

Strategies for improving offender risk prediction: An examination of methods

Ryan M. Labrecque, Ph.D.
Portland State University

Paula Smith, Ph.D.
University of Cincinnati

ABSTRACT

There is a considerable debate in the field with respect to what is the best strategy for improving the predictive accuracy of offender risk assessments. Among these differences of opinion are opposing viewpoints regarding static and dynamic risk factors, the mathematics of computing scores (e.g., adding/removing items, weighting items), and the use of offender responsibility factors (e.g., gender). There are also techniques used in other behavioral science fields (e.g., subgroup norming) that have yet to be explored on offender populations. This study examines the ability of three approaches (criminal history, selected items, subgroup weighted) in improving the predictive accuracy of the Level of Service Inventory- Revised (LSI-R) on the same sample of offenders. This study did not find that any of these methods were able to produce a statistically significant improvement over the standard LSI-R assessment score alone. Implications of these findings and recommendations for future research are discussed.

Keywords: Risk assessment, Level of Service Inventory-Revised, LSI-R, community corrections, correctional policy, subgroup norming

INTRODUCTION

Offender risk and need assessments have often been described as the cornerstone of evidence-based practices, which are designed to reduce the risk of recidivism and enhance public safety (Lowenkamp, Latessa, & Holsinger, 2006). From this perspective, predicting whether an offender will (or will not) engage in crime is an important component of decision-making in the criminal justice system. For example, criminal justice officials often have to make difficult choices about whether or not an offender should be sentenced, released, revoked, sent to treatment, and how intensively they should be supervised. It has been argued that risk assessments can help guide these decisions by providing officials with reliable information in a systematic and objective manner (Bonta, 2002). However, there is also a need for caution whenever risk instruments are used to influence decisions related to individual liberty and public safety (Singh, Fazel, Gueorguieva, & Buchanan, 2014).

The primary purpose of offender risk assessments is to predict an offender's probability for recidivating and/or engaging in institutional misconduct (Van Voorhis, 2009). Therefore, despite any other potential benefits (e.g., identifying treatment targets; allocating resources for case management), it is important that the risk instrument displays a high-level of predictive validity (Austin, 2006). That is, the tool must be able to accurately separate which offenders are more likely to reoffend from those who are less likely to reoffend (Bonta, 2002).

The accurate assessment of risk is not only important for the creation of a fair and equitable system, but is also described as an essential step in the successful reduction of recidivism/misconduct (Andrews & Bonta, 2010). Therefore, it is imperative that continuous efforts are made to improve the predictive accuracy of these tools. However, there currently exists a considerable amount of debate in the field with respect to what is the best strategy for achieving such improvements in prediction (see Andrews & Bonta, 2010; Austin, 2006; Barnoski, 2003; Baird, 2009; Van Voorhis, Wright, Salisbury, & Bauman, 2010). Among the differences of opinion discussed in the literature are opposing viewpoints regarding static and dynamic risk factors, the mathematics of computing scores (e.g., adding/removing items, norming cutoff scores to local populations, weighting items), and the use of offender responsibility factors (e.g., gender). There are also techniques used in other behavioral sciences fields, such as subgroup norming, that have yet to be explored within the field of offender risk prediction.

The Evolution of Offender Risk Assessments

During the last four decades, there have been several advancements made in the area of offender risk assessments (for a review see Andrews & Bonta, 2010). Bonta (1996) describes these improvements in terms of "generations." The *first-generation* of risk tools are based on professional judgments—a method that has been discredited as subjective, inconsistent, and for producing poor predictive accuracy (Brennan, Dieterich, & Ehret, 2009). The *second-generation* actuarial instruments, which are still used in many jurisdictions, use static factors, such as age, gender, and criminal history items. Examples of these tools include the Salient Factor Score (SFS; Hoffman, 1984), Statistical Inventory on Recidivism (SIR; Nuffield, 1982) and Static-99 (Hanson & Thornton, 1999). Second-generation instruments have empirically been shown to produce much higher predictive validity than first-generation assessments (Grove, Zald, Lebow,

Snitz, & Nelson, 2000); however, these tools have also been highly criticized for their reliance on static (i.e., unchangeable) factors (Hannah-Moffat, 2005).

