge (n = 971)			All offenders (N = 1,928)		
AUC	*SE*	*d*	AUC	*SE*	*d*
	(base rate = 3.9%)			(base rate = 2.6%)	
$.62^{*}$	.05	.45	$.58^{*}$	.04	.30
.50	.05	−.02	.49	.05	−.10
$.60^{*}$	.04	−.37	$.64^{**}$	.04	−.44
.53	.05	.10	.51	.04	.03
	(base rate = 4.3%)			(base rate = 3.2%)	
$.63^{**}$	.05	.49	$.60^{**}$	.04	.36
.51	.05	.15	.52	.04	.01
$.60^{*}$	.04	−.35	$.64^{**}$	.03	−.43
.53	.05	.11	.50	.04	.00
	(base rate = 7.6%)			(base rate = 6.1%)	
$.63_{b}^{**}$	.03	.46	$.57^{*}$	.02	.21
$.61^{**}$	.04	.33	.54	.03	.11
$.65^{**}$	.03	−.50	$.69^{**}$	.02	−.63
$.66^{**}$	.03	.53	$.62^{**}$	.03	.42
	(base rate = 10.8%)			(base rate = 8.3%)	
$.62_{b}^{**}$	.03	.44	$.57^{**}$	.02	.22
$.57^{*}$	.03	.21	.52	.03	.03
$.64^{**}$	.04	−.50	$.68^{**}$	.02	−.60
$.63^{**}$	.03	.41	$.60^{**}$	.02	.32
	(base rate = 15.9%)			(base rate = 11.8%)	
$.61_{b}^{**}$	.02	.39	$.55^{**}$	.02	.16
$.58^{**}$	.03	.26	$.55^{*}$	.02	.17
$.64^{**}$	.02	−.48	$.67^{**}$	.02	−.57
$.61^{**}$	.03	.41	$.59^{**}$	.02	.35
	(base rate = 24.2%)			(base rate = 18.3%)	
$.64_{a}^{**}$	.02	.49	$.57^{**}$	.02	.23
$.62^{**}$	.02	.41	$.58^{**}$	.02	.26
$.68^{**}$	.02	−.62	$.70^{**}$	.02	−.69
$.68^{**}$	.02	.59	$.65^{**}$	.02	.48

recidivists vs. non-recidivists). In the current study, positive *d* values indicate that recidivists had a higher mean score than non-recidivists, whereas negative values indicate that non-recidivists had a higher mean score. We expected positive *d* values and *AUC* values greater than .50 for all variables except age at release, because existing research suggests that the likelihood of recidivism decreases as age increases (a negative relation). However, to facilitate comparison of AUC values for age to AUC values from other predictors, we have reflected the AUC values for age so that significant negative associations between age and recidivism result in AUC values greater than .50. The sign of the d value indicates whether an increase in the predictor was associated with an increase (positive sign) or decrease (negative sign) in recidivism.

We present the univariate effects for risk measures in Table 3 to show the overall pattern of effects across recidivism categories and release groups. Table 3 also summarizes effects from analyses examining whether risk and offender variables were significantly better at predicting recidivism in the discharge subsample than the mandatory supervision subsample (i.e., non-overlapping 95% CIs). With one exception (described below), questions about whether one measure was a significantly better predictor than another within the same group of offenders were examined through multivariate analyses.[7]

The findings in Table 3 were striking in that they revealed consistently modest effects for ARAIs, especially when effects were calculated based on the entire sample (as opposed to release groups). For example, the largest AUC value for an ARAI risk-related measure in the total sample of offenders was .60 ($d = .36$), for the STATIC-99 predicting any sexual recidivism. Although the STATIC-99 was statistically significant predictor of the recidivism categories most relevant to sex offender screening in the total sample (sexually violent, combination of violent or sexually violent), the MnSOST-R was not (see Table 3). Indeed, the MnSOST-R was a somewhat stronger predictor of nonviolent recidivism (AUC = .58, $p <$ .01) than types of recidivism that included sexual offenses (AUC range = .49 to .54, $p > .05$).

Predictive effects did tend to be smaller for offenders who were released under mandatory supervision. This was especially true for the STATIC-99, which was a significantly better predictor of most types of recidivism in the discharged offender sample than in the mandatory supervision sample. The largest AUC value for an ARAI risk-related measure was observed in the discharge sample (AUC = .64, $d = .49$), for the STATIC-99 predicting nonviolent nonsexual recidivism. For the two recidivism categories most relevant to sex offender screening (sexually violent, combination of violent or sexually violent), the largest AUC value for an ARAI was .62 ($d = .45$), for the STATIC-99 predicting violent or sexually violent recidivism in the discharge sample. In the mandatory supervision sample, neither of the ARAIs was a statistically significant predictor of recidivism categories that included sexual offenses.

The two most consistent predictors of recidivism were age at release and total number of prior arrests. Age at release produced the largest effect in the entire study, with an AUC value of .72 ($d = -.73$) for predicting violent recidivism in the mandatory supervision sample. Moreover, both age at release and total number of prior arrests were consistent moderate-sized predictors of recidivism in

[7] Although it is possible to calculate a significance test for the difference between two AUC values within the same group of offenders (see Hanley & McNeil, 1983), six significance tests would be needed to examine these differences for only one type of recidivism in one (sub)sample. With three samples (total, mandatory supervision, discharge) and six recidivism categories, 108 significance tests would be needed to examine these differences. By chance alone, we would expect about six of these comparisons to reach statistical significance. Because we had no a priori rationale for making only a select set of comparisons, we opted to use multivariate analyses to address most questions concerning the potential difference in effects between measures.

both release type samples, although prior arrests were not predictive of recidivism arrests that were clearly sexual in nature (violent sexual, any sexual). In the total sample, age at release was a significantly stronger predictor of sexually violent recidivism than the STATIC-99 and MnSOST-R ($Z > 5.00$ for both measures, $p < .01$; see Hanley & McNeil, 1983). Thus, age at release outperformed these two ARAIs that include items assessing age at release, even when the analysis focused on the type of offenses used to develop the ARAIs.

We used χ^2 analyses to examine the relation between two categorical study variables (prior sexual offending against children, offender race) and sex-offense related recidivism in the total offender sample. With one exception, which is noted below, effects for these categorical measures were similar in the mandatory supervision and discharge samples.

Prior sexual offending against a child was not a statistically significant predictor of sexually violent recidivism, $\chi^2(1, N = 1924) = 0.07$, $p = .93$, OR = .98, or sex offender registry violations, $\chi^2(1, N = 1924) = 3.26$, $p = .07$, OR = 1.30. Prior sexual offending against a child was a statistically significant, but small, predictor of the combined category of violent or sexually violent recidivism, $\chi^2(1, N = 1924) = 3.94$, $p = .05$, OR = .72. Those with no prior offenses against children were somewhat more likely (9.8%) to recidivate than offenders with a history of offenses against children (7.2%). However, this effect was more prominent in the discharge group, where 13.9% of offenders with no history of offending against children recidivated and 9.0% of those with a history of offending against children recidivated, $\chi^2(1, N = 969) = 5.64$, $p = .02$, OR = .61. In the mandatory supervision group, 6.5% of offenders with no history of offending against children recidivated, compared with 5.1% of offenders with a history of offending against children, $\chi^2(1, N = 952) = 0.94$, $p = .33$, OR = .76.

