2007

Adjusted actuarial assessment of sex offenders: the impact of clinical overrides on predictive accuracy

Kathleen Spencer Gore
Iowa State University

Adjusted actuarial assessment of sex offenders:
The impact of clinical overrides on predictive accuracy

by

Kathleen Spencer Gore

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Psychology (Counseling Psychology)

Program of Study Committee
Douglas L. Epperson, Major Professor
Fred Borgen
Norman A. Scott
Douglas G. Bonett
Matthew J. DeLisi

Iowa State University

Ames, Iowa

2007

UMI Number: 3274898

UMI Microform 3274898

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Although tragic, it is not unusual for the victims of particularly heinous and shocking crimes to have their names become synonymous with those crimes, linked forever to the perpetrator and their act of violence. In an attempt to honor these victims and prevent similar outrages in the future, such crimes usually also result in reactive measures. In 1993, 12-year old Polly Klaas became the symbol of a legislative juggernaut aimed at preventing the release of offenders like her killer, Richard Allen Davis, a man repeatedly paroled despite a lengthy and violent criminal history. Davis was sentenced to death for her kidnapping and murder and California passed its infamous "Three Strikes Law," mandating life imprisonment for most offenders after their third felony conviction.

Sadly, over the past decade, a particular class of crime has appeared repeatedly in the national media spotlight. The focus of dramatic attention has centered on the occurrence of recently released sexual offenders committing new offenses, often against children. In the past two years alone, nationwide attention has riveted on several such stories. In March 2005, 9-year old Jessica Lundsford disappeared from her father's home in Florida. Subsequently, John Couey, a 46-year old registered sex offender led authorities to her body and admitted to her rape and murder, for which he was convicted of in March 2007 (Anderson, March 15, 2007). The same week that police recovered the body of Jessica Lundsford, an Iowa girl, Jetseta Gage, was seen leaving her home with Roger Bentley, an acquaintance of the family and a registered sex offender. The remains of the 10-year old were discovered later and Bentley was convicted of first degree murder and kidnapping in January 2006 (McWilliams, February 1, 2006). Finally, in May 2005 a brother and sister in Idaho, Dylan and Shasta Groene, were discovered missing after authorities found their older

brother, their mother, and her boyfriend bludgeoned to death in the family's home. Six weeks later, 8-year old Shasta was spotted with 42-year old Joseph Duncan, a convicted sex offender with two outstanding arrest warrants. Nine-year old Dylan's body was found shortly afterward. Duncan pleaded guilty to three counts of murder in October 2006 and has also been indicted on federal charges, including murder and kidnapping. His federal trial has been set for January 2008. Duncan also faces prosecution for the abduction and murder of another child in California in 1997 (Associated Press, January 9, 2007).

Cases such as these fueled public sentiment and led to an outcry for actions to prevent similar tragedies. Demands for tougher legislation aimed at repeat sex offenders, including more stringent guidelines surrounding their release and supervision, were almost immediate. Politicians in Florida and Iowa were quick to introduce legislation named for Jessica Lundsford and Jetseta Gage respectively. The governor of Florida signed the Jessica Lunsford Act on May 2, 2005, which required 25-year minimum prison terms for those convicted in Florida of certain sex crimes against children and lifetime tracking by global positioning satellite once released (Vansickle, January 2, 2006). Jetseta's Bill," was introduced as federal legislation and called for an increase in the mandatory minimum sentences and penalties for offenders convicted of violent crimes against children. Originally, it also proposed an expansion of the death penalty in certain sexual offense cases, but that provision was dropped as part of a compromise early in 2006 (U.S. Federal News Service, May 4, 2006).

The culmination of many of these laws aimed at repeat offenders appeared in a sweeping and significant piece of federal legislation, which actually included "Jetseta's Bill," and went into effect on July 27, 2006. The Adam Walsh Child Protection and Safety Act,

which passed in the House of Representatives (H. Res. 4472, 2006) just one week after passing in the Senate by unanimous vote, was signed into law on the anniversary of the abduction of its namesake, son of activist John Walsh. While the law increases mandatory minimum incarcerations and penalties and widens funding for law enforcement to pursue computer-based sexual predators, many provisions in the act are aimed at improving tracking of released sex offenders. The legislation organizes sex offenders into three tiers and mandates that offenders in the third and most serious tier update their whereabouts every three months. Any offender who fails to register and/or update his or her information is guilty of a felony under the act. Perhaps one of the most noteworthy elements of the legislation is its creation of a national sex offender registry, which necessitates the posting of identical offender criteria, including name, address, date of birth, and photograph, by each state. The registry itself is named for Dru Sjodin, another victim of a recently released sex offender, and "Dru's Law," was the measure that originally pushed for such a mechanism.

Unfortunately, despite good intentions, these types of reactive responses by governments at both the state and federal level often lead to legislation that is flawed. First, the bills passed usually apply to <u>all</u> sex offenders, which impedes their effectiveness because the requirements are often not enough for truly high risk offenders. At the same time, they are usually too tough and expensive for low risk offenders. In some instances, low risk offenders may even have their risk increased by such measures. For example, juveniles are an inherently low risk group, but when adult policies are applied to them indiscriminately, a socially constructive reintegration into society becomes almost impossible. Consequently, their initial very low risk for reoffense may actually increase. Second, even if this type of legislation was only applied to high risk offenders, there is no evidence to support its

effectiveness. Proximity statutes that forbid sexual offenders from living with a certain distance of schools, day care centers, and certain other locations are one example. This legislation has proven to be extraordinarily expensive and difficult to enforce, and there is no evidence that it is effective. Such legislation assumes that sexual offenders offend only in their own immediate neighborhood and/or are incapable of traveling 2,000 feet. Empirical data are inconsistent with both assumptions. Instead of making communities safer, this expensive legislation has driven sexual offenders underground with registration compliance substantially declining.

Despite these concerns, it appears legislation such as the Adam Walsh Act, will continue to be implemented and supported in the foreseeable future. To increase the utility and applicability of such laws, it is necessary to identify those offenders most at risk to commit additional crimes. Doing so would allow additional resources, like increased supervision and community notification, both of which are required by law, to be directed towards such high-risk offenders, with the aim of preventing subsequent crimes. Additionally, if an offender recidivates despite the presence of added deterrents, such as intensive supervision, the argument may be made that he should now be the focus of more stringent sentencing. However, in order to proceed with such programming, useful and effective methods to assess the risk of sexual recidivism are required, in order to effectively identify high-risk offenders. Over the past decade, the field of sexual offender risk assessment has developed and grown in response to this critical need, with efforts continuously made to improve the methods used.

Historical Context of Sex Offender Risk Assessment

Despite the dramatic and intense media coverage and the resulting legislative action engendered by the tragic cases discussed above, concern about repeat offenders is not a new phenomenon. The possibility of criminals recently released from prison committing new acts of violence has obviously concerned the criminal justice system for years. Andrews and Bonta (1998) surmised that, "The prediction of criminal behavior is perhaps one of the most central issues in the criminal justice system. From it stems community safety, prevention, treatment, ethics, and justice" (p. 211). Because of these far-reaching implications, the legal system must rely on assessments of risk made by psychologists and other practitioners. Such assessments often influence various decisions, like determining sentence length and release or supervision guidelines. Given their criticality, assessments of risk have undergone intense examination, from both a legal standpoint and an empirical one.

Almost from the beginning, legal challenges were made regarding the constitutionality of violence prediction, a corollary of risk assessment. Opponents argued that predictions of future criminal behavior were so inaccurate that to give them a legal effect would violate the due process and equal protection guaranteed by the Fourteenth Amendment (Monohan, 1996). Despite this resistance, courts throughout the United States overwhelmingly dismissed this idea. Monohan (1996) referenced two cases in which the unpredictability of future behavior was rejected as a legal concern. In the first, seven justices of the Supreme Court denied the notion that it is impossible to predict future behavior and ruled that the likelihood of a defendant committing future crimes was an acceptable reason for imposing a death sentence (Jurek v. Texas, 1976). In a later case, Schall v. Martin (1984), Supreme Court Justice Rehnquist rejected the argument that predictions of future

criminal conduct were inherently unattainable from a legal standpoint. With the judicial support demonstrated by favorable rulings such as these, Grove and Meehl (1997) concluded, "... it seems well established that there is no constitutional impediment to using predictions of dangerousness in legal proceedings" (p. 36).

Once the constitutionality of laws that relied on violence prediction was upheld, the focus shifted to empirical questions, including how such predictions could be improved (Grisso & Appelbaum, 1992). As noted earlier, decisions based on violence predictions have crucial implications. Therefore, it is imperative that those decisions be based on the most accurate information and assessments. Researchers began to examine ways to enhance the accuracy of both, including studying factors that influence risk and evaluating the methods of prediction. Prior to this, risk assessment solely relied on the judgment of clinicians, who often used their experience and intuition as the only means to evaluate potential risk. According to Monahan and Steadman (1994), many researchers found the development of actuarial methods, which utilized the common risk factors found in empirical studies, to be the answer to improving clinical predictions. However, the clinical versus statistical debate has been on-going, and recently attempts have been made by some practitioners to combine actuarial tools with clinical judgment, which is discussed in a later section.

As mentioned previously, the origination of actuarial methods lay in the study of risk factors. Results from the MacArthur Violence Risk Assessment Study (Monohan & Steadman, 1994), on which many risk assessment tools for violence are based, demonstrated the ability of researchers to determine factors that influence risk. Using such risk factors, several tools now exist to assess an individual's risk for committing a new violent offense.

The most widely used violence risk assessment instrument, the Violence Risk Appraisal Guide (VRAG; Quinsey, Rice, Harris, & Cormier, 1998), is an example of such a tool.

Specialization of Sex Offender Risk Assessment

Assessing the risk of reoffense for sex offenders in particular presents additional challenges, however. Hanson and Bussiere (1998) noted that evidence indicates sexual offending is different from other crimes. This means that risk factors for violence and general recidivism may not be the same as those for sexual recidivism. For example, while general criminal history was the best predictor of general recidivism, a range of studies identified a number of stronger variables as potential contributors to the prediction of sexual recidivism. In their 1998 meta-analysis of 87 articles based on 61 data sets, Hanson and Bussiere found that in addition to general criminal history, the following variables were predictive of sexual recidivism: deviant sexual arousal, deviant sexual attitudes, number of prior sex offenses, victimization of a stranger, victimization of an unrelated person, victimization of a male, early onset of sexual offending, any sex offender treatment failure, young age, and never having been married. These findings were essentially replicated in Hanson and Morton-Bourgon's (2004) updated meta-analysis, with some additional clarification regarding deviant sexual arousal and attitudes. Because of differences such as these, the assessment of sexual offenders has necessarily developed into a specialized area (Becker & Murphy, 1998).

Specialization was also required as the need to assess risk in sex offenders increased dramatically with the advent of certain federal and state legislation, which were precursors to the recent legislation already discussed. Unfortunately, tragedies like the Lunsford, Gage, and Groene cases were not the first incidents of a recently released sexual offender

committing another offense. In the 1990s, a number of disturbing, high profile sexual crimes committed by offenders after their release from prison precipitated increased fear about repeat offenders. Although sex offenders as a group generally reoffend at a lower rate than other violent offenders, society seems particularly troubled by this population, at least in part due to the high number of children who are victimized (Becker & Murphy, 1998). In response to increasing public pressure, spearheaded by victims' families and their communities, lawmakers hurriedly enacted statutes aimed at protecting the public from known sex offenders. One of the most prominent, the federal legislation of "Megan's Law" in 1996, required all fifty states to implement registration and community notification laws regarding released sex offenders (Epperson, et al., 1998). Additionally, court decisions like Kansas v. Hendricks (U.S. Supreme Court, 1997), which upheld the Sexually Violent Predator Act in Kansas, opened the door for the use of civil commitment for sex offenders deemed to be a high risk for reoffense (Hanson, 1998). To effectively implement these initial laws, it became obvious that there was a need to identify those offenders most at risk for reoffending (Becker & Murphy, 1998) since such offenders warrant an increased amount of community notification and higher levels of supervision, up to and including possible civil commitment. As discussed previously, it is also essential to identify individuals at low risk to reoffend, as a means of protecting both those individuals and the communities to which they return. By effectively identifying low risk offenders, it is possible to facilitate their constructive re-entry into society and to prevent any increase in their risk of reoffense that could result from applying procedures that are unwarranted or ineffective. Effective implementation of risk assessment involves a balancing act between applying the necessarily more stringent criteria to high risk offenders, while minimizing the damage possible to low

risk offenders as a result of the application of overly harsh measures. With the continued introduction of legislation regarding registration, notification, and civil commitment for sex offenders, the stakes have become even higher for all parties involved. Therefore, accurate assessments of recidivism risk for sex offenders have become crucial, both for the sake of protecting the community as a whole and safeguarding individual liberty.

MnSOST-R. Unfortunately, at the time the first registration and notification laws were enacted in the mid-1990s, methods to assess sexual recidivism were woefully inadequate, relying predominantly on techniques that were often ineffective or required information that was frequently unavailable for many sex offenders. In response to the critical need for a more formal and uniform process to assess violent and predatory sex offenders, the Minnesota Sex Offender Screening Tool (MnSOST) was developed (see Epperson, Kaul, Huot, Hesselton, Alexander, & Goldman, 1998, 2000; Epperson, Kaul, Huot, Alexander, & Goldman, 2000; Epperson, Kaul, Huot, Goldman, & Alexander, 2003). This predictive instrument proved to be reliable and valid and utilized only information routinely available to corrections personnel.

The ultimate purpose of the MnSOST was to bring increased precision and utility to risk assessments of sex offenders and the risk management decisions based on those assessments. In addition, there were several specific objectives of the MnSOST. First, it sought to rely fully on actuarial data. Second, it endeavored to be relatively brief and simple to use. Finally, it would apply to all incarcerated sex offenders and predominantly rely on behaviorally anchored items.

The tool was developed from a sample of sex offenders released in Minnesota between 1988 and 1990. The development sample was comprised of 166 non-sexual

recidivists and 90 sexual recidivists, with sexual recidivism defined as re-arrest for a "hands-on" sex-related offense. Offenders were excluded if their only sex offenses were intrafamilial and none of those offenses were the equivalent of rape. In 1998, the MnSOST underwent a major revision, which refined empirical methods for selection and scoring. This research yielded the 16-item MnSOST-R (see Appendix A for score sheet descriptions of each item), which was presented at the 1998 annual research and treatment convention of the Association for the Treatment of Sexual Abusers (ATSA; Epperson, Kaul, & Hesselton, 1998). The MnSOST-R has since been adopted by several states as the primary tool used to assess the risk of sexual recidivism.

The reliability of the MnSOST-R has been demonstrated repeatedly and has been shown to be generally above .80 by both internal and external studies (Epperson, et al., 2003; Langton, Barbaree, Harkins, Seto, & Peacock, 2002). Specifically, in a comparison of eight instruments, Langton, et al. (2002) reported inter-rater reliabilities ranging from .75 to .94, with the MnSOST-R having a reliability of .83. In studies of the MnSOST-R alone, interclass correlations (ICC) for relative agreement ranged from .80 to .87 (Epperson, et al., 2003). The ICC for absolute agreement in the same studies ranged from .76 to .86.

Additionally, several studies have examined the validity of the MnSOST-R. Most commonly, the results were reported as Receiver Operator Characteristics (ROC) – Area Under the Curve (AUC). The ROC curve is generated by plotting the proportion of recidivists correctly classified as high risk (sensitivity) against the proportion of non-recidivists incorrectly classified as high risk (1-specificity) for each possible cut score. For each possible cut score, offenders scoring at or above the specified cut score were classified as high risk and offenders scoring below the specified cut score were classified as low risk.

The area under the ROC curve reflects the overall accuracy of the instrument. An ROC-AUC of .50 reflects chance level accuracy. Values significantly greater than .50 are significantly greater than chance. A value of 1.0 reflects perfect accuracy.