Given the notion behind rehabilitation is that offenders are capable of reducing their propensity toward criminal behavior, it has been argued that risk assessments should be capable of assessing this change (Bonta, 1996). In response, *third-generation* risk/needs scales have been designed to capture the aspects of an offender's life that can be altered through effective correctional treatment to reduce his/her risk for recidivism (Andrews & Bonta, 2010). Although some static factors, such as criminal history, remain important features of third-generation tools, additional dynamic (i.e., changeable) factors, or criminogenic needs, have been added to these instruments (e.g., antisocial attitudes, antisocial personality, antisocial peers, employment, substance abuse). Two examples of these assessments include the Level of Service Inventory-Revised (LSI-R; Andrews & Bonta, 1995) and Wisconsin Risk and Needs assessment instrument (Baird et al., 1979).

Fourth-generation tools adopt a similar approach to third-generation instruments. However, fourth-generation assessments (e.g., Level of Service/Case Management Inventory [LS/CMI], Andrews, Bonta, & Wormith, 2004; and Correctional Offender Management Profile for Alternative Sanctions [COMPAS], Brennan & Oliver, 2000) seek to more directly integrate case management priorities (Andrews & Dowden, 2006). Meta-analytic investigations of third- and fourth-generation assessments suggest that both types display good predictive validity (Andrews, Bonta, & Wormith, 2006; Brennan et al., 2009; Campbell, French, & Gendreau, 2009; Gendreau, Little, & Goggin, 1996; Gendreau, Goggin, & Smith, 2002).

Improving Offender Risk Prediction

Offender risk prediction has advanced considerably from the early days of simply relying on professional judgments to make classification decisions (Bonta, 1996). However, there currently exists a considerable amount of disagreement regarding what strategies should be taken to improve prediction even further. This study examines the different viewpoints that have emerged in the areas of risk and need factors, scoring risk assessments, responsivity factors, and norming/subgroup norming.

Risk and need factors.

The contemporary position on offender risk assessments holds that risk/needs instruments are important not only for determining likelihood for reoffending, but also for identifying treatment target areas (Andrews & Bonta, 2010; Van Voorhis, 2009). The added goal of "risk reduction," rather than just "risk prediction," means risk instruments must contain both static and dynamic factors (Schwalbe, 2008). Meta-analytic findings indicate that higher predictive accuracy is achieved when both risk and need items are included in the same assessment (Gendreau et al., 1996). More recent investigations have also revealed support for the dynamic validity of risk assessment instruments; that is, changes to offender risk scores between initial assessment and reassessment can be used to further increase predictive accuracy of either score alone (Labrecque, Smith, Lovins, & Latessa, 2014; Vose, Smith, & Cullen, 2013). It has therefore been recommended from supporters of this position that correctional agencies should use, at a minimum, a third-generation risk/needs assessment (Bonta, 2007).

It should, however, be noted that not all scholars are in agreement with the mainstream position on the inclusion of risk *and* need items in the same assessment. For example, Baird (2009) has raised the question of whether or not “the promise of innovation has trumped actual performance” (p. 2). Baird (2009) notes that whereas early actuarial risk instruments generally consisted of relatively few items (less than a dozen), it is not uncommon for new third- and fourth-generation tools to contain between 25 and 100 factors. It is his contention that the inclusion of many of these additional items often creates problems with reliability, especially among the dynamic risk factors (e.g., peer relationships, marital/family, use of leisure time), which introduce noise into the prediction models, thereby decreasing, rather than increasing, predictive accuracy (see also Austin, 2006; Austin, Coleman, Peyton, & Johnson, 2003; Barnoski, 2003; 2006). According to this perspective, risk and need items should not be combined into one composite measure, but rather should be separated into two distinct assessments (one for assessing risk and one for identifying treatment target needs). Proponents of this viewpoint maintain third- and fourth-generation assessments are not worth the investment in additional time and effort it takes to administer them and argue for a return to the more simple and parsimonious second-generation actuarial risk assessment, which would consist of relatively few static criminal history risk factors (Baird, 2009).

Scoring risk assessments.

The standard method for scoring offender risk tools, known as the Burgess method (see Burgess, 1928), involves adding one point if the characteristic is present and adding no points if it is not. When the total points are summed, offenders with more points are thought to be more likely to reoffend than those with fewer points. There is an implicit assumption in this approach that all of the items are predictive of outcome for all offenders. However, some researchers have voiced concerns that many of the factors found within the risk/needs assessments tools do not maintain statistically significant univariate relationships with outcome (e.g., Austin, 2006; Baird, 2009). Therefore, from a purely statistical standpoint, it has been suggested that one way to improve the predictive accuracy of these tools is to remove all of the non-significant items.