Survival Analysis

Although neither ARAI was a strong univariate predictor of recidivism using AUC analyses, the STATIC-99 appeared to outperform the MnSOST-R for predicting the types of recidivism most relevant to SVP evaluations (sexually violent recidivism and the combination of violent or sexually violent recidivism). This finding raises the question of whether Texas evaluators gain anything by administering the MnSOST-R in addition to the STATIC-99. In other words, does the MnSOST-R account for a significant amount of variance in recidivism once STATIC-99 scores are taken into account? The univariate analyses also suggested that offender background characteristics, such as age at release and prior arrests, may be more useful predictors of recidivism than either ARAI, which raises the question of whether the STATIC-99 accounts for a significant amount of variance in recidivism once offender background characteristics are taken into account? We used Cox regression, a form of survival analysis, to examine these two questions.

Cox regression is an appropriate model for examining recidivism when follow-up time differs for offenders and when some cases are right-censored (i.e., some offenders do not recidivate by the end of the study period). Cox regression provides information about whether predictors entered into the regression model predict the outcome (e.g., recidivism) better than a model based on time alone.

This includes a chi-square test examining the statistical significance of the entire model, as well as a test of individual predictors. Information about the effect of individual predictors is provided by a regression coefficient and an OR or hazard ratio (with 95% CIs). The OR tells the odds of an event occurring faster (ratios greater than 1.00) or slower (ratios less than 1.00) given the covariate (Garson, 2008). For a dichotomous predictor like release type (0 = *mandatory supervision*, 1 = *discharge*), an OR of 1.20 would mean that the odds are 1.20 to 1.00 that a person in the discharge group is more likely to reoffend than a person in the mandatory supervision group. For a continuous predictor like the STATIC-99, an OR of 1.20 would mean that the probability of recidivism increases by 20% for each one point increase on the STATIC-99 (see Tabachnick & Fidell, 2007).

We focused the Cox regression analyses on the two recidivism variables that are most relevant to SVP risk assessment: violent sexual recidivism and the combination of violent or sexually violent recidivism. All continuous measures were centered before being entered into the Cox regression models. Dummy coding was used for categorical variables (see Table 4).

Incremental validity of MnSOST-R over STATIC-99. We focused on the combination of violent and sexually violent offending for analyses examining the incremental validity of the MnSOST-R over the STATIC-99 because the Mn-SOST-R was not a significant univariate predictor of sexually violent recidivism in either release group. We used STATIC-99 scores as the initial predictor in the Cox regression model. Because STATIC-99 scores were somewhat more effective for predicting recidivism in the discharge sample, we also entered release type

Table 4
Summary of Cox Regression Analyses for Predicting Recidivism

Recidivism/predictor	*B*	*SE*	OR	Lower	Upper	χ^2
Incremental validity of MnSOST-R over STATIC-99						
Violent or sexually violent recidivism (N = 1,569)						31.39**
STATIC-99	.03	.09	1.03	0.86	1.22	
Release type	.77**	.20	2.17	1.47	3.19	
STATIC-99* release type	.20*	.10	1.22	1.00	1.50	
MnSOST-R	−.02	.02	0.98	0.94	1.02	
Incremental validity of STATIC-99 over offender characteristics						
Sexually violent recidivism (N = 1,922)						26.51**
Age at release	−.02	.01	0.98	0.95	1.00	
Release type	1.18**	.34	3.27	1.68	6.34	
STATIC-99	.18*	.08	1.20	1.03	1.40	
Violent or sexually violent recidivism (N = 1,907)						91.08**
Age at release	−.05**	.01	0.95	0.94	0.97	
Release type	.63**	.18	1.87	1.32	2.66	
Prior arrests	.60**	.01	1.06	1.04	1.08	
STATIC-99	.06	.05	1.07	0.97	1.17	

Note. OR = odds ratio; MnSOST-R = Minnesota Sex Offender Screening Tool–Revised. Release type (0 = *mandatory supervision*, 1 = *discharge*).
* $p < .05$. ** $p < .01$.

(0 = *mandatory supervision*, 1 = *discharge*) and the interaction between release type and STATIC-99 scores into the initial regression model. We then entered MnSOST-R scores as the final predictor. Because the focus of this analysis was on the incremental validity of the MnSOST-R (final predictor) we present results for only the final regression model in Table 4. The findings in Table 4 indicate that the MnSOST-R did not make a significant contribution to the prediction of violent and sexually violent recidivism once STATIC-99 scores and release type were entered into the model. The only significant effect for an ARAI was the interaction between STATIC-99 scores and release type, which is consistent with the earlier finding that STATIC-99 scores were more effective predictors of recidivism for offenders who had been discharged, as opposed to those released on mandatory supervision.

Incremental validity of STATIC-99 over offender characteristics: Predicting sexually violent recidivism. STATIC-99 scores, age at release, and release type (mandatory supervision vs. discharge) were the only statistically significant univariate predictors of sexually violent recidivism. We conducted several preliminary analyses to determine whether the final Cox regression models needed to contain higher order interaction terms involving age at release or release type. Hanson (2006) found a curvilinear effect for the relation between age at release and sexual recidivism in a Cox regression model that controlled for STATIC-99 scores. We attempted to replicate Hanson's finding by constructing a Cox regression model that included STATIC-99 scores, age at release, and age at release squared (to model a curvilinear effect). The curvilinear effect did not approach statistical significance ($B < .001$, $SE = .001$, OR = 1.00, OR 95% CI = .998 to 1.002). For release type, we considered whether STATIC-99 scores were more effective for predicting recidivism for discharged offenders by examining a model that included STATIC-99 scores, release type, and the interaction between STATIC-99 scores and release type. The interaction term did not approach statistical significance ($B = -.13$, $SE = .17$, OR = 0.89, OR 95% CI = .64 to 1.24).

We examined the incremental validity of the STATIC-99 over offender background characteristics by adding offender characteristics (age at release, release type) as predictors in an initial model, and STATIC-99 scores as an additional predictor in a second model. The Cox regression results in Table 4 are from the second model, examining the incremental validity of the STATIC-99 over offender background characteristics. In the final model, the STATIC-99 (OR = 1.20) continued to be a small, but statistically significant predictor of sexually violent recidivism after age at release and release type were entered into the model. Release type was a strong predictor of recidivism in this model, with the odds of recidivating for offenders who were discharged being 3.27 higher than the odds of recidivating for offenders who were released under mandatory supervision. Age at release was no longer a statically significant predictor of sexually violent recidivism, which may be due in part to age being considered on the STATIC-99 and the fact that offenders who were discharged were, on average, 5 years younger than those released under mandatory supervision.