In the development sample for the MnSOST-R, Epperson, et al. (1998, 2003) reported an ROC-AUC of .77 for the total sample, .79 for rapists, and .74 for molesters. In three subsequent validation studies with different populations, the ROC-AUCs for the total samples were consistently at or above .73 (Epperson, et al., 2000, 2003). The first of these studies used 220 incarcerated sex offenders released from Minnesota prisons in 1992. The period of time offenders were at risk for recidivism was six years. In that study, the ROC-AUC was .73. A second study examined the validity of the MnSOST-R with a North Dakota prison sample. This study used 182 sex offenders released in North Dakota between 1989 and 1998. The median risk period for these offenders was eight years. With that sample, the ROC-AUC was .76. The third study used 271 offenders placed on probation in North Dakota for a sex offense from 1989 to 1998. The median period of time offenders were at risk for recidivism was just under eleven years. In that study, the ROC-AUC was .77.

Other researchers have also demonstrated the validity of the MnSOST-R. One such validation study of the MnSOST-R was conducted with a Canadian sample (Langton, Barbaree, Seto, Peacock, Harkins, & Hansen, 2007). This study expanded on the preliminary results from Barbaree, Seto, Langton, & Peacock (2001). In the updated, larger sample of 354 sex offenders who were at risk for an average of 5.9 years, which largely subsumed their first sample, the MnSOST-R yielded an ROC of .70.

The MnSOST-R, along with several other tools, was also successfully validated with a sample of 599 sexual offenders referred to the Massachusetts Treatment Center for evaluation

for possible civil commitment (Knight & Thornton, 2007). Because this sample included only those offenders considered for possible commitment, it largely excluded lower risk sexual offenders. Despite the resulting restricted range on the MnSOST-R, the area under the ROC curve was .68 at three-year follow-up and .67 at ten-year follow-up, and .66 at fifteen-year follow-up.

Bartosh, Garby, Lewis, and Gray (2003) conducted a validation of several risk assessment tools, including the MnSOST-R, with a sample of 186 sex offenders in the state of Arizona who were at risk for approximately 5 years. The MnSOST-R ROC of .58 missed the threshold for statistical significance in this study, though it was only slightly lower than the ROC values for the other instruments assessed. Several studies (Epperson et al., 2003, Langton et al., 2007, Barbaree et al., 2001, Knight & Thornton, 2007, & Bartosh, et al., 2003) also assessed other sex offender risk assessment tools in addition to the MnSOST-R, including the Rapid Risk Assessment for Sexual Offense Recidivism (RRASOR) and the Static-99, and none of the studies identified significant differences between the three tools.

Overall, the MnSOST-R has unmistakably contributed to a more formal and uniform review process for incarcerated sex offenders in Minnesota and a number of other states. Evidence indicates that the MnSOST-R, like other actuarial measures of sexual recidivism, provides a reliable and accurate assessment of risk when used appropriately and in keeping with specified guidelines.

The Clinical Versus Statistical Debate

Unfortunately, it has sometimes proven difficult to assure the appropriate and valid use of actuarial measures, such as the MnSOST-R. As mentioned previously, methods to assess risk prior to the mid-1990s were often inadequate and unreliable. The development of

actuarial tools, like the MnSOST-R, ushered in a new era of risk assessment. However, the debate between the relative accuracy of clinical judgment versus actuarial prediction, first seen in Meehl's classic book, *Clinical Versus Statistical Prediction* (1954), appears to be on-going. Initially, early predictions in most domains, including violence and recidivism, relied upon clinical judgment. This has been succinctly defined as "the typical procedure long used by applied psychologists and physicians, in which the judge puts data together using informal, subjective methods" (Grove, Zald, Lebow, Snitz, & Nelson, 2000, p. 19).

Unfortunately, despite being the longest-used method for the prediction of violence, numerous researchers have demonstrated several problems with unaided clinical judgment. Specifically, Douglas, Cox, and Webster (1999) cited clear and convincing findings of the inability of mental health professionals to accurately predict violence based solely on clinical judgment. More precisely, in his review of clinical prediction, Monohan (1981) found that only one in three positive predictions of violence by clinicians was accurate. Various explanations have been given for the problems clinicians face when trying to make predictions. Hilton and Simmons (2001) pointed out a fundamental difficulty, noting that "clinicians using only unaided clinical judgment can be subject to the same errors and biases as lay persons" (p. 394). Webster, et al. (2000) also noted the susceptibility of humans to many errors in clinical judgment, but further observed that many clinicians rarely receive any feedback on the accuracy of their predictions, which compounds the problem because any changes in their maladaptive methods are prevented.

In contrast to clinical judgment, "The superiority of mechanical prediction holds across many prediction domains," (Grove, et al., 2000, p. 24) and "is well-specified" (Webster, et al., 2000, p.19). Ideally, this type of tool provides hard actuarial data on the

probability of violence (Borum, 1996). For example, based on the cut scores, the MnSOST-R is used to classify offenders as being at low, moderate, or high risk to commit another sexual offense. In addition, Hart (1998) noted that possibly the most significant advantage of actuarial prediction is its improvement in the accuracy and consistency of predictions. Hilton and Simmons (2001) reiterate this, stating that "Research has shown that actuarial assessments of violence are consistently more accurate than unaided judgments by clinicians" (p. 393). The successful tests of reliability and validity on actuarial tools like the MnSOST-R confirm such statements.

Over the past decade, research has consistently supported the use of actuarial prediction over clinical judgment in decision-making across various domains (Harris, Rice & Quinsey, 1993; Grove & Meehl, 1996; Grove, et al., 2000), including risk assessment. Furthermore, actuarial tools have been ruled as scientifically valid in U.S. courts and have been used in U.S. penal settings for several years (Dolan & Doyle, 2000). Monohan concluded in his 1997 editorial on the subject that "Future research on the validity of predictions of violence ... is likely to stress a reliance on actuarial approaches as the best hope to improve predictive accuracy" (p. 168).

Adjusted Actuarial Assessment

Despite repeated research demonstrating the inaccuracy of clinical or subjective judgment, especially in contrast to the accuracy of actuarial tools, many clinicians remain hesitant to accept methods of actuarial prediction (Douglas, Cox, and Webster, 1999). Some practitioners argue that they deserve the freedom to make such judgments. According to Monohan (1996), clinicians may see reliance on actuarial methods as a criticism of their judgment or a threat to self-esteem. Hanson (1998) noted that clinicians may question the

comprehensiveness of actuarial tools and "Those skeptical of actuarial predictions will always find reasons to adjust actuarial estimates" (p. 65).

Other researchers have fueled the debate further with the argument that there is an overestimation of the deficiencies of clinical judgment because no distinction is made between clinical judgments based on self-reports and those based on objective reports (Westen & Weinberger, 2004). They would further posit that almost all observations involving psychology are ultimately clinical, as they involve some informal aggregation across time by someone (Westen & Weinberger, 2004). Douglas, et al. (1999) stated such an opinion explicitly, maintaining that "The use of actuarial tools also does not remove the necessity of clinical skill and judgment" (p. 164).

Researchers have also raised concerns about potential necessary reasons to adjust an actuarial assessment. Hanson (1998) argued that even the strongest proponents of actuarial prediction (Grove & Meehl, 1996; Meehl, 1954) believed that adjustments to statistical predictions could be justified in certain circumstances. Grubin (1997) furthers this argument by contending that mechanical methods may miss rare variables in individual cases that are essential to the prediction. However, Grove and Meehl (1996) refer to this objection as the 'broken leg case." In their argument, a broken leg is a clear objective fact with obvious implications, but individual cases with such variables are relatively infrequent. While such a variable may warrant overriding the actuarial findings, giving "broken leg" status to a large number of variables inconsistently applies potentially mistaken weight to such variables.

It is arguments such as these that have lead many clinicians to argue for the use of a method referred to as adjusted actuarial assessment. Quinsey, Lalumiere, Rice, and Harris (1995) explained adjusted actuarial assessment as beginning with an actuarial prediction, but

allowing expert evaluators to adjust the actuarial prediction after considering potentially important factors not included in the original actuarial measure. This method is more structured than the clinical judgment approach, but more flexible than actuarial approaches (Douglas & Kropp, 2002). However, since this approach does not place restrictions on including, weighting, or combining factors, it still fits Grove and Meehl's (1996) definition of "subjective, impressionistic" decision-making (p.293). Accordingly, some would label the adjustment of an actuarially determined risk level based on clinical judgment as a clinical override. Also, to date, there has not been a single empirical study to assess the reliability and validity of an adjusted actuarial approach to risk assessment.

Preliminary Research Indications

Despite the lack of empirical data to support the use of an adjusted actuarial approach to assessment, it appeared that such methods were already being utilized in some places. In the state of Minnesota, a number of offenders whose risk level had been determined by an actuarial measure were released at a different level after an adjustment or override was made. According to Minnesota Department of Corrections records (personal communication, William Donnay, March 11, 2004), 1,770 sex offenders, whose risk had been assessed using the MnSOST-R, were scheduled for release into Minnesota communities beginning in September, 1999. Of those offenders, 304 scored in the high risk level on the MnSOST-R. However, only 154 were classified as high risk by the ECRC. This example, and others like it, prompted Epperson and Gore (2004) to investigate the possible use of adjustment actuarial assessment with the MnSOST-R.

The purpose of the investigation by Epperson and Gore (2004) was to determine the frequency of and reasons for clinical overrides of the MnSOST-R by Minnesota DOC staff.

The results of that study demonstrated that the MnSOST-R clearly anchors the risk assessment process in the Minnesota DOC, as shown by the 75% of cases in which the risk level assigned reflects the presumptive risk level associated with the MnSOST-R score. However, despite this high rate of congruence, clinical overrides occurred in a large number (25%) of cases overall. In addition, the high percentage of cases with no override was driven by the high congruence rate (90%) in low risk cases. The percentage of clinical overrides was actually much higher in cases with a presumptive risk level in the moderate (53%) and high (49%) ranges.

Upon further analysis, more specific patterns in overrides were discovered. For upward and downward overrides, adjustments were usually one level (e.g. high to moderate, low to moderate, or moderate to either high or low), as opposed to two levels (e.g. low to high or high to low). However, downward overrides occurred twice as often as upward overrides, so the majority of one-level overrides were from high to moderate or moderate to low.

The reasons for the overrides varied widely. The authors of the MnSOST-R provided nine guidelines (see Appendix B) that were acceptable reasons for overriding the MnSOST-R score (Epperson, et al., 1998, 2000), eight of which addressed reasons to override a score in an upward direction, as they could increase the likelihood of sexual recidivism. Only one guideline addressed an acceptable reason to override the MnSOST-R score in a downward direction (see Appendix C), as it could decrease the likelihood of sexual recidivism. In addition to these guidelines, the primary author of the MnSOST-R acknowledged one additional reason, low intellectual functioning of an offender, as acceptable in making an upward override (personal communication, Douglas Epperson, May 31, 2004). He further

acknowledged that there were two additional reasons commonly used by DOC personnel when adjusting downward: when the offender had committed only an intra-familial offense or when a higher level of community notification was unnecessary due to offense circumstances. While these reasons do not affect the likelihood an offender will recidivate, they do impact the level of community notification required by an offender's release (personal communication, Douglas Epperson, May 31, 2004), which is largely based on the assigned risk level.

In the analyses of the Epperson and Gore (2004) study, all of the reasons previously established or acknowledged as acceptable by the authors of the MnSOST-R were referred to as established reasons. Although these reasons were used in many overrides (11% - 69%, depending upon the type of override), a larger percentage of cases used unestablished reasons (29% - 89%; see Appendix D for the list of unestablished reasons). Furthermore, cases in which there was a downward override were less likely to have an established reason given (11% - 35%), as opposed to cases in which there were upward overrides (52% - 69%), and, in many of the cases overridden in a downward direction (62%), the ECRC did not cite any reasons at all. Appendix E provides the details from Epperson and Gore (2004) regarding the type of reasons used and the direction of the override.

It is important to note that the DOC also had their own internally approved list of 14 special considerations that could justify an upward override. Virtually 100% of the "unestablished reasons" for upward overrides listed in Appendix D logically map onto one of the DOC's approved 14 special considerations. The "unestablished reasons" for downward overrides listed in Appendix D are not as logically connected to the four established DOC

reasons for downward overrides. In addition, a high proportion of the downward overrides cited no reasons at all.

Given that overrides of an offender's risk level by DOC personnel occur at some level in one out of every four cases, it seems important to use only reasons that have been established as "broken legs." Otherwise, clinical judgment may supersede the empirically supported actuarial measure and, as acknowledged earlier, the degree to which this may damage the instrument's predictive accuracy is unknown. Unfortunately, the consequences of any damage could be serious for individuals put at risk and society as a whole.

The Current Study

The primary limitation of the Epperson and Gore (2004) study was that the effect of the clinical overrides on predictive accuracy was not addressed. However, the research was a requisite first step into the investigation of adjusted actuarial assessment. To empirically determine the actual effects of adjusted actuarial assessment, it was necessary to evaluate the frequency of clinical overrides of an actuarial instrument and the reasons for such overrides. Epperson and Gore (2004) established that adjustments to MnSOST-R risk levels have been made in a significant percentage of cases since the full implementation of the MnSOST-R in Minnesota in 1999. Furthermore, they made clear that a large portion of these overrides were not conducted in accordance with guidelines established by the authors of the MnSOST-R.

While the purpose of the study by Epperson and Gore (2004) was to lay a foundation, the current study takes the next step by examining the impact of the reported overrides on predictive accuracy. Specifically, this research examined the overall accuracy of the purely actuarial assessments versus the adjusted actuarial assessments. It also looked at the impact adjusted actuarial assessment may have on the profile of accuracy as a whole, addressing

effects to both sensitivity and specificity. The research further investigated the accuracy of adjusted actuarial assessments if only one direction of adjustment is examined (ignoring upward or ignoring downward) and by the type of reason given for the adjustment (established or unestablished).

Overview and research questions. This study expands upon the earlier study by Epperson and Gore (2004) and provides critically needed additional information by examining the impact on accuracy of clinical adjustments made to risk levels determined by an actuarial tool, the MnSOST-R (Epperson, et al., 1998, 2000, 2003). In Minnesota, prior to their release from confinement, sex offenders are assessed twice using the MnSOST-R. This is done first by a psychologist to determine if the offender meets criteria for referral to the county attorney for possible civil commitment. For offenders who score 13 or above on the MnSOST-R, the presumptive action is referral for civil commitment review. However, any offender in the high risk category of the MnSOST-R may be referred. In some cases, the psychologist has overridden the MnSOST-R, either by referring offenders who did not score in the high risk category or by not referring offenders who scored in the very high risk subset. Then, the psychologist recommends a risk level (low, moderate, or high; see Appendix F for presumptive risk levels and associated cut scores) to be assigned based on the first MnSOST-R score. In some cases, the psychologist has overridden the MnSOST-R and recommended a different level of risk based on that override.

For offenders who are not civilly committed, the ECRC scores a second MnSOST-R and assigns the level of risk. In some cases, the ECRC has overridden the MnSOST-R and designated a different level of risk. As discussed earlier, Epperson and Gore (2004) determined the frequency of clinical overrides by psychologists and the ECRC and the

reasons for such adjustments in a sample of sexual offenders released in Minnesota since August 1999. What remained to be determined was the impact this has had on the predictive accuracy of the MnSOST-R. In this study, the primary research question asks: Is there a significant difference in the accuracy of the purely actuarial assessments versus the adjusted actuarial assessments? To provide additional information, additional research questions in this study ask: Is there a significant difference in the accuracy profile of the purely actuarial assessments versus the adjusted actuarial assessments, specifically with regard to sensitivity and specificity? Is there a significant difference in the accuracy of the adjusted actuarial assessment if only one direction of adjustment is examined (ignoring upward or ignoring downward)? And, is there a significant difference in the accuracy of the adjusted actuarial assessment by the type of reason given for the adjustment (established or unestablished)?