There have been a handful of studies conducted in several different jurisdictions that indicate such improvements to prediction are achieved when non-statistically significant items have been removed from the standard risk instruments (see Austin et al., 2003; Austin, 2006; Barnoski, 2006). For example, in Pennsylvania, Austin et al. (2003) found the predictive accuracy of the LSI-R was improved by using only eight of the 54 total items, and in Vermont, Austin (2006) found improvements when 13 of the LSI-R's 54 items were used. Similarly, in Washington State, Barnoski (2006) found better results in predicting five-year sex recidivism when only five of the LSI-R's 54 items were used.

Responsivity factors.

The mainstream risk assessments used today (e.g., LSI-R) assume the predictors of crime are general. That is, the items in these assessments predict equally well for all offenders, regardless of factors such as gender, race, age, country of origin, etc. (Smith, Cullen, & Latessa, 2009; Vose, Cullen, & Smith, 2008). In addition, the specific responsivity principle maintains that treatment modalities and services ought to be adapted (or matched) to the specific characteristics of the offender (e.g., personality, emotions, cognitive abilities, intellectual

functioning; Andrews & Bonta, 2010; Andrews, Bonta, & Hoge, 1990; Gendreau, 1996). It is a widely held belief that specific responsivity factors do not directly relate to antisocial behavior; rather these factors mediate the relationship between criminogenic needs (e.g., antisocial attitudes, antisocial personality, antisocial peers) and criminal behavior. Accordingly, the focus of treatment should be on addressing criminogenic needs.

Although the research on specific responsivity is underexplored (Andrews & Bonta, 2010), there is a growing speculation that the predictors of crime may be related to some offender responsivity factors. This challenge to the conventional thinking is most evident in the gender responsive area where there is a growing body of gendered pathways research. This suggests that women offenders take different trajectories toward crime than men (Brennan, Breitenbach, Dieterich, Salisbury, Van Voorhis, 2012; Jones, Brown, Wanamaker, & Greiner, 2014; Salisbury & Van Voorhis, 2009). The feminist pathways literature makes a distinction between two types of risk factors, those that are “gender neutral,” or are associated with both male and female recidivism (e.g., antisocial attitudes, antisocial personality, antisocial peers) and those that are “gender responsive,” or are particularly relevant to female offending (e.g., victimization, relationship dysfunction, mental health; DeHart, Lynch, Belknap, Dass-Brailsford, & Green, 2014). It has been cautioned by this group that reliance on risk instruments that fail to consider gender responsive factors will lead to less than optimal levels of predictive validity for some female offenders and may ultimately lead to the selection of inferior rehabilitation strategies (Bloom, Owen, & Covington, 2003; Holtfreter & Cupp, 2007; Salisbury, Van Voorhis, & Spiropoulos, 2009; Van Voorhis et al., 2010).

Although qualitative research in the area of gender-responsive offending abounds, quantitative investigations have been less frequent (Salisbury & Van Voorhis, 2009). More recently, however, gender-responsive scholars have begun to produce more empirical evaluations (e.g., Salisbury et al., 2009; Van Voorhis et al., 2010). For example, Van Voorhis et al. (2010) developed a supplemental assessment that included gender-responsive needs (the Women’s Supplemental Risk/Needs Scale), which is administered to female offenders in conjunction with a standard risk assessment. Van Voorhis et al. (2010) determined that there was incremental validity gained by using the supplemental assessment (i.e., the use of both tools together was more accurate than the standard assessment alone).

Another responsivity concern that has been raised is whether risk assessments are equally valid across offender race/ethnicity categories. Some researchers have questioned whether offender risk instruments, which have primarily been developed on whites, are predictively valid for both whites *and* non-whites (Schlager & Simourd, 2007). The current empirical research in this area is unfortunately limited to only a handful of studies and the results are not conclusive. Some studies have reported the LSI-R both over-classifies (i.e., misclassifies low-risk offenders as high-risk) and under-classifies (i.e., misclassifies high-risk offenders as low-risk) some race groups more often than others (Fass, Heilbrun, DeMatteo, & Fretz, 2008; Holsinger, Lowenkamp, & Latessa, 2003; Whiteacre, 2006). Schlager and Simourd (2007) found the LSI-R was predictive for whites and Blacks, but not for Hispanics. Finally, Chenane, Brennan, Steiner, and Ellison (2015) discovered the LSI-R was predictive across race groups for prevalence of institutional misconduct, but not for incidence of misconduct.