Incremental validity of STATIC-99 over offender characteristics: Predicting violent or sexually violent recidivism. The offender characteristics that were statistically significant univariate predictors of violent or sexually violent recidivism were release type, total number of prior arrests, prior sexual offending against children, and age at release.[8] We examined the incremental validity of the STATIC-99 over offender background characteristics by adding offender characteristics (age at release, release type, prior arrests, prior offending against a child) as predictors in an initial model, and STATIC-99 scores as an additional predictor in a second model. Initial analyses indicated that prior sexual offending against a child was not a significant predictor of recidivism once age at release and release type were taken into account. Thus, we removed this predictor from the final regression model.

The final Cox regression model for predicting the combination of violent or sexually violent recidivism is summarized in Table 4. The STATIC-99 was not a statistically significant predictor of the combined category of violent or sexually violent recidivism once offender characteristics were taken into account. However, younger age at release, more prior offenses, and being discharged as opposed to released under mandatory supervision were all significant predictors of recidivism.

Local Norms for Actuarial Risk Measures

Clinicians who use the STATIC-99 and MnSOST-R often report an offender's score on the measure alongside the observed recidivism rate for offenders with that score in the instrument development sample. Indeed, the 2003 STATIC-99 coding rules (Harris et al., 2003) provided risk communication templates to help evaluators communicate their findings in terms of an expected recidivism rate from the normative sample. However, there are many reasons to suspect that these normative rates do not apply to other jurisdictions, especially when the base rate of reoffending in a jurisdiction differs from the normative sample (see Mossman, 2006). The low base rate of sexually violent recidivism in our Texas sample (2.6%) suggests that the recidivism rates from the STATIC-99 normative sample (18% at 5 years) will overestimate the likelihood of recidivism for most Texas offenders.

Table 5 contains 5-year recidivism rates for the STATIC-99 among Texas offenders, for both sexually violent recidivism and the combination of violent or sexually violent recidivism. When calculating these rates, we included only offenders with at least 5 years of opportunity to reoffend (i.e., at least 5 years between release and collection of recidivism data), to facilitate comparison with the 2003 STATIC-99 norms. We defined recidivism as a reoffense within 5 years.

Table 5 also lists 5-year normative recidivism rates (i.e., sexual recidivism) for STATIC-99 reported by the instrument authors in 2003 (Harris et al., 2003) and 2009 (Helmus et al., 2009). A primary reason for the development of the 2009 norms was that recent declines in crime rates (including sexual offending) make it likely that the recidivism rates in the 2003 STATIC-99 manual overestimate rates of recidivism in contemporary samples of offenders (Helmus et al., 2009). Although the 2009 norms are based on ongoing research and may be updated in the near future, Helmus et al. (2009) recommend using the new norms instead of

[8] As with sexually violent recidivism, there was not a statistically significant curvilinear relation between age at release and violent or sexually violent recidivism when controlling for STATIC-99 scores ($B = -.001$, $SE = .001$, OR = 1.00, OR 95% CI = .998 to 1.001).

Table 5

Recidivism Rates for STATIC-99 Scores: Reoffense Within 5 Years

	STATIC-99 Norms			Texas		
Recidivism type/STATIC-99 score	2003	2009: Routine risk samples	2009: High risk samples	Mandatory supervision	Discharge	All offenders
Sexually violent recidivism						
1 (*n* = 65, 60, 125)	6.0	6.2	10.3	1.5	5.0	3.2
2 (*n* = 91, 75, 166)	9.0	4.3	12.8	1.1	2.7	1.8
3 (*n* = 64, 49, 113)	12.0	5.7	15.7	1.6	4.1	2.7
4 (*n* = 129, 77, 206)	26.0	7.7	19.1	1.6	3.9	2.4
5 (*n* = 58, 35, 93)	33.0	10.2	23.1	1.7	8.6	4.3
6+ (*n* = 50, 14, 64)	39.0	NA	NA	4.0	14.3	6.3
Base rate	18.0	7.0	21.9	2.0	5.5	3.4
Violent or sexually violent recidivism						
1 (*n* = 65, 60, 125)		10.8	20.0	3.1	11.7	7.2
2 (*n* = 91, 75, 166)		13.6	24.1	7.7	12.0	9.6
3 (*n* = 64, 49, 113)		17.0	28.6	6.3	10.2	8.0
4 (*n* = 129, 77, 206)		21.1	33.7	7.8	16.9	11.2
5 (*n* = 58, 35, 93)		25.8	39.1	3.4	17.1	8.6
6+ (*n* = 50, 14, 64)		NA	NA	8.0	28.6	12.5
Base rate		18.8	35.3	7.4	14.8	10.4

Note. NA = not applicable. *n* values are for the Texas sample. Normative 5-year recidivism rates for sexually violent recidivism are from Harris et al. (2003) and Helmus et al. (2009). Base rates for the 2009 norms are fixed follow-up rates from an October 2008 document titled *Detailed Recidivism Tables* posted on the static99.org Web site. The 2009 norms do not report scores for the group of offenders scoring > 6. Offenders were not included in the calculation of the Texas rates unless they had at least 5 years of opportunity to reoffend, defined as years after release for the offense(s) that led to the STATIC-99 being scored by correctional officials. Offenders whose first reoffenses occurred after 5 years were counted as nonrecidivists for the rates reported.

the 2003 norms. Table 5 lists recidivism rates for specific STATIC-99 scores from both the 2003 and 2009 norms. The 2003 and 2009 norms differ in several ways. The 2003 norms are based on observed rates of recidivism, whereas the 2009 norms are predicted recidivism rates based on the results of logistic regression analyses. The 2003 norms grouped offenders with scores of 6 or above into a single high-risk group, whereas the 2009 norms report recidivism rates for all possible STATIC-99 scores. The 2003 norms provided recidivism rates for only sexually violent offending, whereas the 2009 norms report predicted recidivism rates for both sexually violent and violent (including sexually violent) recidivism. Finally, the 2009 norms report predicted recidivism rates for two groups of offenders: routine risk offenders and preselected high-risk offenders. High risk offenders are those who warranted "exceptional measures," such as being ordered to treatment or denied statutory release (see Helmus et al., 2009).

Comparison to 2003 norms. As Table 5 reveals, the 5-year rates for sexually violent recidivism in Texas do not match those from the STATIC-99 normative sample described in the 2003 scoring manual. Overall, observed rates of reoffending were much lower than those reported in the 2003 manual, which is not surprising given the relatively low base rates of sexually violent recidivism in our sample. For example, examining the Texas sample overall, only 6.3% of high-risk (STATIC-99 score ≥6) offenders recidivated, compared with 39.0% in the 2003 normative sample. Examining the discharge sample only, the observed recidivism rate for high-risk offenders (14.3%) was closer, but still well below, the 39% figure reported in the 2003 STATIC-99 norms. For offenders released under mandatory supervision, the mismatch was even more dramatic, with only 4.0% of high-risk offenders recidivating.