Benefits. There are several benefits to this research. First, to the best of the primary researcher's knowledge, no study has yet examined the effect of clinical overrides of a purely actuarial measure on predictive accuracy. Without conducting such research, it is impossible to determine whether such adjusted actuarial assessment strengthens or weakens the predictive utility of an instrument. As long as the criminal justice system continues to require predictions of dangerousness, upon which decisions such as release and confinement are based, researchers must respond with the most reliable methods available. To do this, such methods must be studied using actual case data and release outcomes.

While benefits to the science of risk assessment, such as those discussed above, encourage the pursuit of research by academics, it is the practical benefits that motivate practitioners and their organizations. The benefits of this study to the DOC include assisting the department with organizing and interpreting their own data, analyzing the effectiveness

of current departmental practices, and improving public relations. Compiling a data set of released offenders provides the DOC with an available source for department statistics about risk levels assigned and the corresponding recidivism data. This type of research also provides a means by which the DOC can either demonstrate the effectiveness of its current assessment and release procedures or show its commitment to constantly investigating and improving upon existing measures. Finally, building upon the previous idea, the DOC can express to the public its commitment to the safety and welfare of the community and rehabilitation of released offenders.

The importance of research into the reliable prediction of sexual recidivism cannot be overstated. With the continued press for more legislation such as the Adam Walsh Act, and the established use of legislated actions, like community notification and sex offender registration, it is imperative that risk assessments be as accurate as possible. The cost of mistaken assessment carries equally devastating outcomes to both offenders and communities. By mistakenly classifying offenders as high risk, an individual may unjustly be deprived of his liberty. On the other end, mistaken classification of a dangerous offender as low risk could result in the victimization of individuals and their communities in a particularly traumatic way. As noted earlier, a balance must be struck in order to effectively implement risk assessment. This requires an application of strict criteria across the domains of supervision, registration, and notification for high risk offenders that must be juxtaposed with the need to minimize damaging consequences to low risk offenders as a result of applying measures which are too harsh, inapplicable, and possibly even increase risk. Therefore, any and all efforts to improve upon predictions of recidivism should be wholeheartedly explored and supported.

METHOD

Participants

The full sample for the Epperson and Gore (2004) study was an exhaustive sample of 1,770 Minnesota sex offenders released since August 1999, for whom all necessary data could be accumulated. From that sample, they identified 449 cases in which there appeared to have been a clinical override of the MnSOST-R in the assigned risk level, either by the reviewing psychologist or the ECRC. Because more information became available regarding the original Epperson and Gore (2004) full sample, more stringent criteria were applied and cases in which data were missing were eliminated. This resulted in a total of 383 cases in which an override had occurred and those 383 cases comprise the sample for the current study. It should be noted that this sample is an exhaustive, and therefore representative, sample of cases with clinical overrides of the MnSOST-R by DOC personnel, but it is not a representative sample of all released sex offenders.

Materials and Procedures

In order to conduct this research, it was necessary to obtain data from a review of Minnesota DOC records and the state and national crime index data bases. Procedures utilized by the Minnesota DOC determined much of the required data. Since September 1999, all sex offenders in Minnesota have been assessed using the MnSOST-R prior to their scheduled release from confinement. Before the scheduled release of a sex offender in Minnesota, two sequential processes take place. The first is a review to determine whether the offender should be referred to the county attorney for possible civil commitment. The second is a review by the End of Confinement Review Committee (ECRC) to determine the risk level assigned to an offender, which influences how much community notification and

supervision the offender's release requires. In the event that the county attorney successfully pursues civil commitment, the second review by the ECRC does not take place. Decisions made at both reviews are anchored by the offender's MnSOST-R score(s). Therefore, collected data included initial MnSOST-R scores from the reviewing psychologist and the risk level recommended; ECRC MnSOST-R scores and the risk level assigned; and any rationale(s) for the clinical override(s). Recidivism data for each offender and basic demographic information for the offenders such as race, age, and level of education, were also obtained. Some of the data listed above was available through computer records, although a review of paper files was also conducted.

It is important to note that once the data was collected and matched, all identifiers were removed, rendering it anonymous. All data was collected by a trained team of researchers, including the author and a doctoral-level psychologist with graduate and undergraduate students in psychology from Iowa State University. All researchers involved in obtaining and analyzing the necessary archival data completed human subjects research training, as well as background checks required by the DOC, to further ensure the confidentiality of the records to be reviewed. This study also received full approval from both the Iowa State University Institutional Review Board (IRB; approved on January 5, 2004, IRB ID # 03-901) and the Minnesota DOC (approved January 23, 2004).

Analyses

The analyses focused on two different types of changes to predictive accuracy that are possible, changes in overall accuracy and changes in the profile of accuracy. Various statistical analyses and corresponding software were utilized.

Changes in overall accuracy. Utilizing MedCalc for Windows, version 9.2.0.0 (MedCalc Software, Mariakerke, Belgium), ROC analyses and paired comparisons were used to test for significant differences in overall accuracy. As discussed previously, the ROC curve is generated by plotting the proportion of recidivists correctly classified as high risk (sensitivity) against the proportion of non-recidivists incorrectly classified as high risk (1-specificity) for each possible cut score. For each possible cut score, offenders scoring at or above the specified cut score are classified as high risk and offenders scoring below the specified cut score were classified as low risk. The area under the ROC curve reflects the overall accuracy of the instrument. An ROC of .50 reflects chance level accuracy, whereas values significantly greater than .50 are significantly greater than chance. A value of 1.0 reflects perfect accuracy.

Changes in profile of accuracy. To test for significant differences in the profile of accuracy between purely actuarial assessment and adjusted actuarial assessment, McNemar tests (Zar, 1984) were conducted. Specifically, McNemar tests were first used to determine if the difference in the number of true positive predictions obtained by each type of assessment was significant. True positives are defined here as cases in which the individual was predicted to be at high risk to recidivate and actually did recidivate. In order to conduct the McNemar test, it was necessary to collapse the three levels of the MnSOST-R into two levels for comparison. As discussed previously, the MnSOST-R utilizes three levels of risk, high, moderate, and low. Because the McNemar Test only makes comparisons across two levels, cases falling into the moderate level on the MnSOST-R were combined with cases in the low level. This allowed for a more stringent test of true positives, as only the cases

actually falling in the high risk category constituted positive predictions. The formula for the McNemar Test for differences in true positive classifications is:

$$(F_{act\text{-}TP} - F_{adj\text{-}TP}) / \text{Square root } (F_{tool\text{-}TP} + F_{clin\text{-}TP})$$

In the above equation, $F_{act\text{-}TP}$ represents the frequency of true positive classifications for the actuarial tool and $F_{adj\text{-}TP}$ represents the frequency of true positive classifications for the adjusted actuarial method. Keep in mind that the assessment and risk assignment process begins with the actuarially determined risk level, which is either adjusted or not adjusted by the clinician. Consequently, there can be agreement on true positive classifications for the two methods, and both are given credit for the true positive. $F_{tool\text{-}TP}$ stands for the frequency of unique true positive classifications for the actuarial tool and $F_{clin\text{-}TP}$ stands for the frequency of unique true positive classifications for the adjusted actuarial method. For these last two variables, true positive classifications on the tool that were **not** adjusted by the clinician were disallowed and not counted in as a unique true positive for either method.

McNemar tests were then used to determine if the number of true negative predictions for the actuarial tool and the adjusted actuarial method were significantly different. True negatives are defined here as cases in which the individual was predicted to be a low risk to recidivate and actually did not recidivate. Again, in order to conduct the McNemar test, it was necessary to collapse the three levels of the MnSOST-R into two levels for comparison. However, now cases falling into the moderate level on the MnSOST-R were combined with cases in the high level. This allowed for a more stringent test of true negatives, as only the cases actually falling in the low risk range were counted as true negative predictions. The formula for the McNemar Test for differences in true negative classifications is:

$$(F_{act\text{-}TN} - F_{adj\text{-}TN}) / \text{Square root } (F_{tool\text{-}TN} + F_{clin\text{-}TN})$$

In the above equation, $F_{act\text{-}TN}$ represents the frequency of true negative classifications for the tool and $F_{adj\text{-}TN}$ represents the frequency of true negative classifications for the adjusted actuarial method. $F_{tool\text{-}TN}$ stands for the frequency of true negative classifications for the actuarial tool only and $F_{clin\text{-}TN}$ stands for the frequency of true negative classifications for the adjusted actuarial method only.

The McNemar test produces a *z* statistic, so the critical value for a significant difference at the .05 level of probability is 1.96. Absolute values that equal or exceed 1.96 are statistically significant ($p \leq .05$). Significant positive values indicate that the actuarial method outperformed the adjusted actuarial method. Conversely, significant negative values indicate that the adjusted actuarial method out performed the actuarial method.

RESULTS

As Epperson and Gore determined in their study (2004), it is clear that risk level assignments in the DOC are anchored by MnSOST-R scores. However, prior to conducting analyses regarding predictive accuracy, it seemed prudent to review the results obtained by Epperson and Gore (2004). After reexamining the full sample from the Epperson and Gore study in 2004 to confirm the overall frequency and direction of clinical overrides by DOC personnel, analyses of the override sample were also conducted again. The override sample taken from the original full sample in the Epperson and Gore (2004) study became the full sample for this research. Because more information was available regarding the full sample and more stringent criteria applied in eliminating cases in which data were missing, this sample is smaller than the override sample from the original study (Epperson & Gore, 2004). Therefore, the analyses were repeated to verify the patterns in the number and types of overrides made by psychologists and the ECRC, as well as in the number and types of reasons for the overrides found by Epperson and Gore (2004), with the refined override sample. Prior to reexamining their separate overrides, the agreement between the psychologists' MnSOST-R scores and the ECRC's MnSOST-R scores was again investigated.

Agreement of MnSOST-R Scores from the Reviewing Psychologists and the ECRC

As described earlier, psychologists score the MnSOST-R about six months before it goes to the ECRC. Upon receipt of the case, the ECRC scores the MnSOST-R a second time, with knowledge of the psychologist's score. Bivariate correlational analyses were conducted again to determine the level of agreement between the psychologists' and the ECRC's MnSOST-R scores. Overall, there was very strong agreement ($r = .94$, $p < .01$)

between the two scores. Although this is not a strict test of inter-rater reliability because the two scores were not completely independent, this high degree of agreement is consistent with formal studies of inter-rater reliability on the MnSOST-R, which have generally produced reliability indices in the .80 to .89 range.

Overrides by Reviewing Psychologists

Reviewing psychologists within the DOC score their MnSOST-R first, and then use it as a basis for recommending an appropriate level of risk. Table 1 shows the total number of overridden and congruent cases according to psychologist's MnSOST-R scores and psychologist's recommended risk levels. Each of the 383 cases in the sample for this study involved an override by the reviewing psychologist and/or the ECRC. Complete data were available to evaluate overrides by psychologists in 381 cases within the total override sample. Psychologists overrode the presumptive MnSOST-R risk level in 272 (71%) of these cases. Psychologists overrode cases in a downward direction more frequently than in an upward direction. Downward overrides occurred in 165 cases (61%) and upward overrides occurred in 107 cases (39%).

Table 1. Psychologist Overrides

	MnSOST-R Presumptive Risk Level		
Recommended Risk Level	Low (3 & Below)	Moderate (4-7)	High (8 and above)
Low	18 (19%)	69 (48%)	7 (5%)
Moderate	64 (68%)	42 (29%)	89 (61%)
High	11 (11%)	32 (22%)	49 (34%)
Totals	93	143	145

Two-level overrides in either direction were relatively uncommon among the psychologists. In cases with a presumptive MnSOST-R score in the high range, only seven cases (5%) were overridden to the low range. Similarly, in cases with a presumptive MnSOST-R score in the low range, only 11 cases (11%) were overridden to the high range.

The largest percentage of clinical overrides within a risk level among the reviewing psychologists was from low risk to moderate risk, with 64 cases (68%) adjusted this way. The second largest percentage of overrides was from high risk to moderate risk, with 89 cases (61%) being overridden in such a manner. Based on these percentages, the reviewing psychologists demonstrated a tendency to adjust actuarial scores towards moderate levels of risk for their recommendations, both in an upward and downward direction.

Overrides by ECRC

While reviewing psychologists make a recommendation as to which risk level they think is appropriate for an offender, it is the ECRC that actually assigns the risk level to an offender prior to his release. This assigned risk level is based on a second MnSOST-R score from the ECRC. Table 2 shows the total number of overridden and congruent cases according to ECRC MnSOST-R scores and assigned risk levels. Complete data were available to evaluate overrides by the ECRC in 379 cases within the total override sample. Out of those cases, a total of 364 (96%) had a clinical override, substantially more than the adjustments made by psychologists. In fact, clinical overrides by the ECRC occurred in almost every case. The ECRC overrode cases in a downward direction nearly twice as often as they overrode cases in an upward direction. Downward overrides occurred in 242 cases (66%) and upward overrides occurred in 122 cases (34%).

Table 2. ECRC Overrides

	MnSOST-R Presumptive Risk Level		
Assigned Risk Level	Low (3 & Below)	Moderate (4-7)	High (8 and above)
Low	4 (4%)	112 (75%)	12 (9%)
Moderate	82 (85%)	8 (5%)	118 (89%)
High	11 (11%)	29 (19%)	3 (2%)
Totals	97	149	133

Differences also existed among the ECRC overrides. Like the overrides by psychologists, two level overrides in either direction were relatively uncommon by the ECRC. In cases with a presumptive MnSOST-R score in the high range, only 12 cases (9%) were overridden to the low range. Similarly, in cases with a presumptive MnSOST-R score in the low range, only 11 cases (11%) were overridden to the high range.

The trend in ECRC clinical overrides appeared similar to that of the psychologists. The largest percentage of clinical overrides within a risk level by the ECRC was from high risk to moderate risk, with 118 cases (89%) adjusted this way. The second largest group was from low risk to moderate risk, with 82 cases (85%) being overridden in such a manner. Based on these data, the ECRC also demonstrated a tendency to adjust actuarial scores toward moderate levels. However, to accomplish this, the ECRC made slightly more downward adjustments from high risk while the psychologists made slightly more upward adjustments from low risk. The single biggest difference between psychologists and ECRC in patterns of overrides was that ECRCs made 33% more overrides than did psychologists.

Trends between Mnsost-R Scores and Overrides

As reported originally by Epperson and Gore (2004), one possibility behind the overrides was that MnSOST-R scores near cut-offs for each risk level were being overridden to the next level. For example, were scores of 4 (just above the cut point of 3 for low risk) being adjusted downward more frequently than higher scores in the moderate range, and were scores of 7 (just below the cut point of 8 for high risk) being adjusted upward more frequently than lower scores in the moderate range (see Appendix F for MnSOST-R cut points)?

There was a tendency for upward overrides to increase as scores increased towards 3, the cut point between low risk and moderate risk. Among the upward overrides by psychologists, 6.8 % had a score of 0, 9.8% had a score of 1, and 16.5% had a score of 2. Among the upward overrides for the ECRC, 6.1% had a score of 0, 13.0% had a score of 1, and 14.5% had a score of two. There was also a tendency for downward overrides to increase as scores decreased towards 8, the cut point between high risk and moderate risk. Among the downward overrides by psychologists, 5.6% had a score of 11, 8.9% had a score of 10, and 10.9% had a score of 9. Among downward overrides for the ECRC, 4.8% had a score of 11, 7.7% had a score of 10, and 10.1% had a score of 9. Downward overrides also had a similar tendency to increase in frequency as scores decreased towards 4, the cut point between moderate and low risk. However, it was very unlikely for scores greater than 12, the cut point for the very high risk subset, to be overridden at all.