Determining whether risk assessment tools are valid for all offenders is important because risk scores are used in sentencing, release, and treatment decisions. If the information from a risk assessment is not accurate for a particular group of offenders, then those individuals will not receive an appropriate system response (e.g., assigned level of supervision, type and

intensity of treatment). Therefore, there is a need for more research on offender specific responsivity characteristics and how these factors may be used to help improve the predictive accuracy of offender risk instruments.

Norming/subgroup norming.

In the application of offender assessments, some criminal justice agencies simply adopt a risk measure developed elsewhere, and fail to validate the assessment on their specific population. The practice of validating risk assessments with target populations (i.e., norming) has been encouraged as an important step in improving the predictive accuracy of the tool (Gottfredson & Snyder, 2005). There are a number of ways offender assessments can be validated and normed within a given offender population. The most common approach includes validating the assessment on the total target population, and if necessary, adjusting the cut-off scores to improve prediction (Flores, Lowenkamp, Holsinger, & Latessa, 2006). The use of the cut-points are linked to a corresponding recidivism probability rate, where higher scores are more likely to recidivate than lower scores.

In contrast to this standard approach, *subgroup norming* involves basing normative reference data on subgroups of a population (e.g., gender, race, age) rather than the total population. Subgroup norming is a form of model development with separate models that are generated for each specific subgroup. The process for the subgroup norming of risk assessments involves the examination of the correlations between individual items/domains on the risk instrument, and an outcome measure of recidivism broken down by subgroup categories of offenders. An *empirical key* is then constructed for each subgroup that weights the scores for each item/domain by the magnitude of the effect size (Sackett & Wilks, 1994). One potential benefit of subgroup norming is that it addresses the potential issue of individual items/domains of a risk instrument not predicting equally well across subgroups of offenders. It is the empiricism of this approach that considers the possibility that individual items/domains can be more effective predictors for some groups, while less predictive for another group. The importance of norming assessment instruments on relevant demographic variables (e.g., age, gender, race) has been well documented in other behavioral science fields, including psychology, education, and psychometrics (see Cicchetti, 1994). In fact, practices involving the norming of assessments to subgroup populations have existed since at least the 1970s, and are now quite commonplace in these fields (Guion & Gottier, 1965; Hunsley & Di Giulio, 2001; Jacobs & Manese, 1977; Sunil Rao, 2000).

Although subgroup norming is not yet widely used in the criminal justice field, there is some available evidence that suggests it may help improve prediction. For example, Knight, Ronis, and Zakireh (2009) used group differences to identify risk factors for the persistence of sexual offending. They examined unique and common characteristics between three groups of offenders, juvenile sex offenders, adult-juvenile sex offenders, and adult-non-juvenile sex offenders. Knight et al. (2009) found the importance of each risk factor differed by whether the offender first offended as a juvenile or an adult. Therefore, it is possible that subgroup norming could also be beneficial in other areas of corrections, such as with gender and race.

Current Study

It is important that efforts are made to improve the predictive accuracy of offender risk assessments. Four competing views have emerged in the literature regarding how this goal would best be achieved. The first perspective suggests the predictors of crime are general and apply equally to all offenders, the second contends only static criminal history risk items should be included, the third advocates for using only items with strong univariate statistical relationships, and the fourth suggests that the predictors of crime are related to offender responsibility factors (e.g., gender). This study presents another perspective that has not been widely explored in the field of criminal justice: subgroup norming. Unfortunately, the results from competing risk assessment improvement strategies are often not compared with one another in the same study. The significance of this study, therefore, is that it examines the utility of several different approaches of risk improvement by comparing the predictive accuracy between methods on the same sample of offenders.

METHOD

Participants

The participants in this study included a total of 1,549 probationers from one large mid-western state that were assessed with the Level of Service Inventory-Revised (LSI-R; Andrews & Bonta, 1995) by their supervising probation officer. The original database included a total of 1,599 probationers, but 3.1% of the sample (or 50 cases) was excluded due to missing demographic information of gender and/or race. Cases were excluded in this manner because such information was necessary in order to determine which subgroup an offender belonged.

Measures

LSI-R.