Although the mismatch between the Texas recidivism rates and the 2003 normative recidivism rates in Table 5 appears to suggest poor performance for the STATIC-99 in Texas, a measure may have similar discriminative properties in two samples, but lead to different recidivism rates simply because of differences in the base rate of recidivism (Mossman, 2006). Because of this well-known relation between base rates and recidivism rates, Mossman (2006) argued that researchers and clinicians should also consider the discriminative properties of the measure when examining its stability across samples. Specifically, Mossman (2006) recommended calculating a likelihood ratio for a positive result (LR+) for each score on the risk measure. In the context of recidivism and STATIC-99 scores, this LR provides information about the extent to which the odds of recidivism increase when the offender has that score. Mathematically, the LR is equal to sensitivity divided by one minus specificity. Ideally, the discriminative properties of the STATIC-99 would be similar in both the normative and Texas samples, which would mean that the apparent mismatch in recidivism rates was due primarily to the low base rate of recidivism in Texas. In other words, we examined whether specific STATIC-99 scores might still discriminate between recidivists and non-recidivist to the same extent that they did in the STATIC-99 normative sample, even if the actual rates of reoffense are much lower in Texas than in the STATIC-99 normative sample.

Table 6 provides LR values for STATIC-99 scores in Texas, based on 5-year recidivism rates. Because these LR values are based on a subset of the entire sample (those with at least 5 years of opportunity to reoffend), Table 6 also

Table 6
Likelihood Ratios for STATIC-99 Scores: Reoffense Within 5 Years

		Texas		
Recidivism type/ STATIC-99 score	2003 norm (95% confidence interval)	Mandatory supervision	Discharge	All offenders
Sexually violent recidivism				
AUC (*SE*) (*N* = 501, 346, 847)		.56 (.09)	.61 (.07)	.58 (.06)
1 (*n* = 65, 60, 125)	0.30 (0.16–0.58)	0.96	1.09	1.09
2 (*n* = 91, 75, 166)	0.46 (0.29–0.73)	0.68	0.56	0.63
3 (*n* = 64, 49, 113)	0.66 (0.44–0.97)	0.98	0.88	0.93
4 (*n* = 129, 77, 206)	1.65 (1.24–2.19)	0.97	0.84	0.85
5 (*n* = 58, 35, 93)	2.34 (1.59–3.44)	1.08	1.93	1.54
6+ (*n* = 50, 14, 64)	3.00 (2.19–4.13)	2.57	3.44	2.29
Violent or sexually violent recidivism				
AUC (*SE*) (*N* = 501, 346, 847)		.53 (.05)	.63* (.07)	.58* (.06)
1 (*n* = 65, 60, 125)		0.50	0.88	0.80
2 (*n* = 91, 75, 166)		1.31	0.91	1.10
3 (*n* = 64, 49, 113)		1.05	0.76	0.89
4 (*n* = 129, 77, 206)		1.32	1.36	1.29
5 (*n* = 58, 35, 93)		0.56	1.38	0.97
6+ (*n* = 50, 14, 64)		1.37	2.68	1.47

Note. AUC = Area under the receiver operatic characteristic curve. *n* values are for the Texas sample. Normative 5-year likelihood ratios with confidence intervals are from Mossman (2006), based on the normative data reported in Hanson and Thornton (2000) and Harris et al. (2003). Offenders were not included in the calculation of these rates unless they had at least 5 years of opportunity to reoffend, defined as years after release for the offense(s) that led to the STATIC-99 being scored by correctional officials. Offenders whose first reoffenses occurred after 5 years were counted as nonrecidivists for the rates reported.
* $p < .05$.

provides AUC values for the STATIC-99 in this smaller sample. The LR values for STATIC-99 scores in Texas generally fall outside the 95% CIs for STATIC-99 LR values (from Mossman, 2006)[9] in the 2003 normative sample (see Table 6). For low STATIC-99 scores, the LR values tended to exceed those based on the normative sample. For high STATIC-99 scores, the LR values tended to be lower than those expected based on the normative data. For the total sample, only three of the six LR values fell within the expected range of those from the normative sample. In the mandatory supervision sample and the total sample, the LR values do not consistently increase as STATIC-99 scores increase, indicating that the STATIC-99 did not correctly rank the likelihood of recidivism. One promising finding for the STATIC-99 is that the Texas LR values generally matched those from the normative sample for predicting sexually violent recidivism for "high-

[9] Mossman (2006) used recidivism rates from Hanson and Thornton (2000) to calculate LR values for the STATIC-99. The recidivism rates reported by Hanson and Thornton (2000) are identical to those in the 2003 STATIC-99 scoring guidelines (Harris et al., 2003).

risk" offenders (STATIC-99 ≥ 6.0). Indeed, in all three groups (total sample, mandatory supervision, discharge), the LR value for offenders scoring in the high-risk (scores > 6) range fell within the 95% CIs from the normative sample, as reported by Mossman (2006).

Together, the recidivism rate and LR values in Tables 5 and 6 suggest that the STATIC-99 performed differently in our Texas sample than it did in the 2003 normative sample. The recidivism rates in Table 5 do not match the normative 2003 rates because of the low base rate of reoffending and because of the different discriminative properties for several scores on the measure. Nevertheless, the data in Table 5 indicate that the rate of recidivism did tend to increase as STATIC-99 scores increased, especially for offenders who were discharged. In the discharge sample, LR values tended to increase as scores increased as well. Across all groups, scores of 6 or higher did indicate an increased risk of recidivism (LR > 2.00). This pattern of findings is consistent with the earlier univariate effects for the STATIC-99, but also indicate that score interpretations based on the normative data often do not apply to recidivism outcomes in Texas. Recidivism rates from the 2003 norms clearly do not apply to Texas, and, for STATIC-99 scores below 5, a higher score does not always suggest a higher likelihood of recidivism than a lower score. Scores of 6 or above, however, do suggest a relatively higher likelihood of recidivism.

Comparison to 2009 norms. The Texas recidivism rates for each STATIC-99 score group appeared much more consistent with the 2009 STATIC-99 norms than the 2003 norms, at least when Texas offenders are compared with the subsample of the 2009 norm group labeled "routine risk" rather than "preselected high risk." These new STATIC-99 norms still suggested a higher rate of recidivism than was observed in Texas, especially for Texas offenders released under mandatory supervision, but the difference in recidivism rates was less pronounced for the 2009 norms than the 2003 norms. At the time of this study, the data needed to calculate LR values for the 2009 norms were not available.

Recidivism rates for the MnSOST-R in Texas. We did not conduct in-depth analyses of recidivism rates and LR values for the MnSOST-R because it was not a statistically significant predictor of sexually violent recidivism. However, the observed rates of recidivism for the MnSOST-R did not come close to those reported for the normative sample (6 years of follow-up). For example, only 3.8% of Texas offenders with at least 6 years of opportunity reoffend who were classified as high risk by the MnSOST-R were rearrested for a sexually violent offense within 6 years. But data in the MnSOST-R manual would have led us to expect a rate 15 times higher (57%). The recidivism rate for MnSOST-R high-risk offenders was consistent in the mandatory supervision (3.6%) and discharge (4.0%) samples.