Despite the indication from these trends of some adjustment resulting from scores being near cut points, they did not account for a large portion of the override cases. For the psychologists, only 12.8% of upward overrides were made at scores of 3, the cut point

between moderate and low, and 10.5% at scores of 7, the cut point between high and moderate. In looking at downward overrides by psychologists, it was determined that only 12.9% were made at a score of 4, while 12.1% were made at a score of 8. For the ECRC, only 14.5% of upward overrides were made at scores of 3 and 8.4% were made at a score of 7. In looking at downward overrides by the ECRC, it was determined that only 16.9% were made at a score of 4, while 14.1% were made at a score of 8. Looking at this another way, only 23.3% of upward overrides and 25% of downward overrides by psychologists were made at the cut point. For the ECRC, 22.9% of upward overrides and 31% of downward overrides were made at the cut point. Therefore, at most, this trend accounts for less than one third of any category of override.

Types of Reasons Used by Psychologists and the ECRC

Because trends in scoring did not sufficiently explain the number of overrides occurring, it was necessary to examine the reasons given by psychologists and the ECRC for their overrides. The authors of the MnSOST-R provided guidelines regarding acceptable reasons for potentially overriding MnSOST-R risk levels (see Appendix B and C; Epperson, et al., 1998, 2000). Eight of these guidelines addressed acceptable reasons to override a MnSOST-R risk level in an upward direction, as they may increase the likelihood of sexual recidivism. One guideline addressed an acceptable reason to override the MnSOST-R risk level in a downward direction, as it may decrease the likelihood of sexual recidivism. In addition, one other reason, low intellectual functioning of an offender, was cited by the primary author of the MnSOST-R (personal communication, Douglas Epperson, May 31, 2004) as an acceptable reason to make an upward override.

The primary author of the MnSOST-R also acknowledged two additional reasons commonly used by DOC personnel when adjusting downward (personal communication, Douglas Epperson, May 31, 2004). These reasons were when the offender had committed only an intra-familial offense or if a higher level of community notification was unnecessary due to offense circumstances. While these reasons do not affect the likelihood an offender will recidivate, they do impact the level of community notification required by an offender's release. As mentioned previously, one function of risk assessment in the DOC is to determine the level of community notification necessary, with offenders at higher risk for reoffense usually requiring broader community notification. In cases where a broader level of notification is not required due to offense circumstances, such as when an offender committed only an intra-familial offense, an offender's risk level may be adjusted downward. This reflects a somewhat imperfect relationship between risk assessment and community notification, which will be discussed later.

For the purposes of the original study by Epperson and Gore (2004), as well as for the current study, all of the reasons previously established or acknowledged as acceptable by the authors of the MnSOST-R are referred to as established reasons and the frequency of each was quantified for the purposes of analysis. It is important to note that these established reasons do not take into account additional reasons that the DOC may recognize as acceptable rationales for overriding the MnSOST-R.

In many cases, additional reasons beyond those given as guidelines by the authors of the MnSOST-R were cited by psychologists or the ECRC. In the original Epperson and Gore (2004) study, these individual reasons were grouped into appropriate categories by a Q-sort. For that study, as well as the current study, these reasons are referred to as unestablished

reasons (see Appendix D; Epperson & Gore, 2004) and the frequency of each was quantified for the purposes of analysis. Table 3 shows the breakdown of psychologists' use of established and unestablished reasons, as well as the number of cases where no reasons were given. Table 4 shows the same breakdown for reasons given by the ECRC. It is important to bear in mind that both established and unestablished reasons could be used in the same case.

The two general trends that emerged in the reasons used for both reviewing psychologists and the ECRC for this study matched those found previously by Epperson and Gore (2004). First, both groups were more likely to cite reasons overall for upward overrides than for downward overrides. Second, reviewing psychologists and the ECRC were more likely to use unestablished reasons than established reasons for both types of overrides. In addition, the ECRC was much more likely to give no reasons at all for downward overrides.

Table 3. Psychologists' Reasons for Overrides

	Category of Reason Used		
Override Direction	Established	Unestablished	No Reasons Given
Upward	74 (69%)	95 (89%)	3 (3%)
Downward	58 (35%)	122 (74%)	6 (4%)

Table 4. ECRC's Reasons for Overrides

	Category of Reason Used		
Override Direction	Established	Unestablished	No Reasons Given
Upward	63 (52%)	96 (79%)	15 (12%)
Downward	26 (11%)	69 (29%)	122 (50%)

Table 5. Congruent Reasons for Overrides by Psychologists

	Category of Reason Used	
Congruent with Override?	Established	Unestablished
Yes	36 (27%)	121 (56%)
No	96 (73%)	96 (44%)

Table 6. Congruent Reasons for Overrides by ECRC

	Category of Reason Used	
Congruent with Override?	Established	Unestablished
Yes	32 (36%)	108 (66%)
No	57 (64%)	57 (34%)

In confirming the analyses conducted by Epperson and Gore (2004) regarding the use of established and unestablished reasons for clinical overrides, a new observation was made. In cases where reasons were provided, both the reviewing psychologists and the ECRC appeared to use a "check-list" type of approach. Often, in the same case, reasons for overriding in both a downward direction and an upward direction were used, despite the fact that the override could only be in one of those directions. This occurred for established reasons and unestablished reasons. Because of this, it was impossible to determine the driving reason for the overrides. Table 5 and 6 summarize the frequency of psychologists' and ECRC's use of congruent reasons.

Impact of Overrides on Overall Predictive Accuracy

It is important to note that while the crux of the current study was to examine the impact of clinical overrides on predictive accuracy, the results of the research must be viewed as preliminary and exploratory due to the small number of recidivists in the sample. Out of the total sample of 383 offenders, only 19 (5%) recidivated during the study, with recidivism measured as a new arrest for a sex offense. This can partially be attributed to the relatively short follow-up period because 50% of the sample was less than four years post release. In addition, sexual recidivism has decreased since Minnesota implemented formal risk assessment and a tiered risk management system.

Despite the limitation of the small N for recidivists, planned tests for significant differences in overall accuracy were conducted, including ROC analyses and paired comparisons of ROC results for the actuarial and the adjusted actuarial approaches. Additionally, ROC analyses and paired comparisons of ROC results were performed to examine differences in accuracy based on the direction of override, upward or downward. Because of the small sample of recidivists, it was impossible to separately examine upward overrides and downward overrides. The question of whether or not overrides were more accurate in one direction than the other was tentatively addressed by ignoring one type of override (e.g., downward overrides), allowing the analysis to focus on the single remaining direction (e.g., upward overrides).

ROC analyses for total override sample. Using ROC curves, analyses were conducted with the total override sample to examine the predictive accuracy of adjusted actuarial assessment and pure actuarial assessment. Of the offenders in the total override sample, 19 had a new arrest for a sex offense. ROC analyses using risk levels determined by

psychologist MnSOST-R scores and risk levels recommended by psychologists were conducted. ROC analyses using risk levels determined by the ECRC Mn-SOST-R scores and risk levels actually assigned by the ECRC were also conducted. Table 7 summarizes the results of each analysis, including the Area Under the Curve (AUC) and the 95% Confidence Interval (CI) and Figures 1 and 2 show the ROC curves for each analysis.

Pairwise comparisons were then conducted between the ROC curves to determine if any of these differences were significant. The following differences in AUC were examined: MnSOST-R presumptive risk levels for psychologists with recommended risk levels by psychologists (clinical overrides) and MnSOST-R presumptive risk levels for the ECRC with assigned risk levels by the ECRC (clinical overrides) . Table 8 summarizes the results of these comparisons. In neither case was the finding significant ($p>.05$, n.s.).

Table 7. Comparison of MnSOST-R and Override Levels for the Total Sample

Risk Levels	AUC	95% CI
MnSOST-R Presumptive Risk Levels for Psychologists	0.637	0.586 to 0.685
Recommended Risk Levels (Clinical Overrides) by Psychologists	0.626	0.575 to 0.675
MnSOST-R Presumptive Risk Levels for the ECRC	0.637	0.587 to 0.686
Assigned Risk Levels (Clinical Overrides) by the ECRC	0.549	0.497 to 0.600

Figure 1. Comparison of Actuarial Assessment and Adjusted Actuarial Assessment by Psychologists Among All Overrides

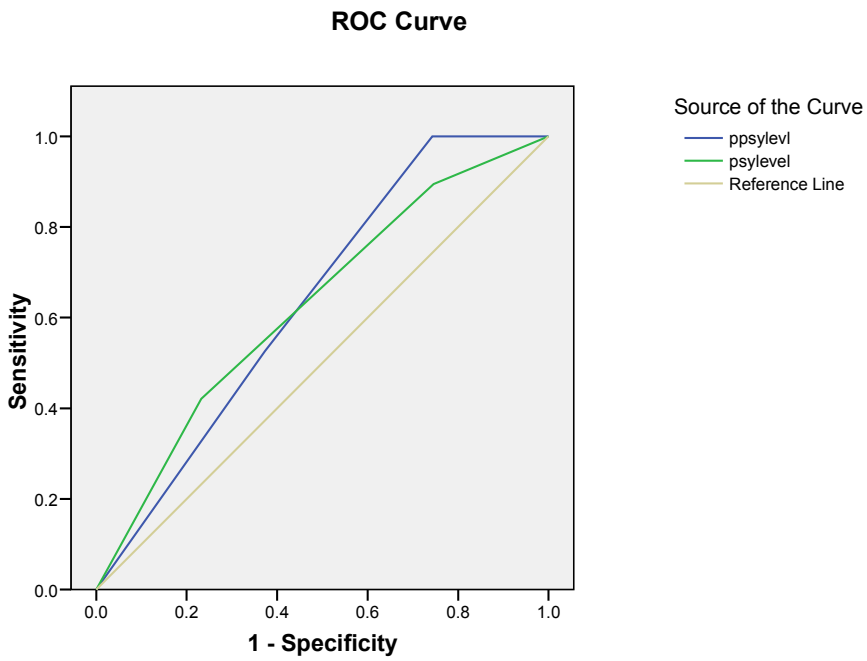


Figure 2. Comparison of Actuarial Assessment and Adjusted Actuarial Assessment by the ECRC Among All Overrides

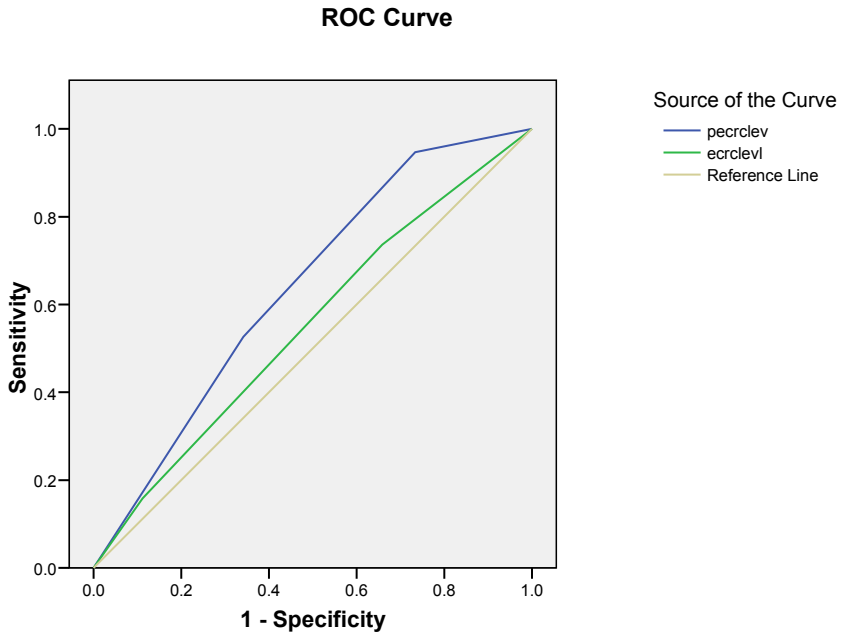


Table 8. Pairwise Comparison of ROC Curves for Total Override Sample

Psychologist Presumptive Risk Level ~ Psychologist Recommended Risk Level (clinical overrides)	
Difference Between Areas	0.011
95% Confidence Interval	-0.171 to 0.192
Significance Level	P = 0.909
ECRC Presumptive Risk Level ~ ECRC Assigned Risk Level (clinical overrides)	
Difference Between Areas	0.088
95% Confidence Interval	-0.090 to 0.267
Significance Level	P = 0.332

ROC analyses for psychologist override sample. A second set of analyses compared the predictive accuracy of adjusted actuarial assessment to pure actuarial assessment within the sample overridden by psychologists. There were 272 offenders in this subset, 12 of whom had a new arrest for a sex offense. ROC analyses were conducted for risk levels determined by psychologist Mn-SOST-R scores and risk levels actually recommended by psychologists. Additional ROC analyses were conducted, focusing on a particular direction of override, by ignoring downward overrides first, and then ignoring upward overrides. Table 9 summarizes the results of each analysis, including the Area Under the Curve (AUC)

Table 9. Comparison of Psychologist MnSOST-R and Override Levels

Risk Levels	AUC	95% CI
MnSOST-R Presumptive Risk Levels	0.649	0.589 to 0.706
Recommended Risk Levels (Clinical Overrides)	0.616	0.555 to 0.674
Recommended Risk Levels, Ignoring Downward Overrides	0.669	0.609 to 0.724
Recommended Risk Levels, Ignoring Upward Overrides	0.703	0.645 to 0.757

Figure 3. Comparison of Actuarial Assessment and Adjusted Actuarial Assessment by Psychologists Among Overrides by Psychologists

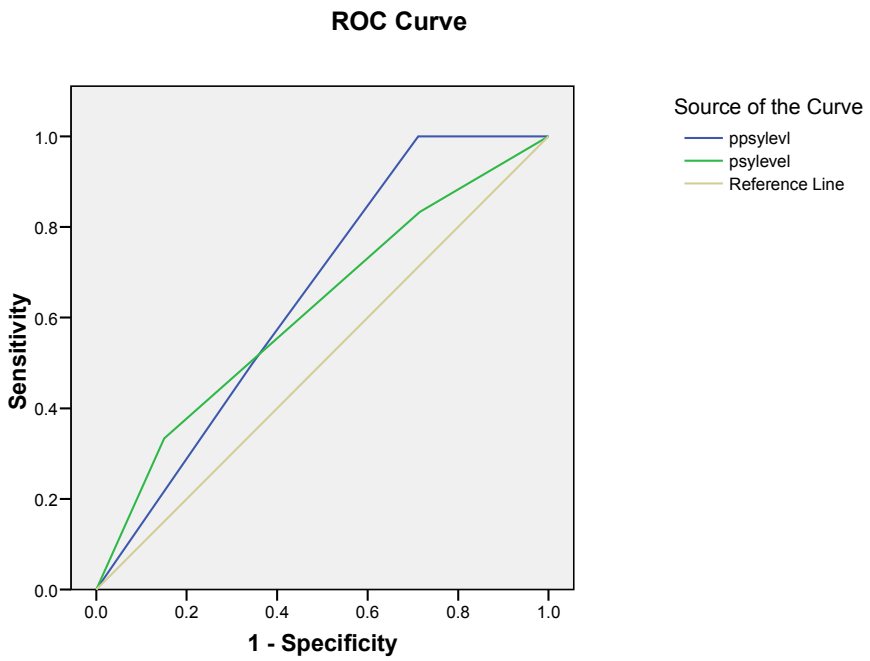


Figure 4. Comparison of Actuarial Assessment and Adjusted Actuarial Assessment, Ignoring Downward Overrides, by Psychologists Among Overrides by Psychologists