The Level of Service Inventory-Revised (Andrews & Bonta, 1995) is a standardized risk/need assessment instrument in which trained raters score each item as *present* (1) or *absent* (0) on the basis of a file review and interview with the offender. The LSI-R contains 54 items that fall into ten domains: Criminal History (10), Education/Employment (10), Financial (2), Family/Marital (4), Accommodation (3), Leisure/Recreation (2), Companions (5), Alcohol/Drugs (9), Emotional/Personal (5), and Attitudes/Orientation (4). An offender's score on this scale determines his or her overall level of risk, where higher scores indicate a greater risk for recidivism than lower scores. The LSI-R is operationalized here in the following four ways: (1) the standard LSI-R total score; (2) the Criminal History domain total score; (3) the selected LSI-R items total score; and (4) the subgroup weighted LSI-R total score. The scoring schemes for the constructed LSI-R variations (3 and 4) are explained below.

Recidivism.

There are many ways to measure recidivism. Some researchers argue that reincarceration is the best measure (Lowenkamp, Holsinger, & Latessa, 2001), while others maintain less

stringent criteria (e.g., technical violations) are acceptable (Bonta, Pang, & Wallace-Capretta, 1995). Further, others insist that the type of recidivism measure used to assess the predictive accuracy of the LSI-R “makes little difference” (Andrews et al., 2006, p. 17). In this study, the probation department opted to select arrest as the outcome of interest. Therefore, the recidivism variable is operationalized here as any new arrest within one-year of the LSI-R assessment date.

Offender subgroups.

Given the limited available research on specific responsivity factors in the corrections literature, each offender in the study was assigned to one of four subgroups based on the characteristics of gender (i.e., *male* or *female*) and race (i.e., *white* or *non-white*). These two factors were selected because they have long been considered important responsivity factors (Andrews et al., 1990). Table 1 describes the four offender subgroup categories, including the number of cases falling within each subgroup (see Appendix). For example, white males represent the largest subgroup ($n = 717$), whereas non-white females represent the smallest subgroup ($n = 109$). Table 1 also reports the percentage of offenders within each subgroup who were arrested during the follow-up time period. The subgroup of non-white males had the highest percentage of recidivists (46.8%), whereas the two female subgroups shared the lowest percentage (33.9%).

Construction and Validation Samples

Using SPSS 21.0, approximately half of the probationers were randomly assigned to the construction sample ($n = 794$) and the other half were randomly assigned to the validation sample ($n = 755$) without replacement. The construction sample was used to construct the scoring schemes for both the selected items LSI-R score and the weighted LSI-R score. The validation sample served as a validity test for the four scoring methods examined here.

A description of the construction and validation samples is presented in Table 2 (see Appendix). Generally speaking, the probationers in this study were predominately white males in their early thirties who were at low-moderate risk for reoffending, according to Andrew and Bonta’s (1995) recommended LSI-R cutoff categories. Table 2 also reports the mean scores for the four LSI-R variations investigated here. Approximately 40% of the offenders in both groups were arrested during the follow-up period. There were no significant differences between the two groups on any of these characteristics based on t tests or chi-square tests.

Selected LSI-R items score.

The selected LSI-R items score was calculated in the following manner. First, phi (ϕ) was calculated between each of the 54 items of the LSI-R (1 = *present* or 0 = *absent*) and recidivism (1 = *arrest* or 0 = *no arrest*). Phi was selected as the measure of association because its value provides a measure of strength between two categorical variables (Blalock, 1979). These analyses revealed that only 24 items had at least a low level of statistical association with recidivism for the construction sample (defined here as $\phi \geq .10$). Next, a regression analysis was run to see which of these items had an independent effect on recidivism (as described in Austin et al., 2003). This resulted in the following seven items remaining in the model:

1. Three or more prior convictions

2. Arrested under age 16
3. Currently unemployed
4. Non-rewarding, parental relationship
5. Criminal family/spouse
6. Alcohol/drug problem that interferes with school/work
7. Poor attitude toward supervision

Only two of the seven items found here were the same as the eight found by Austin et al. (2003)—*arrested under the age of 16* and *alcohol/drug problem that interferes with school/work*. Also, the majority of the statistically significant items were found to be dynamic (5), rather than static (2). The selected LSI-R items score was computed using the Burgess method where one point was added for each of the seven items that was present, with higher scores indicating a higher likelihood for recidivism.

Subgroup weighted LSI-R score

The empirical scoring key for the subgroup weighted LSI-R scores was also developed on the construction sample and involved the following three steps. First, a Pearson r was calculated between each of the 10 LSI-R domains scores and recidivism (1 = *arrest* or 0 = *no arrest*) for each of the four subgroups. Next, the group scores were standardized to make comparisons across subgroups possible. In order to do so, the second step involved dividing 54 (the highest possible score on the standard LSI-R assessment) by the sum of the total possible r values for each subgroup. Each of the domain scores were then multiplied by the quotient derived for their respective subgroup from step two, which produced the weighted value that was used to calculate the new LSI-R scores. Finally, the weighted LSI-R scores were calculated by multiplying an offender's domain score by the corresponding weighted value of that domain in the empirical key, and then summing the values of the 10 domains in order to produce a total subgroup weighted LSI-R score. For a full description of the empirical scoring key by domain and subgroup, please refer to Table 3 in the Appendix.