Discussion

Actuarial risk assessment instruments are a primary part of the process designed to identify and civilly commit offenders as SVPs in the United States. Yet our field has surprisingly little data addressing the accuracy of ARAIs when used for this purpose. Therefore, we investigated ARAIs as applied in routine procedures in a large, state-wide sample of sexual offenders.

ARAIs and Decisions About Offenders

Decision makers appeared to use offenders' STATIC-99 and MnSOST-R scores to help decide which offenders should be considered for SVP commitment. Offenders who were pursued for commitment had much higher scores on these measures than offenders who were screened with ARAIs, but not evaluated further (ds > 1.15). Among offenders who underwent "behavioral abnormality evaluations," the STATIC-99 was also a strong predictor of which offenders the state decided to pursue for commitment ($d = .92$). Of course, these findings cannot prove evaluation and commitment decisions for these offenders were *explicitly* based on ARAI scores. Decision makers had access to detailed records about the offenders, and it is possible that they ignored the ARAI scores but made decisions based on the same behaviors and characteristics scored on the ARAIs, or perhaps other highly correlated factors. Nevertheless, the large effects in this study, along with the standard practice of administering ARAIs in this system, suggest that decision makers probably used offenders' ARAI scores to help make SVP related decisions.

Predicting Recidivism with ARAIs and with Background Variables

Because of the role that ARAIs appear to play in selecting SVP candidates, it is important to know how well they predict SVP-relevant outcomes, particularly sexually violent offending. The most recent meta-analysis of the STATIC-99 and MnSOST-R suggest that both measures are moderate to strong predictors of sexual recidivism (Hanson & Morton-Bourgon, 2009). However, none of the validation or cross-validation studies of these measures provide findings directly relevant to the context of SVP screening in Texas. In this particular context, neither the STATIC-99 nor the MnSOST-R proved to be a strong predictor of recidivism. The STATIC-99 was a statistically significant predictor of sexually violent recidivism ($d = .30$) and the combination of violent or sexually violent recidivism ($d = .22$), but these effects were much smaller than the mean meta-analytic effect size of $d = .67$ reported by Hanson and Morton-Bourgon (2009). Moreover, our Cohen's d effect sizes for the STATIC-99 fell well outside the 95% CI in that meta-analysis (.62 to .72). Although Hanson and Morton-Bourgon (2009) did find that STATIC-99 effects were smaller in U.S. samples, the STATIC-99 effect size in our Texas sample was even smaller than the mean effect for U.S. samples ($d = .60$, CI not reported). Overall, predictive effects in our study were more similar to (although still smaller than) those from the two peer-reviewed published studies examining the STATIC-99 in U.S. samples ($d = .50$ in Bartosh et al., 2003; $d = .43$ in Sreenivasan et al., 2007).

Across all study analyses, findings were weaker for the MnSOST-R, which was no better than chance for predicting sexually violent recidivism ($d = -.10$) or the combination of violent and sexually violent recidivism ($d = .03$) in the overall sample.

The most consistent predictors of recidivism were young age at release, number of prior arrests, and release type. Regarding release type, offenders who were discharged were more than twice as likely to be arrested for a sexually violent offense than offenders released under mandatory supervision. The STATIC-99 was a stronger predictor of recidivism for discharged offenders than those released under mandatory supervision, but effects for sexually violent recidivism ($d = .45$) and the combination of violent or sexually violent recidivism ($d = .44$) were still below the 95% CIs reported by Hanson and Morton-Bourgon (2009).

"Field Validity" of ARAIs

What can we conclude from the relatively weak performance of the MnSOST-R and the well-established STATIC-99 in this large sample, "real world," sex offender screening process? First, *we cannot assume that actuarial measures perform similarly across all contexts*. Second, evaluators cannot claim that a particular ARAI score from an offender corresponds to the same likelihood of recidivism that it did in the original test-development sample. In short, a measure's "field validity,"—or validity as applied in routine practice in new contexts—may differ from the measure's performance in controlled research studies. These simple conclusions have weighty implications for evaluators and systems that rely on ARAIs to draw important inferences about sex offenders. We hope these simple conclusions can buffer against some of the common misunderstandings or misuses of ARAIs.

At the same time, we also want to warn against potential misunderstanding or misuse that may arise from our findings. We emphasize that these results cannot reasonably be used to criticize Texas, or similar state systems, for adopting the STATIC-99 or the MnSOST-R as part of their routine procedures in the late 1990's and early 2000's. At that point, sufficient literature existed to demonstrate that ARAIs were more accurate than unstructured clinical judgment (Grove & Meehl, 1996), and to demonstrate that the STATIC-99 and the MnSOST-R, though far from ideal, were the best-validated available ARAIs for sexual offenders (Doren, 2002). The decision to adopt these ARAIs for routine procedures was a reasonable decision based on the data available at the time. However, as our findings illustrate, any decision to adopt ARAIs *must* be followed by local validity studies to examine how ARAIs work in a new context (see also Hanson & Morton-Bourgon, 2009).

For this reason, we do not claim that our results reflect the way ARAIs perform in every other U.S. state SVP screening process. But our results do suggest that every other state should examine ARAI and reoffense data specific to their own context, rather than assume ARAIs will perform just as they did in normative samples. Until this ideal is common practice, jurisdictions are left to make cautious inferences about which existing recidivism studies best generalize to their jurisdiction (Conroy & Murrie, 2007; Helmus et al., 2009). In considering whether these Texas results can generalize to other contexts, there are several features of our data and study context that are important to consider.

Post-release supervision and opportunity for reoffending. In this study, there were three ways that offenders were released: discharge, mandatory supervision, or civil commitment as an SVP. Ostensibly, the restrictions associated with civil commitment and mandatory supervision mean that civilly committed offenders had fewer opportunities to reoffend than those released under mandatory supervision, who in turn had fewer opportunities than those who were discharged. At the same time, there were significant differences in ARAI scores for these three groups. Civilly committed offenders had much higher ARAI scores than all other offenders, and those placed on mandatory supervision had somewhat higher scores than those who were discharged. Thus, offenders whom ARAIs suggested were at the highest risk for reoffending tended to be released with more community supervision. Indeed, recidivism rates were noticeably lower in the mandatory supervision sample than the discharge sample, even though offenders in the mandatory supervision sample tended to have higher ARAI scores. *These findings suggest that mandatory supervision may be an effective mechanism for reducing reoffending*.