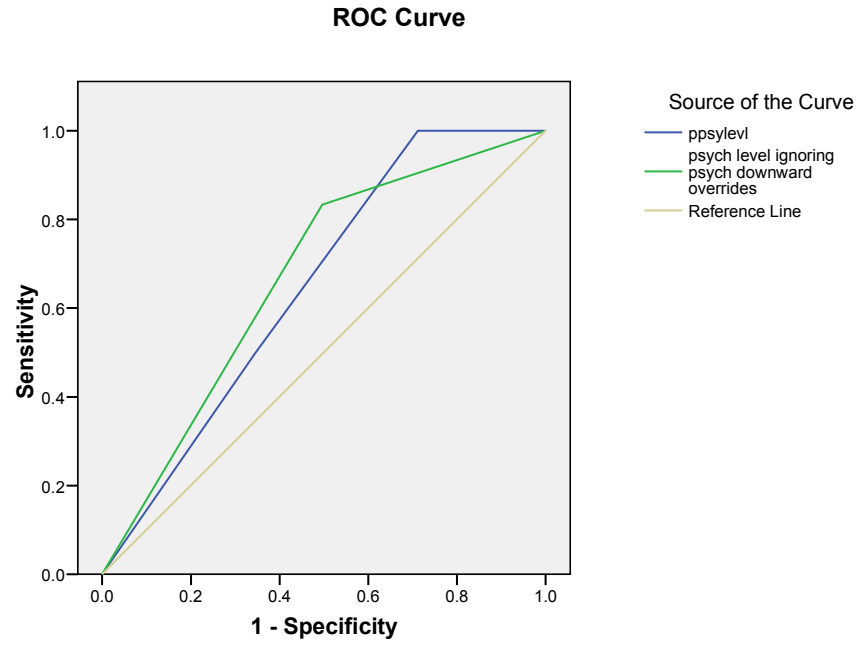
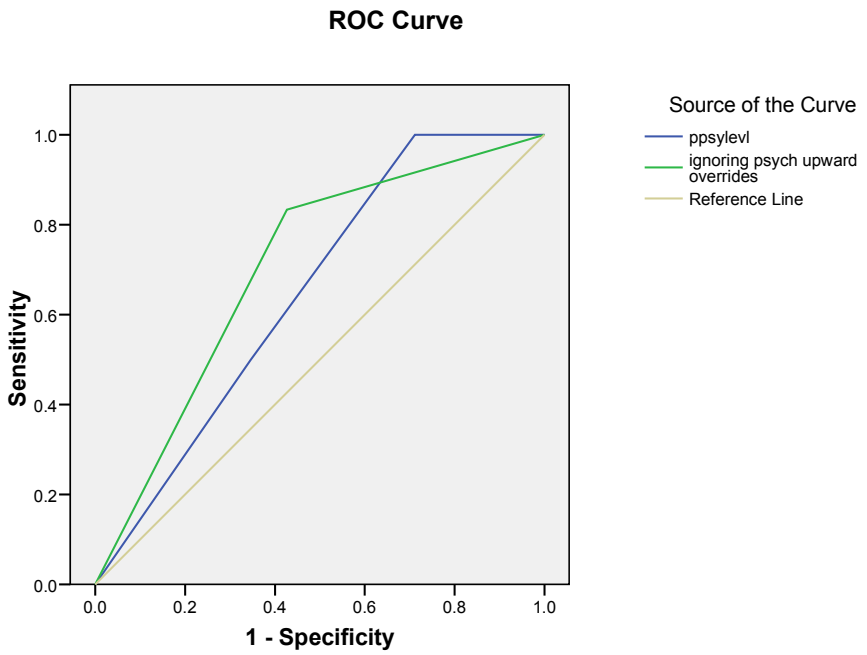


Figure 5. Comparison of Actuarial Assessment and Adjusted Actuarial Assessment, Ignoring Upward Overrides, by Psychologists Among Overrides by Psychologists



and the 95% Confidence Interval (CI) and Figures 3 through 5 display the ROC curves for each analysis.

Pairwise comparisons were then conducted between the ROC curves to determine if any of these differences were significant. The following differences in AUC were examined: MnSOST-R presumptive risk levels with overall risk levels recommended by psychologists (clinical overrides); MnSOST-R presumptive risk levels with psychologist clinical overrides ignoring downward overrides; and MnSOST-R presumptive risk levels with psychologist clinical overrides ignoring upward overrides; and psychologist clinical overrides, ignoring downward overrides, with psychologist clinical overrides, ignoring upward overrides. Table 10 summarizes the results of these comparisons. In all cases, the findings were not significant ($p>.05$, n.s.).

Table 10. Pairwise Comparison of ROC Curves for Psychologist Override Sample

MnSOST-R Presumptive Risk Level ~ Recommended Risk Level (Clinical Overrides)	
Difference Between Areas	0.033
95% Confidence Interval	-0.189 to 0.256
Significance Level	P = 0.769
MnSOST-R Presumptive Risk Level ~ Recommended Risk Level, Ignoring Downward Overrides	
Difference Between Areas	0.020
95% Confidence Interval	-0.148 to 0.187
Significance Level	P = 0.819
MnSOST-R Presumptive Risk Level ~ Recommended Risk Level, Ignoring Upward Overrides	
Difference Between Areas	0.054
95% Confidence Interval	-0.108 to 0.216
Significance Level	P = 0.512
Recommended Risk Level, Ignoring Downward Overrides	~ Recommended Risk Level, Ignoring Upward Overrides
Difference Between Areas	0.035
95% Confidence Interval	-0.032 to 0.101
Significance Level	P = 0.308

ROC analyses for ECRC override sample. A third set of analyses compared the predictive accuracy of adjusted actuarial assessment to pure actuarial assessment within the sample overridden by the ECRC. Of the offenders 364 offenders in this subset, 17 had a new arrest for a sex offense. ROC analyses were conducted for risk levels determined by the ECRC Mn-SOST-R scores and risk levels actually assigned by the ECRC. More ROC analyses were conducted, focusing on a particular direction of override, by ignoring

downward overrides first, and then ignoring upward overrides. Table 11 summarizes the results of each analysis, including the Area Under the Curve (AUC) and the 95% Confidence Interval (CI) and Figures 6 through 8 display the ROC curves for each analysis.

Table 11. Comparison of ECRC MnSOST-R and Override Levels

Risk Levels	AUC	95% CI
MnSOST-R Presumptive Risk Levels	0.634	0.582 to 0.683
Recommended Risk Levels (Clinical Overrides)	0.524	0.472 to 0.577
Recommended Risk Levels, Ignoring Downward Overrides	0.594	0.542 to 0.645
Recommended Risk Levels, Ignoring Upward Overrides	0.628	0.576 to 0.677

Figure 6. Comparison of Actuarial Assessment and Adjusted Actuarial Assessment by the ECRC Among Overrides by ECRC

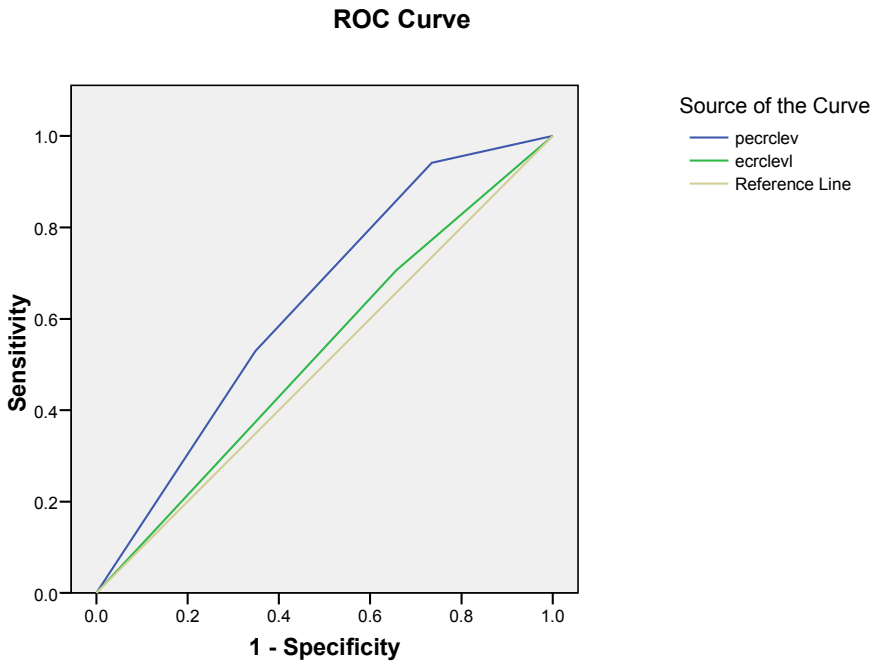


Figure 7. Comparison of Actuarial Assessment and Adjusted Actuarial Assessment, Ignoring Downward Overrides, by the ECRC Among Overrides by the ECRC

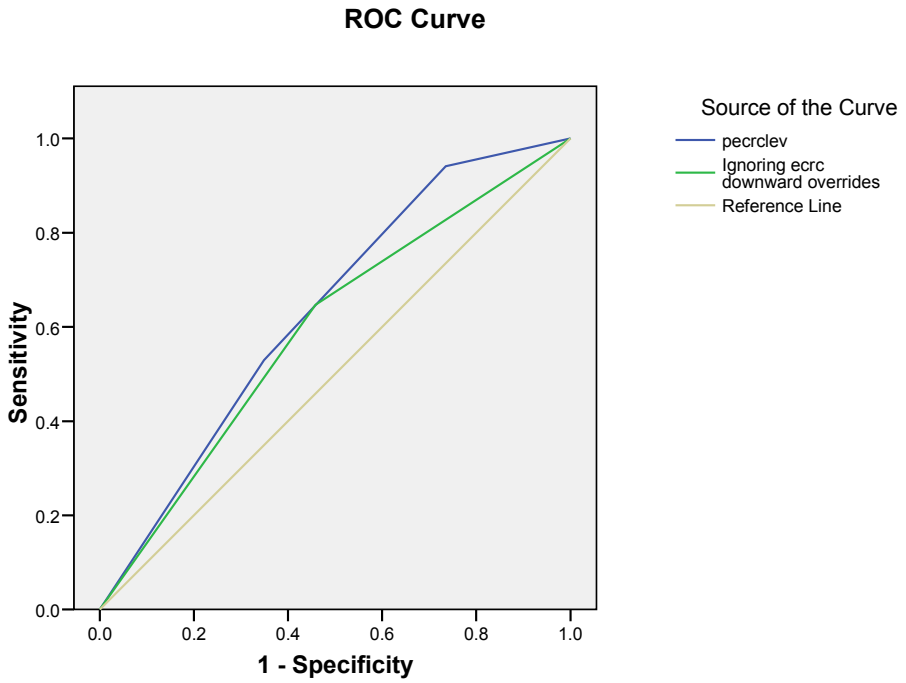
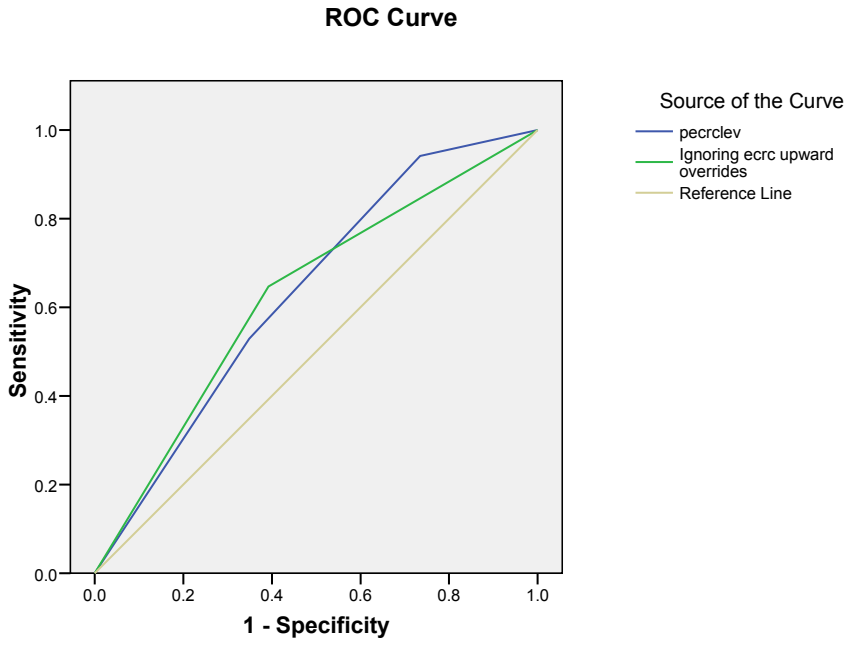


Figure 8. Comparison of Actuarial Assessment and Adjusted Actuarial Assessment, Ignoring Upward Overrides, by the ECRC Among Overrides by the ECRC



Pairwise comparisons were then conducted between the ROC curves to determine if any of these differences were significant. The following differences in AUC were examined: MnSOST-R presumptive risk levels with overall risk levels assigned by the ECRC (clinical overrides); MnSOST-R presumptive risk levels with ECRC clinical overrides ignoring downward overrides; MnSOST-R presumptive risk levels with ECRC clinical overrides ignoring upward overrides; and ECRC clinical overrides, ignoring upward overrides, with

Table 12. Pairwise Comparison of ROC Curves for ECRC Override Sample

MnSOST-R Presumptive Risk Level ~ Assigned Risk Level (Clinical Overrides)	
Difference Between Areas	0.109
95% Confidence Interval	-0.077 to 0.295
Significance Level	P = 0.249
MnSOST-R Presumptive Risk Level ~ Assigned Risk Level, Ignoring Downward Overrides	
Difference Between Areas	0.039
95% Confidence Interval	-0.066 to 0.144
Significance Level	P = 0.464
MnSOST-R Presumptive Risk Level ~ Assigned Risk Level, Ignoring Upward Overrides	
Difference Between Areas	0.006
95% Confidence Interval	-0.098 to 0.110
Significance Level	P = 0.909
Assigned Risk Level, Ignoring Downward Overrides ~ Assigned Risk Level, Ignoring Upward Overrides	
Difference Between Areas	0.033
95% Confidence Interval	-0.021 to 0.088
Significance Level	P = 0.233

ECRC clinical overrides, ignoring downward overrides. Table 12 summarizes the results of these comparisons. In all cases, the findings were not significant ($p>.05$, n.s.).

ROC analyses by reason. While the planned research included analyzing the predictive accuracy of adjusted actuarial assessment based on the use of established or unestablished reasons in addition to the above analyses, this was unable to be completed. The data collected demonstrated the use of a checklist approach by both reviewing psychologists and the ECRC in documenting the reasons used for a clinical override. In numerous cases, the reviewing psychologists and the ECRC used reasons to override a case in both upward and downward directions, despite only being able to override the case in one direction. This checklist of approach, in which all possible applicable reasons were listed, makes it impossible to ascertain what the driving reason for the override was. Without knowing which reasons to use as a basis for examination, further analyses of such reasons were impossible to attempt. Therefore, assessing the predictive accuracy of adjusted actuarial assessment by reason type could not be accomplished as part of this study.

Impact of Overrides on Types of Accuracy

While the ROC analyses were used to ascertain the overall impact on accuracy, further analyses were utilized to gauge the effect on individual aspects of the accuracy profile, specifically the findings of true positives and true negatives by each assessment method. McNemar tests were used to test for significant differences between purely actuarial assessment and adjusted actuarial assessment in the number of true positive and true negative predictions. , True positive predictions are defined here as cases in which the individual was predicted to be at high risk to recidivate and actually did recidivate. The number of true positive predictions is also the numerator in several other indices of accuracy. When this

number is divided by the total number of recidivists, the resulting proportion of recidivists classified as high risk is referred to as sensitivity; when the number of true positives is divided by the total number of positive predictions, the resulting proportion of correct positive predictions is referred to as positive predictive power (Quinsey et al., 1998).

In order to conduct the McNemar test for true positive predictions, it was necessary to collapse the three levels of the MnSOST-R (low, moderate, and high) into two levels for comparison. Because the McNemar Test only makes comparisons across two levels, cases falling into the moderate level on the MnSOST-R were combined with cases in the low level. This allowed for a more stringent test of true positives, as only the cases actually falling in the high risk category constituted positive predictions.