Statistical Analyses

In order to assess the predictive validity of these four different scoring methods (standard LSI-R, only criminal history factors, only statistically significant items, and subgroup normed weighted LSI-R), this study examines the area under the curve (AUC) statistics and 95% confidence intervals between the assessment scores and recidivism, which are also separated by subgroup. The use of the AUC statistic has become increasingly more common in the risk assessment prediction literature in part because it is less base rate sensitive than Pearson correlation coefficient (see Rice & Harris, 2005; Quinsey, Harris, Rice, & Cormier, 1998).

RESULTS

Table 4 examines the predictive validity of the four LSI-R scoring methods, by group type (see Appendix). For the standard LSI-R score, the magnitude of the AUC values varied slightly between groups. For example, the standard LSI-R produced a larger effect for males ($AUC = .68$, $CI = .64$ to $.72$) than for females ($AUC = .65$, $CI = .56$ to $.74$). Further, there was an even larger difference found between white females ($AUC = .62$, $CI = .50$ to $.74$) and non-

white females ($AUC = .68$, $CI = .53$ to $.83$). The CIs from all of the groups examined overlapped with one another.

The four scoring methods presented here were all found to be statistically significant predictors of arrest ($p < .001$). Within each group, the assessment type with the highest predictive validity is bolded. Note that for males, whites, and non-white males there is a tie in efficiency between scoring methods so there are two values bolded for these three groups. The criminal history score produced the lowest values across all of the group types examined, the standard LSI-R score held the advantage for non-whites (and ties for males and non-white males), the selected LSI-R items score produced the advantage for white females (and a tie for whites), and the subgroup weighted LSI-R score had the advantage with females, white males, and non-white females (and ties for males, whites, and non-white males). For the total sample, the predictive accuracy of the standard LSI-R score ($AUC = .67$, $CI = .63$ to $.71$) was higher than the criminal history domain score ($AUC = .62$, $CI = .58$ to $.66$) and selected LSI-R items score ($AUC = .66$, $CI = .62$ to $.70$), but was lower than the subgroup weighted LSI-R score ($AUC = .68$, $CI = .64$ to $.72$). The CIs for all scoring methods overlapped with one another across all of the groups examined.

DISCUSSION

The primary purpose of offender risk assessments is to predict an offender's probability for recidivating and/or engaging in institutional misconduct (Van Voorhis, 2009). Therefore, it is important that offender risk instruments display high levels of predictive accuracy (Austin, 2006). The LSI-R has also been described as the "most widely used and best validated measure of general criminal recidivism" (Hanson, 2005, p. 213). The predictive validity of the LSI-R has the support of large samples, hundreds of studies, and multiple meta-analyses (Flores, Lowenkamp, Smith, & Latessa, 2006; Gendreau et al., 2002; Gendreau et al., 1996; Lowenkamp & Bechtel, 2007; Smith et al., 2009; Vose, Cullen, & Smith, 2008). Andrews et al. (2006) reported the mean AUC value for the LSI-R predicting general recidivism was $.71$ ($r = .36$). This value falls within the CI for the LSI-R predicting arrest in this study ($.63$ to $.71$). Although this work supports the continued use of the LSI-R as an effective offender risk assessment tool, it is nonetheless still important that efforts are made to achieve greater and greater levels of predictive strength.

As discussed earlier, there is currently a great deal of debate in the literature as to what is the best strategy for achieving such improvements in prediction in which four competing views have emerged. The first perspective suggests the predictors of crime are general and apply equally to all offenders, the second contends only static risk items should be included, the third advocates for using only items with strong univariate statistical relationships, and the fourth suggests that the predictors of crime are related to offender responsivity factors (e.g., gender). Further, this study presents a fifth perspective that has not been widely explored in the field of criminal justice, which involves a process of subgroup norming and empirical keying. This study compared the results of these competing strategies with one another on the same sample of probationers. These methods for improving offender risk prediction will now be discussed along with the findings from this investigation.