From a predictive validity perspective, the fact that those with higher ARAI scores tended to receive more intensive post-release supervision may be a possible explanation for the low base rates of recidivism and modest to small ARAI effects in this study. If supervision was successful, the result would be a lower than expected rate of reoffending for those with high ARAI scores, who tended to be in the mandatory supervision sample. This pattern is evident in the recidivism rates for the STATIC-99 in Table 5. Although the varying levels of supervision may help to explain the small effect sizes in this study, it does not mean that the findings should be dismissed as an unfair test of predictive validity, particularly when we are concerned with a measure's field validity, or performance within the contexts in which it is typically used. Rather, these findings underscore that it is essential for clinicians to consider the context to which an offender is released (see Helmus et al., 2009). The data in Table 5 suggest that a clinician's opinion about offender risk should probably vary depending on how that offender is going to be released. Of course, the data in Table 5 speak only to 5-year recidivism rates for the STATIC-99, and it is unclear if the difference in base rates between the discharged and mandatory supervision offenders will continue over time. It may be that the recidivism base rate for the mandatory supervision group will come to approach that of the discharged group as mandatory supervision requirements expire.

Although community supervision appears to be a logical explanation for the lower rate of recidivism and smaller predictive validity effects in the mandatory supervision sample, an alternate explanation is that offenders released under mandatory supervision were truly lower risk offenders. Two factors argue for this interpretation. First, those released under mandatory supervision had behaved well-enough in prison to earn "good time" credit and early release. Second, those released under mandatory supervision tended to be older (M = 45.42 years) than those who were discharged (M = 40.48). Ultimately, findings from the current study cannot identify the true reason or reasons for the smaller effects in the mandatory supervision sample, and it may be that a combination community supervision, older age, and the ability to earn early release credit all played some role.

Low base rates. The supervision requirements that applied to more than half of the sample may be one reason why the sample's base rate for sexually violent recidivism (2.6%) and the combination of violent or sexually violent recidivism (8.3%) was lower than that observed in other studies. For example, Hanson and Morton-Bourgon (2009) reported overall recidivism base rates (average 5.8 years) of 11.5% for sexual offending and 19.5% for violent offending, across their combined sample of 28,757 offenders from 16 countries. Other factors that likely contributed to the low observed base rates of recidivism in Texas may relate to some of the well-known limitations of relying on formal re-arrest records. Obviously, our observed base rates do not reflect reoffenses that law enforcement did not detect. And, unlike the nationalized criminal justice databases common to Canada and Europe, our source for rearrest data was only state-wide rather than national, a problem common to U.S. studies. So, we cannot identify discharged offenders who were rearrested in another state (this limitation is not relevant to offenders on mandatory supervision, who were required to remain in state). However, it is important to emphasize that the lower base rates we observed are probably not attributable only to methodological limitations. Rates of sexual offending (whether measured by victim report or arrest statistics) have decreased in recent years (see, e.g., Federal Bureau of Investigation, 2007; Finkelhor & Jones, 2006; Jones & Finkelhor, 2003; LaFond, 2005), and rates of sexual recidivism observed in other U.S. states are similar to our findings. For example, recent data from Washington (Barnoski, 2005) reveals a 5-year sexual reoffense rate of 2.7% for offenders released to the community between 1994 and 1998. Thus, our low rate (2.6%) over an average of 5 years appear fairly typical of sexual offenders released from U.S. state prisons during recent history.

Whatever the reason for the low base rates of sexually violent recidivism, low base rates make predictions about recidivism difficult. This can be seen most clearly in Table 5, where offenders scoring in the high-risk range on the STATIC-99 were more likely to reoffend than those scoring in lower risk ranges, yet high-risk offenders recidivated at a much lower rate than those with similar scores in the 2003 normative sample, and even those in the 2009 normative sample. Recent materials from the STATIC-99 authors clearly inform evaluators that recidivism rates for particular scores depend on the base rate of recidivism and encourage evaluators to consider what STATIC-99 scores say about *relative* risk, rather than absolute recidivism estimates (Hanson & Thornton, 2008; Helmus et al., 2009). In other words, those with higher scores are more likely to reoffend than those with lower scores. This general trend applies, to some extent, to the STATIC-99 in this study. The STATIC-99 authors are in the process of developing relative risk estimates for specific STATIC-99 scores, as well as materials to help evaluators use them appropriately (Helmus, 2009; Helmus et al., 2009; Harris, Helmus, Hanson, & Thornton, 2008). Thus, findings from Texas may prove to be more consistent with forthcoming recommendations for relative risk interpretations than the effect size and recidivism rate findings suggest.

"Field reliability" of ARAI scores. An additional reason for the weaker than expected performance of the ARAIs in our sample may be measurement error. We were not able to formally study reliability or scoring accuracy among the raters who scored these ARAIs. But, if two evaluators assign different scores to the same offender, it would suggest that either one or both of the evaluators are scoring the measure incorrectly or that scoring instructions for the ARAI do not generalize well to that field setting. Regardless of the reason for scoring disagreements, they increase measurement error, and a high level of measurement error makes prediction difficult. Indeed, the Hanson and Morton-Bourgon (2009) meta-analysis found that predictive validity effect sizes for ARAIs increased as rater agreement increased (r = .21). Although this pattern did not

apply to the STATIC-99, the median level of rater-agreement for STATIC-99 studies in the meta-analysis was very high (.90).

A recent study of ARAI rater agreement among a small sample of offenders being pursued for SVP commitment in Texas (Murrie et al., 2009) suggests that rater-agreement may not be as high in routine SVP evaluation practice as it is in many research settings. Again, we do not know the degree of interrater reliability among raters in this sample, a limitation common to studies that rely on institutional data and examine actual practice. But low reliability is one possible explanation for the modest predictive validity in this study. Nevertheless, even if rater agreement was very low, it would still be important to study the validity of the real-world scores (reliable or not) because these scores were actually used to make decisions about offenders. The fact that we do not have a great deal of information about the reliability of these scores does not diminish the finding that they do not work as well as "optimally designed" research would lead us to expect.

It will be important for state systems and researchers to examine field reliability as part of future studies of field validity. In other words, one way of understanding how well a measure works in a particular context is to examine how rigorously and reliably users employ the measure.

Policy Implications

Few psychological tests or measures are as prominent in public policy as are ARAIs for sexual offenders. But our results highlight some of the complexity in adopting and relying on ARAIs as a matter of policy. *Systems and institutions cannot assume that actuarial data from other jurisdictions apply perfectly to their own*. For example, our data reveal that Texas sex offenders recidivated at substantially lower rates than the widely referenced recidivism rates from the STATIC-99 2003 normative data would lead us to expect. Lower rates of sexual reoffense are excellent news, of course, but these lower rates are a primary reason why published recidivism rates for the two ARAIs in this study overestimated recidivism rates in Texas. Lower reoffense rates and lower ARAI accuracy in this Texas sample underscore the need for all U.S. jurisdictions that rely on ARAIs to carefully examine how well these measures perform in their own jurisdictions. To make accurate and well-informed decisions about offenders in their jurisdictions, systems will need to collect data specific to their jurisdiction and monitor the field reliability and field validity of ARAIs as applied in their system. Decisions about how to improve local procedures will likely be difficult. On the one hand, it will be important for systems to remain open to changing risk assessment procedures over time; the science of risk assessment evolves rapidly and new data may warrant changes in procedure. On the other hand, it will be important that systems do not too hastily adopt new procedures without careful validity studies. Ideally, systems will collect pilot data for new or revised measures (while continuing to use existing measures) in order to decide whether and how to use new measures.