True negatives are defined here as cases in which the individual was predicted to be a low risk to recidivate and actually did not recidivate. The number of true negative predictions is also the numerator in several other indices of accuracy. When this number is divided by the total number of non-recidivists, the resulting proportion non-recidivists correctly classified as low risk is referred to as specificity; when the number of true negatives is divided by the total number of negative predictions, the resulting proportion of correct negative predictions is referred to as negative predictive power Again, in order to conduct the McNemar test, it was necessary to collapse the three levels of the MnSOST-R into two levels for comparison. However, now cases falling into the moderate level on the MnSOST-R were combined with cases in the high level. This allowed for a more stringent test of true negatives, as only the cases actually falling in the low risk range were counted as true negative predictions.

The McNemar test produces a *z* statistic, so the critical value for a significant difference at the .05 level of probability is 1.96. Absolute values that equal or exceed 1.96 are statistically significant ($p \leq .05$). Significant positive values indicate that the actuarial method outperformed the adjusted actuarial method. Conversely, significant negative values indicate that the adjusted actuarial method out performed the actuarial method. The formula for the McNemar tests is provided in the Method section.

Psychologist override sample. The first set of analyses examined overrides by psychologists. The McNemar test was used to compare profiles of accuracy between the psychologists' MnSOST-R and the overrides by psychologists. Table 13 summarizes these results. Results indicated there was no significant difference between the purely actuarial and the adjusted actuarial methods for true positives (McNemar = 0.6325, n.s.) or true negatives (McNemar = 0.0819, n.s.).

Table 13. Effects of Psychologist Overrides

Psychologists - True Positives

TP-Both	TP-Tool Only	TP-Clin Only	NTP-Both	TP-Actuarial	TP-Adjusted	McNemar
0	6	4	262	6	4	0.6325

Psychologists - True Negatives

TN-Both	TN-Tool Only	TN-Clin Only	NTN-Both	TN-Actuarial	TN-Adjusted	McNemar
0	75	74	123	75	74	0.0819

Note: TP-Both = True Positives for Both the Tool and the Clinical Override; TP-Tool Only = True Positives for the Tool Only; TP-Clin Only = True Positives for Clinical Overrides Only; NTP-Both = Not True Positives for Both the Tool and the Clinical Override; TP- Actuarial = True Positives Based on Actuarial Assessment ; TP-Adjusted = True Positives Based on Adjusted Actuarial Assessment; TN-Both = True Negatives for Both the Tool and the Clinical Override; TN-Tool Only = True Negatives for the Tool Only; TN-Clin Only = True Negatives for Clinical Overrides Only; NTN-Both = Not True Negatives for Both the Tool and the Clinical Override; TN- Actuarial = True Negatives Based on Actuarial Assessment; TN-Adjusted = True Negatives Based on Adjusted Actuarial Assessment

Table 14. Effects of Psychologists-Ignoring Downward Overrides

Psychologists - True Positives

TP-Both	TP-Tool Only	TP-Clin Only	NTP-Both	TP-Actuarial	TP-Adjusted	McNemar
6	0	4	262	6	10	-2.0000

Psychologists - True Negatives

TN-Both	TN-Tool Only	TN-Clin Only	NTN-Both	TN-Actuarial	TN-Adjusted	McNemar
0	75	0	197	75	0	8.6603

Note: TP-Both = True Positives for Both the Tool and the Clinical Override; TP-Tool Only = True Positives for the Tool Only; TP-Clin Only = True Positives for Clinical Overrides Only; NTP-Both = Not True Positives for Both the Tool and the Clinical Override; TP- Actuarial = True Positives Based on Actuarial Assessment ; TP-Adjusted = True Positives Based on Adjusted Actuarial Assessment; TN-Both = True Negatives for Both the Tool and the Clinical Override; TN-Tool Only = True Negatives for the Tool Only; TN-Clin Only = True Negatives for Clinical Overrides Only; NTN-Both = Not True Negatives for Both the Tool and the Clinical Override; TN- Actuarial = True Negatives Based on Actuarial Assessment; TN-Adjusted = True Negatives Based on Adjusted Actuarial Assessment

A second series of tests examined differences in true positive predictions between the actuarial risk levels and the psychologists' adjusted risk levels when only upward overrides were permitted and downward overrides were ignored. Table 14 summarizes these results. The adjusted actuarial approach, ignoring downward overrides, was significantly better than the actuarial approach in identifying true positives (McNemar = -2.000, $p<.05$); however, the purely actuarial approach was significantly better than the adjusted actuarial approach in identifying true negatives (McNemar = 8.6603, $p<.01$). Overall, the adjusted actuarial approach, ignoring downward overrides, correctly identified 4 more true positives, an increase in sensitivity from 50% to 83%. However, this was achieved at a cost of 75 fewer true negatives, a decrease in specificity from 29% to 0%.

A third series of tests examined differences between the psychologists' MnSOST-R and the overrides by psychologists, but focused on overrides in a downward direction,

Table 15. Effects of Psychologist-Ignoring Upward Overrides

Psychologists - True Positives

TP-Both	TP-Tool Only	TP-Clin Only	NTP-Both	TP-Actuarial	TP-Adjusted	McNemar
0	6	0	262	6	0	2.4495

Psychologists - True Negatives

TN-Both	TN-Tool Only	TN-Clin Only	NTN-Both	TN-Actuarial	TN-Adjusted	McNemar
75	0	74	123	75	149	-8.6023

Note: TP-Both = True Positives for Both the Tool and the Clinical Override; TP-Tool Only = True Positives for the Tool Only; TP-Clin Only = True Positives for Clinical Overrides Only; NTP-Both = Not True Positives for Both the Tool and the Clinical Override; TP- Actuarial = True Positives Based on Actuarial Assessment ; TP-Adjusted = True Positives Based on Adjusted Actuarial Assessment; TN-Both = True Negatives for Both the Tool and the Clinical Override; TN-Tool Only = True Negatives for the Tool Only; TN-Clin Only = True Negatives for Clinical Overrides Only; NTN-Both = Not True Negatives for Both the Tool and the Clinical Override; TN- Actuarial = True Negatives Based on Actuarial Assessment; TN-Adjusted = True Negatives Based on Adjusted Actuarial Assessment

ignoring upward overrides. Table 15 summarizes these results. Results indicated that the purely actuarial approach was significantly better in identifying true positives than the adjusted actuarial approach, ignoring upward overrides (McNemar = 2.4495, $p<.05$). In contrast, the adjusted actuarial approach, ignoring upward overrides, was significantly better in identifying true negatives (McNemar = -8.6023, $p<.01$). Overall, the adjusted actuarial approach, ignoring upward overrides, identified 74 more true negatives, and increase in specificity from 29% to 57%. However, the cost was not identifying any true positives, as reflected in a decline in sensitivity from 50% to 0%.

In summary, psychologists' adjusted risk levels were no more accurate than the actuarial risk levels in identifying either true positives or true negatives when all overrides were considered. When psychologists' overrides were attended to in only one direction (e.g. upward overrides) and ignored in the other direction (e.g., downward overrides), the adjusted risk levels were significantly more accurate in the direction in which overrides were

permitted (e.g., identification of true positives). However, this improvement in accuracy was achieved at a statistically significant and extreme cost: zero accuracy in the other direction (e.g., identification of true negatives).

ECRC override sample. Additional analyses examined overrides by the ECRC. Again, the McNemar test was used to compare profiles of accuracy, this time between the ECRC's MnSOST-R risk levels and the risk levels assigned by the ECRC. As indicated in Table 16, the purely actuarial approach was significantly better than the adjusted actuarial approach in identifying true positives (McNemar = 2.1106, $p<.05$.). The two approaches did not significantly differ in their ability to identify true negatives (McNemar = -1.8588, n.s.). Overall, the purely actuarial approach identified 7 more true positives than did the adjusted actuarial approach, increasing sensitivity from 12% to 53% without a significant decrease in the number of identified true negatives.

Table 16. Effects of ECRC Overrides

ECRC - True Positives

TP-Both	TP-Tool Only	TP-Clin Only	NTP-Both	TP-Actuarial	TP-Adjusted	McNemar
0	9	2	353	9	2	2.1106

ECRC - True Negatives

TN-Both	TN-Tool Only	TN-Clin Only	NTN-Both	TN-Actuarial	TN-Adjusted	McNemar
0	92	119	153	92	119	-1.8588

Note: TP-Both = True Positives for Both the Tool and the Clinical Override; TP-Tool Only = True Positives for the Tool Only; TP-Clin Only = True Positives for Clinical Overrides Only; NTP-Both = Not True Positives for Both the Tool and the Clinical Override; TP- Actuarial = True Positives Based on Actuarial Assessment ; TP-Adjusted = True Positives Based on Adjusted Actuarial Assessment; TN-Both = True Negatives for Both the Tool and the Clinical Override; TN-Tool Only = True Negatives for the Tool Only; TN-Clin Only = True Negatives for Clinical Overrides Only; NTN-Both = Not True Negatives for Both the Tool and the Clinical Override; TN- Actuarial = True Negatives Based on Actuarial Assessment; TN-Adjusted = True Negatives Based on Adjusted Actuarial Assessment

Table 17. Effects of ECRC-Ignoring Downward Overrides

ECRC - True Positives

TP-Both	TP-Tool Only	TP-Clin Only	NTP-Both	TP-Actuarial	TP-Adjusted	McNemar
9	0	2	353	9	11	-1.4142

ECRC - True Negatives

TN-Both	TN-Tool Only	TN-Clin Only	NTN-Both	TN-Actuarial	TN-Adjusted	McNemar
0	92	0	272	92	0	9.5917

Note: TP-Both = True Positives for Both the Tool and the Clinical Override; TP-Tool Only = True Positives for the Tool Only; TP-Clin Only = True Positives for Clinical Overrides Only; NTP-Both = Not True Positives for Both the Tool and the Clinical Override; TP- Actuarial = True Positives Based on Actuarial Assessment ; TP-Adjusted = True Positives Based on Adjusted Actuarial Assessment; TN-Both = True Negatives for Both the Tool and the Clinical Override; TN-Tool Only = True Negatives for the Tool Only; TN-Clin Only = True Negatives for Clinical Overrides Only; NTN-Both = Not True Negatives for Both the Tool and the Clinical Override; TN- Actuarial = True Negatives Based on Actuarial Assessment; TN-Adjusted = True Negatives Based on Adjusted Actuarial Assessment

Another series of tests examined differences between the ECRC's MnSOST-R risk levels and the overrides by the ECRC, but focused on overrides in an upward direction by ignoring downward overrides. Table 17 summarizes these results. There was not a significant difference in the identification of true positives between the purely actuarial and the adjusted actuarial approach, ignoring downward overrides (McNemar = -1.4142, n.s.). However, the purely actuarial approach was significantly better in identifying true negatives than the adjusted actuarial approach, ignoring downward overrides (McNemar = 9.5917, $p<.01$). The purely actuarial approach identified 92 more true negatives without a significant decrease in the number of identified true positives.

A final series of tests examined differences between the ECRC's MnSOST-R risk levels and the overrides by ECRC, but focused on overrides in a downward direction, ignoring upward overrides. Table 18 summarizes these results. Results indicated there were significant differences between the purely actuarial and the adjusted actuarial methods for

Table 18. Effects of ECRC-Ignoring Upward Overrides

ECRC - True Positives

TP-Both	TP-Tool Only	TP-Clin Only	NTP-Both	TP-Actuarial	TP-Adjusted	McNemar
0	9	0	355	9	0	3.0000

ECRC - True Negatives

TN-Both	TN-Tool Only	TN-Clin Only	NTN-Both	TN-Actuarial	TN-Adjusted	McNemar
76	0	84	204	76	160	-9.1652

Note: TP-Both = True Positives for Both the Tool and the Clinical Override; TP-Tool Only = True Positives for the Tool Only; TP-Clin Only = True Positives for Clinical Overrides Only; NTP-Both = Not True Positives for Both the Tool and the Clinical Override; TP- Actuarial = True Positives Based on Actuarial Assessment ; TP-Adjusted = True Positives Based on Adjusted Actuarial Assessment; TN-Both = True Negatives for Both the Tool and the Clinical Override; TN-Tool Only = True Negatives for the Tool Only; TN-Clin Only = True Negatives for Clinical Overrides Only; NTN-Both = Not True Negatives for Both the Tool and the Clinical Override; TN- Actuarial = True Negatives Based on Actuarial Assessment; TN-Adjusted = True Negatives Based on Adjusted Actuarial Assessment

both true positives (McNemar = 3.0000, $p<.01$) and true negatives (McNemar = -9.1652, $p<.01$), but in opposite directions. The purely actuarial approach identified 9 more true positives than did the adjusted actuarial approach, ignoring upward overrides, increasing sensitivity from 0% to 53%. The adjusted actuarial approach, ignoring upward overrides, identified 84 more true negatives, increasing specificity from 22% to 46%.

In summary, the accuracy of the actuarial risk levels exceeded or equaled the ECRC's adjusted risk levels in identifying both true positives and true negatives. The only exception was in the identification of true negatives when the ECRC's upward overrides were ignored. This increased accuracy was, however, obtained at a severe cost in that no true positives were correctly identified by the ECRC's adjusted risk levels, ignoring upward overrides.

DISCUSSION

The accuracy of sex offender risk assessment has significantly increased with the development and use of empirically validated actuarial risk assessment tools. One such tool, the MnSOST-R (Epperson et al., 1998, 2000, 2003), officially anchors the risk assessment process in Minnesota and a number of other states. Despite the superior accuracy of actuarial methods, many clinicians argue for an adjusted actuarial approach to risk assessment with sex offenders. An adjusted actuarial approach begins with an actuarially determined risk level, but clinicians are permitted to override this and assign a different risk level after considering potentially important factors not addressed by the actuarial tool. Indeed, even Paul Meehl, who pioneered actuarial or statistical assessment tools, acknowledged "broken leg cases" in which a clearly relevant and objective fact could override an actuarial prediction (Grove & Meehl, 1996). However, such circumstances are very infrequent and should not be used as a rationale to routinely substitute less accurate clinical judgment for more actuarial assessment. Consistent with this, the authors of the MnSOST-R identified a short list of potential "broken leg" variables that could possibly justify overriding a MnSOST-R risk level (see Appendix B and C).

Anecdotal information suggested that the MnSOST-R and other actuarial tools were being overridden fairly routinely rather than just infrequently. Epperson and Gore (2004) conducted the first systematic evaluation of the frequency and reasons for overrides of and actuarial tool (MnSOST-R). Using an exhaustive sample of 1,770 sex offenders released by the Minnesota Department of Corrections since August 1999, they found that overrides occurred in approximately 25% of cases overall (n=449), and in almost 50% of cases in which the actuarial risk level was moderate or high.

Using the override sample from the Epperson and Gore (2004) study, the preliminary goals of the current study were to (1) confirm that cases in the sample involved an override of the MnSOST-R risk level by either the reviewing psychologist or by the ECRC, and (2) if the sample changed at all as a result of Point 1 above, re-examine the patterns of reasons given for clinical overrides. The primary goal of the current study was to assess the impact of clinical adjustments to MnSOST-R risk levels on predictive accuracy.

Frequency and Patterns of Overrides

Determination of the rate of clinical overrides in the Epperson and Gore (2004) study was based on a data base kept by the Minnesota Department of Corrections (MDOC). This data base lists the MnSOST-R score and the assigned risk level for each offender. From this data base, Epperson and Gore (2004) compared the MnSOST-R risk level, based on the MnSOST-R score, to the assigned risk level to determine if an override had occurred.