Static Assessments

It has been argued elsewhere that risk and need items should not be included in the same measure (Baird, 2009). The findings in the current study do not support this notion. Specifically, the results from the static criminal history domain ($AUC = .62$, $CI = .58$ to $.66$) were lower when compared to those from the total LSI-R ($AUC = .67$, $CI = .63$ to $.71$). Further, the use of only static criminal history items to predict risk suggests that offenders are not capable of changing for the better (Bonta, 1996). Such a position is clearly at odds with the goals of rehabilitation because it does not view offenders as *dynamic* or *changeable*. However, the inclusion of the other risk/need factors also affords the ability to identify offender treatment targets, which supports the inclusion of the additional risk/need items.

Selected Item Scoring

Some scholars have raised concerns that third- and fourth-generation risk/needs assessments, such as the LSI-R, often include factors that fail to produce statistically significant univariate relationships with recidivism, and have raised questions about the utility of including these non-significant items (e.g., Austin, 2006; Austin et al., 2003; Barnoski, 2003). As predicted by this group of researchers, the predictive strength of the items in the LSI-R was found to vary. This suggests that all risk/needs items may not be equally related to recidivism.

Using the method described by Austin et al. (2003), a selected items score was computed by eliminating the items in the LSI-R that produced a non-significant univariate relationship with recidivism. This process produced a risk assessment with only seven items. This scoring method resulted in slight increases in prediction for some groups as well as some decreases for others. However, this procedure also removed many of the items that could be used to identify treatment targets. It is suggested here that it is important for risk assessments to include both theoretically relevant *and* empirically valid items so that the inclusion of these items (and domains) are not just mechanical (e.g., statistical), but rather are linked to the psychology of criminal conduct.

Specific Responsivity Considerations

Research on the influence of specific responsivity factors in offender risk prediction remains an understudied area. This study adds support to the prior validation studies have found the LSI-R to be predictively valid when the results were separated by the dichotomies of offender gender (as did Flores, Lowenkamp, Smith, et al., 2006; Lowenkamp et al., 2001; Raynor, 2007; Raynor & Miles, 2007; and Vose, Lowenkamp, Smith, & Cullen, 2009) and race (Chenane et al., 2015; Schlager & Simourd, 2007). However, this study was the first to extend the examination of the effectiveness of the tool by subgroup (gender *and* race). The findings of this study indicate that the LSI-R is predictively accurate for males ($AUC = .68$), females ($AUC = .65$), whites ($AUC = .68$), and non-whites ($AUC = .67$); however, when the results are examined by subgroup some differences in predictive accuracy emerge. Specifically, the LSI-R produces lower predictive accuracy for white females ($AUC = .62$) compared to white males ($AUC = .69$).

This result questions the logic of focusing risk assessment research on the role of just one offender characteristic (e.g., gender). Rather, this finding suggests that more research is needed

to examine what role subgroup differences may play in predicting outcome. However, there is no reason that subgroups should be limited to two characteristics examined here. In fact, it is very probable that with a large enough sample and more demographic information available that a more effective subgroup classification scheme could be constructed and might improve prediction even further than the current attempt.

Subgroup Norming

There are many benefits to prediction research, such as increased knowledge of phenomena, increased effectiveness in treatment, and the appropriate planning for different categories of offenders (Gottfredson & Snyder, 2005). Many agencies are looking for an ideal test that fairs well with their population. In this study, subgroup norming was introduced as a potential strategy to improve the prediction of offender recidivism. Further, it was suggested that subgroup norming might help shift the emphasis placed on collective prediction toward that of more individualized prediction. Although the results of the study are encouraging, they did not indicate that significant improvements were achieved via this method compared to the other strategies employed here (i.e., the CIs for all of the scoring methods examined overlapped between groups).

It is important to note that this study found the strengths of the associations between the LSI-Rs domain scores and recidivism varied both within and between offender subgroups. The practical implications of this finding suggest that criminogenic needs with higher correlations may be better targets for treatment, when compared to those with lower correlations. As such, the ideal treatment targets may differ among offenders depending upon which subgroup they belong. Further, unlike the criminal history and selected LSI-R items methods, the subgroup norming method retains all of the items in the LSI-R. Therefore, the subgroup norming method has the potential to be used for better-informed case management decisions.

Although this study contributes to the discussion on methods for improving the predictive validity of offender risk and need assessments, there are some limitations that should be considered in interpreting the findings. First and foremost, there is the issue of the studies sample size. Although the study began with 1,599 probationers, once the offenders were separated into subgroups and randomly assigned to either the construction or validation group, the cell frequencies in some of the subgroups became quite small. This was particularly evident with respect to the female subgroups (e.g., there were only 55 non-white females in the validation sample). Future subgroup norming investigations can address this limitation by ensuring that a more adequate representation of offenders is assigned to each subgroup.