Another broad implication from this study involves the importance of considering release conditions during risk assessment. Offenders in the mandatory supervision group demonstrated unusually low rates of reoffense. Risk assessment should consider context (e.g., Borum & Verhaagen, 2006; Conroy & Murrie, 2007), and these data illustrate that decision-makers should probably draw different inferences about an offender's level of risk depending on whether he is discharged from prison or released under mandatory supervision. More generally, although we emphasize this was not a formal study of supervision effectiveness, these results certainly appear to suggest that mandatory supervision arrangements reduced reoffense, as compared with simple discharge from prison.

Practice Implications

How can these results inform clinicians who use ARAIs in routine forensic evaluations? First, there appear to be implications for selecting actuarial measures. A fair view of the existing literature suggests that no single measure outpredicts all others across all contexts, and combinations of measures generally do not outperform the single best predicting measure (Hanson & Morton-Bourgon, 2009; Seto, 2005). However, local validity studies, such as this one, might reveal that some measures perform better than others in certain contexts. Within the Texas system, no measure was a strong predictor, but the STATIC-99 appeared to perform better than the MnSOST-R. Our analyses also suggest there was no advantage to combing the STATIC-99 and MnSOST-R. Thus, for evaluators who opt to use ARAIs in this Texas system, the STATIC-99 seems a more defensible choice; we could identify no reason to administer the MnSOST-R. Ideally, evaluators will select only instruments that have demonstrated validity in their jurisdiction, or comparable jurisdictions.

For evaluators who use the STATIC-99, there are also some clear implications. Our results suggest that many of the *specific inferences* about precise recidivism rates and predictive accuracy, which evaluators draw based on the STATIC-99 normative data, may not hold true in new jurisdictions. Risk estimates based on the original STATIC-99 normative data grossly overestimated the true re-arrest rates in this sample. This does not mean that *general* inferences about the implications of ARAI scores cannot hold true. For example, offenders scoring in the high-risk rage (scores $\geq$ 6) on the STATIC-99 were, indeed, somewhat more likely to reoffend than offenders with lower scores.

Currently, evaluators who use the STATIC-99 face a decision as to whether they reference the norms published in 2003 or the recently disseminated 2009 norms (see Helmus et al., 2009; www.static99.org) when communicating their findings. Certainly in our sample, the 2009 norms, which reflect lower recidivism rates from more contemporary samples, came much closer to the recidivism rates we observed in Texas. Given overall declines in sexual recidivism across the United States, and the few state data available (e.g., Barnoski, 2005) it seems reasonable to recommend that evaluators use the 2009 norms. However, there are separate 2009 norms for high-risk and routine risk offenders, and the nature of the Texas sample (offenders convicted of more than one sexual offense, and at least one sexually violent offense) would suggest that Texas recidivism rates should have more closely matched the high-risk norms than the routine risk norms. The finding that Texas rates more closely matched the routine risk norms once again highlights the need for local validity studies to help evaluators make informed decisions about the appropriate use of ARAI norms.

Conclusion

Correctional systems, forensic evaluators, and courts have—with good reason—increasingly relied upon ARAIs to draw inferences about sexual offenders' risk for reoffense. However, the inferences one can draw from ARAI scores may be more or less accurate depending on the extent to which they are administered in a context similar to the context in which the ARAIs were developed. This study suggests that two popular ARAIs tend to work less well, and overestimate reoffense risk, in one large U.S. jurisdiction with fairly low recidivism rates. Results illustrate a crucial need for "local data" to inform decisions about whether and how to use risk measures.

References

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Association for the Treatment of Sex Abusers. (2001). *Civil commitment of sexually violent predators.* Retrieved from http://www.atsa.com/ppcivilcommit.html

Barbaree, H. E., Seto, M. C., Langton, C. M., & Peacock, E. J. (2001). Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders. *Criminal Justice and Behavior, 28,* 490–521.

Barnoski, R. (2005). *Sex offender sentencing in Washington State: Recidivism rates.* Olympia: Washington State Institute for Public Policy (Document No. 05–08–1203). Retrieved from http://www.wsipp.wa.gov/series.asp?seriesid=1

Bartosh, D. L., Garby, T., Lewis, D., & Gray, S. (2003). Differences in the predictive validity of actuarial risk assessments in relation to sex offender type. *International Journal of Offender Therapy and Comparative Criminology, 47,* 422–438.

Borum, R., & Verhaagen (2006). *Assessing and managing violence risk in juveniles.* New York: Guilford Press.

Campbell, T. (2007). *Assessing sex offenders: Problems and pitfalls* (2nd ed.). Springfield, IL: Charles C. Thomas.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Conroy, M. A., & Murrie, D. C. (2007). *Forensic assessment of violence risk: A guide to risk assessment and risk management.* Hoboken, NJ: Wiley & Sons.

Council on Sex Offender Treatment. (2007, October). *Heath & Safety Code, Chapter 841. Civil Commitment of Sexually Violent Predators.* Retrieved from http://www.dshs.state.tx.us/csot/csot_solaws.shtm

Doren, D. M. (2002). *Evaluating sex offenders: A manual for civil commitment and beyond.* London: Sage Publications.

Epperson, D. L., Kaul, J. D., Huot, S. J., Hesselton, D., Alexander, W., & Goldman, R. (1998). *Minnesota Sex Offender Screening Tool-Revised (MnSOST-R).* St. Paul, MN: Minnesota Department of Corrections.

Epperson, D. L., Kaul, J. D., Hout, S., Goldman, R., Hesselton, D., & Alexander, W. (2003). *Minnesota sex offender screening tool–Revised (MnSOST-R) technical paper: Development, validation, and recommended risk level cut scores.* St. Paul, MN: Minnesota Department of Corrections. Retrieved from http://www.psychology.iastate.edu

Federal Bureau of Investigation. (2007). *Crime in the United States, 2007.* Retrieved from http://www.fbi.gov/ucr/cius2007/index.html

Finkelhor, D., & Jones, L. (2006). Why have child maltreatment and child victimization declined? *Journal of Social Issues, 62,* 685–716.

Garson, G. D. (2008). Cox regression. Retrieved from http://faculty.chass.ncsu.edu/garson/PA765/cox.htm

Gottfresson, S. D., & Moriarty, L. J. (2006). Statistical risk assessment: Old problems and new applications. *Crime & Delinquency, 52,* 178–200.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law, 2,* 293–323.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12,* 19–30.

Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology, 148,* 839–843.

Hanson, R. K. (2006). Does the Static-99 predict recidivism among older sexual offenders? *Sex Abuse, 18,* 343–355.