For the current study, information from the case files regarding MnSOST-R scores and assigned risk levels was carefully compared to the MDOC data base and some inconsistencies emerged. In trying to understand and resolve those inconsistencies, it was discovered that the MDOC data base was a dynamic data base in which the assigned risk level was overwritten each time there was a change in risk level subsequent to officially assigned risk level by the End of Confinement Review Committee (ECRC). For example, once the offender's risk level was assigned by the ECRC, he could appeal it through an administrative law judge. If the risk level was changed by the judge, the assigned risk level in the data base was overwritten to show the new risk level rather than adding this information in separate column. Similarly, after an offender had been in the community for a sufficient period of time, he could apply for a reduction in his risk level. If this application

was successful, his assigned risk level was overwritten with his new risk level. These changes in the data base sometimes gave the appearance of an override when one had not actually occurred. Because this study focused on overrides based on the clinical judgment of the reviewing psychologists and ECRC, these "apparent overrides" in the data base were disallowed. As a result, there were 383 instances (instead of 449) in which a reviewing psychologist or ECRC overrode a MnSOST-R risk level. Thus, the overall rate of overrides in our sample was 22%, slightly lower than the originally reported rate of 25%.

Despite this slight reduction in the percentage of cases overridden, overall patterns in directions of and reasons for overrides were very similar to those reported by Epperson and Gore (2004). Reviewing psychologists and ECRCs rarely made two-level overrides from low to high or from high to low. Such overrides generally occurred in only about 10% of cases in which a two-level override was possible. Downward overrides occurred at nearly twice the rate of upward overrides for both reviewing psychologists and ECRCs.

Epperson and Gore (2004) discussed that a possible reason for the large number of downward overrides may partially be due to the imperfect relationship between risk level and community notification level. Risk level is conceptualized as the likelihood of an offender to recidivate. Community notification level refers to the need for members of a community to know about an offender's release. However, such a consideration is really a component of risk management, not a risk factor. A particular risk level is a reason for releasing an offender into a supervised setting. It is this supervision that reduces the threat presented to the community. This may decrease the need to broadly inform the community, but it does not change the risk inherent in the individual.

Even considering overrides that may reflect the sometimes murky relationship between risk assessment and risk management, there is still another potential explanation for the high level of downward overrides. Epperson and Gore (2004) noted it is also possible that decisions to adjust risk levels are unduly influenced by resource considerations. This possibility could explain why there are more downward overrides than upward overrides. Concerns about whether the DOC is responsible if a released offender recidivates would reflect a consideration of liability and would likely result in more upward overrides to reduce this possibility. However, concerns about whether there is enough money and personnel to adequately supervise offenders at a high risk level reflect a consideration of DOC resources and would likely result in more downward overrides to conserve such resources.

Although the pattern of overrides was similar for reviewing psychologists and ECRCs, one important difference was that the ECRC overrode 33% more cases than did the psychologists (364 versus 272). Consequently, if any biases affected overrides, the effect was more pervasive with the ECRC. Also, given the greater frequency of overrides by the ECRC, one might hypothesize that the ECRC overrides would have a more negative impact on predictive accuracy.

Reasons for Overrides

Patterns in the reasons given for overrides of MnSOST-R risk levels are difficult to identify and interpret because of the checklist approach that was used to list reasons. Essentially, reviewing psychologists and the ECRC listed any possible reason for a potential override in either direction without any indication of which of the specific reasons drove their

decision to override in the selected direction. There was a pattern for established reasons[1] to be cited more frequently for upward overrides than downward overrides by both reviewing psychologists and ECRC. Again, however, some of these established reasons could have been for an override in the opposite direction because only 27% of all cited established reasons supported the direction of the actual override. One surprising pattern that emerged was that the ECRCs did not provide any reasons for 50% of its downward overrides.

Because of the checklist approach to citing reasons that was used by the reviewing psychologists and the ECRC, it is impossible to interpret or attribute meaning to these patterns. Future research would benefit from greater specificity in the reports. A checklist approach could be part of the report, but the report would be more informative if it explicitly indicated which single reason or group of reasons drove the decision to override in the direction selected. This greater specificity may also help avoid situations in which no reason is cited for an override.

While more difficult to ascertain because of the aforementioned checklist approach, the pattern among reasons used to justify clinical overrides also seems to support the possibility that resource concerns may impact adjustments. Overall, fewer reasons were given for downward overrides than for upward overrides. If the availability of resources were the primary concern, then any use of those resources would need to be fully explained and supported. Upward overrides clearly require more resources and those are the overrides for which the most reasons were given. In contrast, if liability were the main concern, then

[1] It is important to note again that these were reasons established by the authors of the MnSOST-R and that the MDOC had additional reasons that they considered to be grounds for an override. Overwhelmingly, the reasons cited by psychologists and ECRC fell within the larger group of reasons approved by the MDOC.

downward overrides would need to be carefully justified and more reasons would have been used in those cases.

The pattern among the types of reasons used as a rationale for adjustments also appears to support the possible influence of resource considerations. Overall, established reasons were less likely to be used than unestablished reasons. However, this trend was most pronounced for downward overrides. This is consistent with a resource bias. Since there are far fewer established reasons for downward overrides, fewer cases meet the criteria necessary to support adjusting risk levels downward. If resource concerns motivated more downward overrides, then it was necessary to use additional reasons in order to justify those overrides. The large number of cases (50%) in which the ECRC cited no reasons for adjusting risk levels downward also seems to demonstrate less concern for offering any rationale for downward overrides versus upward overrides and is clearly not consistent with liability concerns. Instead, the absence of reasons in so many cases also appears to support the influence of resource concerns in decisions regarding overrides.

Impact of Overrides on Predictive Accuracy

The primary purpose of this study was to determine the impact of clinical overrides on predictive accuracy. The argument supporting an adjusted actuarial approach rests on its presumed greater accuracy relative to a purely actuarial approach. Because of this argument, justification for an adjusted actuarial approach would require that it result in significantly greater predictive accuracy, not just in an equivalent level of accuracy. Other factors supporting a requirement of significantly greater accuracy for an adjusted actuarial approach are its greater cost, in terms of time and personnel, and the fact that this approach at least opens the door for inappropriate overrides.

Impact of overrides on overall accuracy. ROC analyses failed to reveal any statistically significant differences in overall accuracy between the purely actuarial approach (MnSOST-R risk levels) and the adjusted actuarial approach (clinically overridden risk levels). In fact, when all overrides were considered, irrespective of direction, the accuracy of the adjusted actuarial approach of both psychologists and the ECRC failed to even nominally exceed the accuracy of the actuarial approach. Nominally, the lowest ROC was for the ECRC's, which had the greatest number of overrides.

Although there were no statistically significant differences in overall accuracy, it was at least theoretically possible that adjustments in one direction could greatly increase overall accuracy, but be offset by dramatically decreased accuracy for adjustments in the other direction. This possibility was assessed through a series of analyses that ignored overrides in one direction (e.g. downward), thereby ignoring errors in that direction, to assess the accuracy of overrides in the other direction (e.g., upward) for psychologists and the ECRC. The results of these analyses indicated that the degree of accuracy achieved in either direction was equivalent for both psychologists and ECRCs. Furthermore, none of the areas under the ROC curve associated with adjustments in one direction by either psychologists or ECRC were significantly greater that those for the actuarial approach.

Impact of overrides on types of accuracy. Although there were no differences in overall accuracy, it remained possible that adjusted actuarial and purely actuarial approaches would yield significantly different patterns of true positive and true negative predictions. When all overrides were included in the analyses, there were no significant differences in the pattern of true positive or true negative classifications for the actuarial risk levels and the psychologists' adjusted risk levels. However, the actuarial risk levels produced significantly

more true positive predictions than did the ECRC's adjusted levels. The actuarial risk levels resulted in 9 true positive predictions compared to 2 for the ECRC risk levels. This 450% increase is clearly significant and is associated in an increase in sensitivity from 12% to 53%. Substantial increases in sensitivity are often associated with substantial decreases in specificity, but this increase was achieved with only a relatively small and statistically insignificant decrease of 8% in specificity.

A different pattern emerged when only one direction of overrides was assessed by ignoring overrides in the opposite direction. For psychologists, this resulted in the adjusted risk levels significantly outperforming the actuarial risk levels in the direction of the assessed overrides, with the opposite pattern occurring in the other direction. So, when upward overrides were assessed and downward overrides were ignored, psychologists' adjusted risk levels resulted in significantly more true positive predictions and significantly fewer true negative predictions. Specifically, the number of true positive predictions increased from 6 to 10, with an associated increase in sensitivity from 50% to 83%. However, this gain in true positive predictions was associated with a dramatic decline in the number of true negative classifications from 75 to 0, which represents a decline in specificity from 27% to 0%. In other words, <u>all</u> non-recidivists were false positive predictions.

Similarly, when downward overrides by psychologists were assessed by ignoring their upward overrides, psychologists' adjusted risk levels resulted in significantly more true negative predictions and significantly fewer true positive predictions. The number of true negative predictions nearly doubled from 75 to 149, an increase in specificity from 29% to 57%. Again, the cost of this increased specificity was dramatic, with sensitivity dropping from 50% to 0%; <u>all</u> recidivists were false negative predictions.

The pattern for the ECRC was similar to that for psychologists when attending to downward overrides and ignoring upward overrides. The adjusted risk levels again significantly increased the number of true negative predictions by a factor of two, but true positive classifications dropped significantly from 9 to 0. Unlike psychologists, the ECRC derived no significant increase in true positive predictions when upward overrides were attended to and downward overrides were ignored, but they shared the same fate of 0% specificity.

Conclusions and Implications

The results of this study confirm that the MnSOST-R anchors the risk assessment process in the Minnesota Department of Corrections, given that the risk levels recommended by psychologists and assigned by ECRCs are the same as the presumptive risk levels based on MnSOST-R scores 78% of the time. It is also clear that the overrides that occur in the other 22% of cases are the result of some structured process in that the vast majority of reasons cited for overrides are consistent with MDOC guidelines even though they did not all conform with reasons established by the MnSOST-R authors. Conceptually, it is exactly this kind of structured process that could lead to greater accuracy for an adjusted actuarial approach relative to a purely actuarial approach. Given the small number of recidivists in the sample, however, our results must be considered as preliminary. Based on those preliminary results, some tentative conclusions can be made.

The primary goal of this study was to assess the impact of an adjusted actuarial approach on the accuracy of an actuarial tool, specifically the MnSOST-R. Results demonstrated that adjusted actuarial assessment did not differ significantly from the purely actuarial approach. Looking at all overrides, regardless of direction, the adjusted actuarial

approach failed to even nominally exceed the MnSOST-R in terms of overall predictive accuracy. It also did not yield any significant difference in the pattern of true positive and true negative classifications.

Again, in order to justify the use of an adjusted actuarial approach, it must be shown that such an approach would significantly increase the predictive accuracy of the original actuarial measure. Otherwise, the increased cost of using adjusted actuarial assessment cannot be supported, as it would make no sense to utilize more time and personnel without an improvement in results. Also, as noted, the use of adjustments raises the possibility of the application of inappropriate overrides, such as those that do not conform either to the reasons established by the authors of the MnSOST-R or those consistent with MDOC guidelines.

In this study, the standard of greater accuracy was clearly not met, either when examining overrides overall or attending to only one direction of override. It is possible that the lack of improvement in accuracy may be related, at least partially, to the large number of overrides that occurred (22%). This would be especially evident in overrides by the ECRC, given that they overrode 33% more cases than the reviewing psychologists. It is possible that a smaller number of overrides, such as those referred to by Grove and Meehl's (1996) discussion of "broken legs," could result in greater accuracy.

There is also the possibility that a pattern exists among the reasons for overrides, which would have an effect of the accuracy of the adjusted actuarial approach. Unfortunately, any such pattern is unable to be identified or interpreted at this point due to the checklist approach utilized by both the reviewing psychologists and the ECRC to justify their overrides, which limits understanding the rationale behind the override. If greater details were provided by the psychologists and the ECRC regarding which reason or reasons

informed their decision to override, analyses of any patterns that emerged could be attempted and possible relationships to accuracy could be explored.

Limitations and Future Research

The primary limitation of this study is clearly the small number of recidivists in the sample. As noted earlier, out of the total sample of 383 offenders, only 19 (5%) recidivated, with recidivism measured as a new arrest for a sex offense. This is a substantially smaller percentage of recidivists than found in previous studies and is most likely due to the relatively short time frame examined. Rates of recidivism predictably increase with longer follow-up periods (Hanson, 1998). For example, Hanson and Bussiere (1998) reported a 13.4% rate of recidivism in a four to five year follow-up period and Hanson (1998) discussed rates of recidivism increasing from 8% at two years to 15% at 5 years to 22% at 10 years. However, in this study, 50% of offenders in the sample were less than four years post release and the maximum amount of time any offender had been released was six years. In addition, sexual recidivism has decreased over the past decade since Minnesota implemented formal risk assessment and a tiered risk management system.

Another limitation of this study was the use of a rearrest for a sex offense as the predictor variable. While a rearrest for a sex offense is easily measured and accounts for the probability that offense-like behavior occurred even without a new conviction, it cannot take into account incidents of reoffense by offenders which did not result in arrest. Therefore, it is possible that a greater number of recidivists actually existed among the sample, but had not yet been caught.

A final limitation of this study was the use of archival records. Because data in this study were obtained solely through file review, information was limited to the written record

the process investigated. Therefore, it was not possible to fully examine the entire decision-making process behind the clinical overrides. In addition, it was also not possible to investigate the full details of the reoffense behavior resulting in rearrest.

The purpose of this study was to examine the impact adjustments to the MnSOST-R had on its predictive accuracy to establish whether adjusted actuarial assessment enhances or impedes the predictive accuracy of a specific actuarial tool. It was a follow-up to previous research that had established overrides of the MnSOST-R were occurring with some frequency within the Minnesota DOC. While the results of this study demonstrated that adjusted actuarial assessment does not enhance predictive accuracy overall, it will be necessary to conduct a follow-up study in several years. At that time, it is likely the number of recidivists within the overall sample will have increased and this larger sample size should yield sufficient power to confirm or disconfirm these results.

The overarching criterion for implementing clinical overrides should be that their use increases predictive accuracy. It is not simply enough for adjusted actuarial assessment to perform as well the actuarial measure because it introduces several problematic elements. Adjusted actuarial assessment requires more resources and introduces the possibility of clinician error. Furthermore, the current lack of clearly delineated reasons for the majority of override decisions impedes the use of adjusted actuarial assessment in any measurable fashion, which also raises the possibility of numerous problems. A longitudinal follow-up study incorporating more recidivism data will be useful in determining if the preliminary findings from this research remain valid or whether new information may demonstrate increased accuracy for adjusted actuarial assessment. However, if the results of this study hold up, the argument would remain that only purely actuarial approaches should be utilized

given their accuracy, objectivity, and effectiveness in maximizing the use of resources, such as time and personnel. Even now, since it cannot be clearly demonstrated that adjusted actuarial approaches to assessment improve upon the use of the original empirically-validated actuarial tools, they should only be implemented in rare instances. The pure actuarial methods should be relied in the vast majority of cases or the accuracy of risk assessment may be put at risk.

REFERENCES

Anderson, C. (March 15, 2007). Jury: Jessica's killer should die [Electronic version]. *Associated Press*.

Andrews, D.A & Bonta, J. (1998). Prediction of criminal behavior and classification of offenders. In D.A. Andrews and J. Bonta (Eds.), *The psychology of criminal conduct*, pp. 211-247. Cincinnati, OH: Anderson Publishing Company.

Associated Press. (January 9, 2007). Idaho grand jury expected to reach decision soon on Duncan [Electronic version].

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, Special Issue: Psychopathy and risk assessment, 28*, 490-521.

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, Special Issue: Sex Offender Treatment, 47*, 422-438.

Becker, J. V. & Murphy, W.D. (1998). What we know and do not know about assessing and treating sex offenders. *Psychology, Public Policy, and Law, 4*, 116-137.

Borum, R. (1996). Improving the clinical practice of violence risk assessment. *American Psychologist, 51*, 945-953.

Dolan, M., & Doyle, M. (2000). Violence risk prediction: Clinical and actuarial measures and the role of the PCL-R. *British Journal of Psychiatry, 177*, 303-311.