Hanson, R. K., Harris, A. J. R., Scott, T. L., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project* (Corrections Research User Report No. 2007–05). Ottawa, Ontario, Canada: Public Safety Canada.

Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis. *Psychological Assessment, 21,* 1–21.

Hanson, R. K., & Thornton, D. (1999). Static-99: *Improving actuarial risk assessments for sex offenders* (User Report 99–02). Ottawa: Department of the Solicitor General of Canada.

Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior, 24,* 119–136.

Hanson, R. K., & Thornton, D. (2008, October). *Recommendations for interpreting multiple norms for the STATIC-99.* Paper presented at the Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Atlanta, GA.

Harris, A. J. R., Helmus, L., Hanson, R. K., & Thornton, D. (2008, October). *Are new norms needed for the STATIC-99?* Paper presented at the Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Atlanta, GA.

Harris, A., Phenix, A., Hanson, R. K., & Thornton, D. (2003). *Coding rules for the Static-99 Revised. Corrections research: Manuals and forms*. Ottawa: Department of the Solicitor General of Canada.

Helmus, L. (2009). *Relative risk for STATIC-99 scores.* Retrieved from http://www.static99.org/pdfdocs/relativerisktablejanuary2009.pdf

Helmus, L., Hanson, R. K., & Thornton, D. (2009). *The Forum: Newsletter for the Association for the Treatment of Sexual Abusers, 21,* 38–45.

Interstate Commission for Adult Offender Supervision. (2007). *Sex offender assessment information survey.* Retrieved from www.interstatecompact.org/resources/surveys/survey_results/SexOffender_Assessment_042007.pdf

Jackson, R. L., & Hess, D. T. (2007). Evaluation of civil commitment of sex offenders: A survey of experts. *Sexual Abuse: A Journal of Research and Treatment, 19,* 425–448.

Jackson, R. L., & Richards, H. J. (2008). Sex offender civil commitment evaluations. In

R. L. Jackson (Ed.), *Learning forensic assessment* (pp. 183–209). New York: Routledge/Taylor & Francis.

Janus, A. E., & Prentky, R. A. (2003). Forensic use of actuarial risk assessment with sex offenders: Accuracy, admissibility, and accountability. *American Criminal Law Review, 40,* 1443–1499.

Jones, L. M., & Finkelhor, D. (2003). Putting together evidence on declining trends in sexual abuse: A complex puzzle. *Child Abuse & Neglect, 27,* 133–135.

Kansas v. Hendricks, 117 S. Ct. 2072 (1997).

LaFond, J. Q. (2005). *Preventing sexual violence.* Washington, DC: American Psychological Association.

Levenson, J. S. (2004). Sexual predator civil commitment: A comparison of selected and released offenders. *International Journal of Offender Therapy and Comparative Criminology, 48,* 638–648.

Levenson, J. S., & Morin, J. W. (2008). Factors predicting selection of sexually violent predators for civil commitment. *International Journal of Offender Therapy and Comparative Criminology, 50,* 609–629.

McGrath, R. J., Cumming, G. F., & Bouchard, B. L. (2003). *Current practices and trends in sexual abuser management: The Safer Society 2002 Nationwide Survey.* Brandon, VT: Safer Society Press.

Meehl, P. E. (1954). *Clinical vs. statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis: University of Minnesota Press.

Miller, H. A., Amenta, A. E., & Conroy, M. A. (2005). Sexually violent predator evaluations: Empirical evidence, strategies for professionals, and research directions. *Law and Human Behavior, 29,* 29–54.

Monahan, J. (2003). Violence risk assessment. In A. M. Goldstein (Ed.), *Handbook of psychology: Forensic psychology* (Vol. 11, pp. 527–540). New York: John Wiley & Sons.

Monahan, J. (2006). A jurisprudence of risk assessment forecasting harm among prisoners, predators, and patients. *Virginia Law Review, 92,* 391–435.

Mossman, D. (2006). Another look at interpreting risk categories. *Sexual Abuse: A Journal of Research and Treatment, 18,* 41–64.

Mossman, D. (2008). Analyzing the performance of risk assessment instruments: A response to Vrieze and Grove (2007). *Law and Human Behavior, 32,* 279–291.

Murrie, D. C., Boccaccini, M. T., Turner, D., Meeks, M., Woods, C., & Tussey, C. (2009). Rater (dis)agreement on risk assessment measures in sexually violent predator proceedings: Evidence of adversarial allegiance in forensic evaluation? *Psychology, Public Policy, and Law, 15,* 19–53.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). New York: Harcourt.

Phenix, A., Hanson, R. K., & Thornton, D. (2000). *Coding rules for the Static-99. Corrections research: Manuals and forms.* Ottawa: Department of the Solicitor General of Canada.

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. *Law and Human Behavior, 29,* 615–620.

Rice, M. E., Harris, G. T., & Lange, C. (2006). Violent sex offenses: How are they best measured from official records? *Law and Human Behavior, 30,* 525–541.

Seto, M. (2005). Is more better? Combining actuarial risk scales to predict recidivism among adult sex offenders. *Psychological Assessment, 17,* 156–167.

Sreenivasan, S., Garrick, T., Norris, R., Cusworth-Walker, S., Weinberger, L. E., Essres, G., et al. (2007). Predicting the likelihood of future sexual recidivism: Pilot study findings from a California sex offender risk project and cross-validation of the Static-99. *Journal of the American Academy of Psychiatry and the Law, 35,* 454–468.

Tabachnick, B. G., & Fidell, L. S. (2007). Using multivariate statistics (5th ed.). Boston: Allyn & Bacon.

Texas Health & Safety Code § 841.000–841.150 (2000).

Va. Code Ann. § 37.2–903(c) (2005).

Vrieze, S. I., & Grove, W. M. (2008). Predicting sex offender recidivism. I. Correcting for item over-selection and accuracy overestimation in scale development. II. Sampling-error inducted attenuation of predictive validity over base-rate information. *Law and Human Behavior, 32,* 266–278.

Witt, P. H., & Conroy, M. A. (2009). *Evaluation sexually violent predators*. New York: Oxford University Press.

Wollert, R. (2002). The importance of cross-validation in actuarial test construction: Shrinkage in the risk estimates for the Minnesota Sex Offender Screening Tool-Revised. *Journal of Threat Assessment, 2,* 87–102.

Wollert, R. (2003). Additional flaws in the Minnesota Sex Offender Screening Tool-Revised: A response to Doren and Dow. *Journal of Threat Assessment, 2,* 65–78.

Wollert, R. (2006). Low base rates limit expert certainty when current actuarials are used to identify sexually violent predators. *Psychology, Public Policy, & Law, 12,* 56–85.

Woodworth, G. G., & Kadane, J. B. (2004). Expert testimony supporting post-sentence civil incarceration of violent sex offenders. *Law, Probability, & Risk, 3,* 221–241.

Received March 24, 2009
Revision received July 13, 2009
Accepted July 14, 2009 ■