Douglas, K.S. & Kropp, P.K. (2002). A prevention-based paradigm for violence risk assessment: Clinical and research applications. *Criminal Justice and Behavior, 29*, 617-657.

Douglas, K.S., Cox, D.N. & Webster, C.D. (1999). Violence risk assessment: Science and practice. *Legal and Criminological Psychology, 4*, 149-184.

Epperson, D.L. & Gore, K.S. (2004, October). *Investigation of clinical overrides of the MnSOST-R and the resulting impact on predictive accuracy*. Paper presented at the 23rd research and treatment convention of the Association for the Treatment of Sexual Abusers, Albuquerque, NM.

Epperson, D.L., Kaul, J.D., & Hesselton, D. (1998, October). *Final report on the development of the Minnesota Sex Offender Screening Tool – Revised (MnSOST-R)*. Paper presented at the 15th annual research and treatment convention of the Association for the Treatment of Sexual Abusers: Vancouver, BC.

Epperson, D.L., Kaul, J.D., Huot, S.J., Alexander, W., & Goldman, R. (2000, November). *Cross-validation of the Minnesota Sex Offender Screening Tool – Revised (MnSOST-R)*. Paper presented at the 17th annual research and treatment convention of the Association for the Treatment of Sexual Abusers: San Diego, CA

Epperson, D.L., Kaul, J.D., Huot, S.J., Goldman, R., & Alexander, W. (2003). *Minnesota Sex Offender Screening Tool – Revised (MnSOST-R): Development, validation, and recommended risk-level cut scores*. Technical paper, St. Paul, MN: Minnesota Department of Corrections.

Epperson, D.L., Kaul, J.D., Huot, S.J., Hesselton, D., Alexander, W., & Goldman, R. (1998, 2000). *Minnesota Sex Offender Screening Tool – Revised (MnSOST-R)*. St. Paul, MN: Minnesota Department of Corrections.

Grisso, T. & Appelbaum, P.S. (1992). Structuring the debate about ethical predictions of future violence. *Law and Human Behavior, 17*, 482-485.

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.

Haga, C. (April 19, 2004). April 18, 2004: Dru Sjodin's body found [Electronic version]. *Minneapolis Star-Tribune*.

Haga, C., McKinney, M. & Lopez, P. (December 13, 2003). Psychologist raised Rodriguez warning [Electronic version]. *Minneapolis Star-Tribune*.

Hanson, R.K. (1998). What do we know about sex offender risk assessment? *Psychology, Public Policy & Law, 4*, 50-72.

Hanson, R.K, & Bussiere, M.T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66*, 348-362.

Hanson, R.K., & Morton-Bourgon, K. (2004). *Predictors of sexual recidivism: An updated meta-analysis.* (User Report 2004-02). Ottawa: Public Safety and Emergency Preparedness Canada.

Harris, G.T., Rice, M.E., & Quinsey, V.L. (1993). Violent recidivism of mentally disordered offenders. *Criminal Justice and Behavior, 20*, 315-335.

Hart, S.D. (1998). The role of psychopathy in assessing risk for violence: Conceptual and methodological issues. *Legal and Criminological Psychology, 3*, 121-137.

Hilton, N.Z. & Simmons, J.L. (2001). The influence of actuarial risk assessment in clinical judgments and tribunal decisions about mentally disordered offenders in maximum security. *Law and Human Behavior, 25*, 393-408.

H. Res. 4472, 109th Congress: Adam Walsh Child Protection and Safety Act of 2006. (2006) (enacted).

Janus, E.S., & Meehl, P.E. (1997). Assessing the legal standard for predictions of dangerousness in sex offender commitment proceedings. *Psychology, Public Policy & Law, 3*, 33-64.

Knight, R. A., & Thornton, D. (2007). Evaluating and improving risk assessment schemes for sexual recidivism: A long-term follow-up of convicted sexual offenders. Retrieved http://www.ncjrs.gov/pdffiles/nij/grants/217618.pdf.

Langton, C.M., Barbaree, H.E., Harkins, L., Seto, M.C., & Peacock, E.J. (2002). *Evaluating the predictive validity of seven risk assessment instruments for sex offenders*. Paper presented at the 21st annual research and treatment convention of the Association for the Treatment of Sexual Abusers: Montreal, QC.

Langton, C.M., Barbaree, H.E., Seto, M.C., Peacock, E.J., Harkins, L., & Hansen, K.T. (2007). Actuarial assessment of risk for reoffense among adult sex offenders: Evaluating the predictive accuracy of the Static-2002 and five other instruments. *Criminal Justice and Behavior, 34*, 37-59.

McWilliams, M. (February 1, 2006). Bentley guilty [electronic version]. *Iowa City Press-Citizen*.

MedCalc Software. (2007). MedCalc for Windows (Version 9.2.0.0) [computer software]. Mariakerke, Belgium.

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

Monohan, J. (1981). Predicting *violent behavior: An assessment of clinical techniques*. Beverly Hills, CA: Sage.

Monohan, J. (1996). Violence prediction: The past twenty and the next twenty years. *Criminal Justice and Behavior, 23*, 107-119.

Monohan, J. (1997). Actuarial support for the clinical assessment of violence risk. *International Review of Psychiatry, 9*, 167-169.

Monahan, J. & Steadman, H.J. (1994). Towards a rejuvenation of risk assessment research. In J. Monohan & H.J. Steadman (Eds.), *Violence and mental disorder: Developments in risk assessment*, pp. 1-17. Chicago, IL: University of Chicago Press.

Quinsey, V.L., Lalumiere, M.L., Rice, M.E., & Harris, G.T. (1995). Predicting sexual offenses. In J.C. Campbell (Ed.) *Assessing dangerousness: Violence by sexual offenders, batterers, and child abusers* (pp. 1114-1137). Thousand Oaks, CA: Sage.

Rice, M.E., Harris, G.T., & Quinsey, V.L. (1990). A follow-up of rapists assessed in a maximum-security psychiatric facility. *Journal of Interpersonal Violence, 5*, 435-448.

U.S. Federal News Service. (May 4, 2006). Senate passes 'Jetseta's Bill' included in child sex offender legislation [Electronic version].

Vansickle, A. (January 2, 2006). Losing Jessica: A crime awful beyond belief [Electronic version]. *St Petersburg Times*.

Webster, C.D., Douglas, K.S., Belfrage, H., & Link, B. (2000). Capturing change: An approach to managing violence and improving mental health. In S. Hodgins and R. Muller-Isberner (Eds.), *Violence among the mentally ill: Effective treatment and management strategies* (pp. 119-144). Dordrecht, the Netherlands: Kluwer/Academic.

Westen, D. & Weinberger, J. (2004). When clinical description becomes statistical prediction. *American Psychologist*, 7, 595-613.

Zar, J. H. (1984). Biostatistical Analyses (2nd Ed.), pp. 156-157. Englewood Cliffs, NJ: Prentice Hall.

APPENDIX A

MnSOST-R Score Recording Sheet Items (Epperson, et al., 1998, 2000)

Historical/Static Variables

1. Number of sex/sex-related convictions (including current conviction):
   One……...0
   Two or more……+2

2. Length of sexual offending history:
   Less than one year…….-1
   One to six years……+3
   More than six years…….0

3. Was the offender under any form of supervision when they committed any sex offense for which they were eventually charged or convicted?
   No…….0
   Yes……..+2

4. Was any sex offense in a public place?
   No…….0
   Yes……..+2

5. Was force or the threat of force ever used to achieve compliance in any sex offense?
   No force in any offense……-3
   Force present in at least one……..0

6. Has any sex offense involved multiple acts on a single victim within any single contact event?
   No……..-1
   Yes……....+1

7. Number of different age groups victimized across all sex/sex-related offenses (charged or convicted):
   - Age 6 or younger
   - Age 7 to 12 years
   - Age 13 to 15 years and the offender is more than five years older than the victim
   - Age 16 or older

   No age group/only one age group...0
   Two or more age groups……...+3

8. Offended against a 13- to 15-year-old victim and the offender was more than five years older than the victim:
   No…….0
   Yes……..+2

9. Was the victim a stranger in any sex/ sex-related offense (charged or convicted)?
   No victims were strangers……..-1
   At least one was a stranger…….+3
   Uncertain/missing information…….0

10. Is there any evidence of adolescent antisocial behavior in the file?
    No indication...……-1
    Some relatively isolated acts……....0
    Persistent, repetitive pattern……...+2

11. Pattern of substantial drug or alcohol abuse (12 months prior to arrest for instant offense or revocation)?
    No……....-1
    Yes……..+1

12. Employment history (12 months prior to arrest for instant offense):
    Stable employment >1 year...……..-2
    Homemaker, retired, full-time student or disabled…….....-2
    Part-time, seasonal, unstable…….0
    Unemployed…..……..+1
    File contains no information…….0

Historical/Static Subtotal:

Institutional/Dynamic Variables

13. Discipline history while incarcerated (does not include treatment failure):
    No major reports or infractions……..0
    One or more major reports……...+1

14. Chemical dependency treatment while incarcerated:
    No treatment recommended……0
    Treatment recommended and successfully completed……..-2
    Treatment recommended, but offender refused or quit…….+1
    Treatment recommended, but terminated by staff…….....+4

15. Sex offender treatment history while incarcerated:
    No treatment recommended……0
    Treatment recommended and successfully completed……..-1
    Treatment recommended, but offender refused or quit……..0
    Treatment recommended, but terminated by staff…….....+3

16. Age of offender at time of release:
    Age 30 or younger…….+1
    Age 31 or older……...-1

Institution/Dynamic Subtotal

Total Score (static + dynamic)

APPENDIX B

MnSOST-R Special Considerations (Epperson, et al., 1998, 2000)

Special Concerns That <u>May</u> Increase Likelihood of Sexual Recidivism

1. The offender has made statements, as documented in file, indicating an intent to reoffend.

2. The offender's release conditions have been restructured or revoked for failure to follow treatment directives, inappropriate sexual behavior, or behavior reflective of offense dynamics.

3. Ten or more major discipline reports during the current incarceration modestly increases the likelihood of sexual recidivism (approximately equivalent to a one-point increase in MnSOST-R score).

4. Five or more separate sentencing occasions for all sex/sex-related offenses (juvenile sex offenses, gross misdemeanor or felony sex offenses) ***<u>plus</u>*** all adult non-sexual felony offense may modestly increase likelihood of sexual reoffending (Document actual number, including current sentencing occasion).

   > Sentencing occasions are *distinct* court appearances when the offender was sentenced for criminal convictions. The number of charges, counts, or convictions is irrelevant because only distinct sentencing occasions are counted.

5. Credible evidence in the file indicates that the offender has victimized an extraordinarily high number of people (50 or more) in hands-on sex offenses (charged or uncharged).

   > For example, an offender reports over 70 victims of hands-on sex offenses, and this self-report is documented in the file and deemed credible.

6. The offender refused, quit, or did not pursue sex offender or chemical dependency treatment during the current incarceration (items #14 and #15 on the MnSOST-R) ***<u>and</u>*** has a pattern of <u>repeated</u> prior treatment failures (being terminated from treatment, quitting treatment, and/or sexual re-offending after completing treatment).

7. The offender has demonstrated a pattern of inappropriate sexual behavior while incarcerated, as officially documented in the file.

   > Examples of inappropriate sexual behavior would include discipline reports for masturbation, exposing, sexual activity, or sexual assault; collection of inappropriate sexual material; grooming of targeted victims; and other sexual activity that violates institutional rules.

8. The current or past sex/sex-related offenses involved a significant degree of harm in the form of physical injury or death to the victim(s).

   > This item reflects the ***<u>seriousness</u>*** of the risk involved rather than the ***<u>likelihood</u>*** of risk. Therefore, this item does not increase the likelihood of reoffending, but it may increase their risk level because of the increased potential of serious physical harm to victims.

APPENDIX C

MnSOST-R Special Considerations (Epperson, et al., 1998, 2000)

Special Concerns That <u>May</u> Decrease Likelihood of Sexual Recidivism

1. The offender has developed an incapacitating illness or physical condition that decreases motivation or ability to sexually offend (e.g. later stages of a terminal illness, etc.)

APPENDIX D

Unestablished Reasons Used to Override MnSOST-R (Epperson & Gore, 2004)

Unestablished Reasons Used for Upward Overrides by Psychologists and ECRC

1. History of supervision/release failures (did not rise to the level of release revocation or restructure)
2. Deviant sexual orientation
3. Nature of victim pool requires broader notification
4. Evidence of extensive number of victims for which offender was not charged (less than 50)
5. Predatory offense behavior and/or high-risk grooming behaviors
6. Long history of previous sex offenses, civil commitment referral, severe violence, or extensive criminal history
7. Unlikely to have stable, well-supervised release conditions
8. Mental health and/or addiction concerns
9. Denial of index offense and/or lack of offender insight
10. Egregious circumstances during offense (e.g., gratuitous violence, injury to children)

Unestablished Reasons Used for Downward Overrides by Psychologists and ECRC

1. Currently in or successful completion of sex offender or chemical dependency treatment
2. Only one known sex offense
3. Concerns about MnSOST-R scoring or risk level assigned (i.e., questions raised about scoring or offender does not appear to rise to associated level of risk)
4. Stable and/or well-supervised release conditions
5. Sexual contact for index offense did not involve force, was non-predatory, involved an acquaintance, and/or was consensual/mutual
6. Subject's age at time of release or time elapsed since index offense
7. Indications of offender insight or rehabilitation potential

APPENDIX E

Overview of Types of Reasons Used in Clinical Overrides (Epperson and Gore, 2004)

Psychologists' Reasons for Overrides

	Category of Reason Used		
Override Direction	Established	Unestablished	None
Upward	74 (69%)	95 (89%)	4 (4%)
Downward	58 (35%)	122 (74%)	15 (9%)

ECRC's Reasons for Overrides

	Category of Reason Used		
Override Direction	Established	Unestablished	None
Upward	62 (52%)	95 (80%)	13 (11%)
Downward	26 (11%)	70 (29%)	149 (62%)

APPENDIX F

Presumptive Risk Levels and Associated MnSOST-R Cut Scores
(Epperson, et al., 2003, p. 24)

Presumptive Risk Level	**MnSOST-R Score**
1 (low)	3 and below
2 (moderate)	4 to 7
3 (high)	8 and above
Refer to county attorney[a]	13 and above

Note. [a] The referral group is a subset of the high risk group.

ACKNOWLEDGEMENTS

This dissertation was possible only because of the generous support and cooperation of the Minnesota Department of Corrections, which is committed to improving their policies and procedures through empirical research. I am particularly grateful for the assistance provided by DOC staff members Bill Donnay, Dwight Close, and Grant Duwe. I also deeply appreciate the high quality work of the undergraduate research assistants on this project.

I would also like to express my sincere gratitude to all who provided assistance, in the form of expertise, encouragement, empathy, or all of the above. Specifically, I wish to thank and acknowledge the following people for the crucial contributions they made: Douglas L. Epperson, Ph.D., my major professor, for all the guidance, affirmation, and shared meals throughout graduate school and for frequently going above and beyond as a mentor and friend; Jonathan S. Gore, Ph.D., my husband, for his unending supply of patience and his continued willingness to take on extra household duties and to answer any statistics-related query; Michael S. Spencer, Ph.D., my father, for setting a fine example of scholarship and perseverance and for his willingness to commiserate with me throughout my own academic pursuits; Susan C. Spencer, my mother, for being the best role model to whom any woman could aspire and for her willingness to always listen and provide chocolate; all the colleagues who provided assistance, expertise and feedback; all the family and friends whose love and support has helped shape everything good in me, especially Cynthia and Jeff Gillette, Kelsey Spencer, Andrew Spencer, Austin Spencer, Leonard Phillips, Herb Johnson, Jeff and Amy Gore, David Basham, Matthew Stovall, Lyle Stefanich, Lori Trull, and Carolyn Moore; and all the clients with whom I have worked, who remind me, despite the overwhelming investment of time and energy, of the privilege it is to become a psychologist. Thank you.