Second, the data used in this investigation were limited to the items found in the LSI-R and did not include any information on additional gender-responsive items. It is therefore unknown whether the gender-responsive approach may have produced better results than the methods examined here. Future investigations in this area should compare the results achieved via gender-responsive assessments with other strategies to see which (if any) is capable of producing the highest levels of predictive accuracy. Further, it is currently unknown, but likely, if the subgroup norming method described here were applied to the additional items found in the gender-responsive tools, that further increases in predictive accuracy might be achieved. This process would also aid in determining which of the gender-responsive and gender-neutral needs have the highest effect sizes and thus warrant priority in treatment case planning.

Third, this study constructed subgroups based on the demographic characteristics of gender and race. There are certainly many other ways to categorize offenders that might prove to be more beneficial. There are also practical, legal and ethical considerations that arise when offenders are scored and treated differently based on demographic characteristics such as gender and race. Other researchers have approached the use of demographics by assigning points to an offenders risk score for simply possessing/not possessing a specific characteristic (e.g., gender, race, age, marital status, education level, etc.; see Austin et al., 2003; Austin, 2006; Barnoski, 2003). It should be noted that these researchers found that the inclusion of these items further improved the prediction of the assessment tool beyond its standard means. However, subgroup norming arguably seems a more ethical way to make use of offender demographic information to improve prediction, by simply comparing each offender to similarly matched offenders rather than giving additional points to offenders for possessing certain traits.

Fourth, the outcome variable examined here included any arrest within one year of the initial assessment. It would be beneficial for future studies to include longer follow-up periods (e.g., 36 months) and more diverse outcomes (e.g., institutional misconducts, incarceration, violent offenses). Fifth, the participants in this study were probationers from one Midwestern state. It is therefore unknown if the findings will generalize to other types of offenders and/or locations.

Sixth, this study used one particular method for quantifying factors into a weighted assessment scale. However, there are many other statistical techniques available (e.g., bootstrapping, classification trees, random forest modeling) that should also be explored in future studies of the topic (see Ridgeway, 2013; Ritter, 2013; and Steadman et al., 2000). Finally, this study used AUC as the performance indicator for its comparisons. There are certainly other ways to measure predictive performance (e.g., accuracy, calibration, clinical usefulness) that should be examined in future studies to help correctional authorities determine which offender risk assessment best meets their needs (see Jacobson & Truax, 1991; and Tollenaar & van der Heijden, 2013).

In conclusion, continued work in this area is critical in order to further advance the concept of risk and need assessment with offender populations. Such studies hold the potential to increase predictive validity of risk instruments, add to the generalizability of the risk assessments across offender group types, and point towards the best treatment targets for individual offenders. Luckily, given that correctional agencies use risk assessment instruments (e.g., LSI-R) so widely in the field and the information needed for such an investigation is commonly collected, this research could be easily done.

REFERENCES

- Andrews, D. A., & Bonta, J. (1995). *The Level of Service Inventory-Revised*. Toronto, Canada: Multi-Health Systems.
- Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). Cincinnati, OH: Anderson Publishing.
- Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior*, 17(1), 19-52.
- Andrews, D. A., Bonta, J., & Wormith, J. S. (2004). *The Level of Service/Case Management Inventory (LS/CMI)*. Toronto, Canada: Multi-Health Systems.
- Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime and Delinquency*, 52(1), 7-27.
- Andrews, D. A., & Dowden, C. (2006). Risk principle of case classification in correctional treatment: A meta-analytic investigation. *International Journal of Offender Therapy and Comparative Criminology*, 50(1), 88-100.
- Austin, J. (2006). How much risk can we take? The misuse of risk assessment in corrections. *Federal Probation*, 70(2), 58-63.
- Austin, J., Coleman, D., Peyton, J., & Johnson, K. D. (2003). *Reliability and validity study of the LSI-R risk assessment instrument*. Washington, DC: The Institute on Crime, Justice and Corrections at The George Washington University.
- Baird, C. (2009). *A critique of risk assessment models used in the criminal justice system*. Special Report, National Council on Crime and Delinquency.
- Baird, C., Heinz, R., & Bemus, B. (1979). *The Wisconsin Case Classification/Staff Deployment Project: Two-year follow-up report*. Madison, WI: Wisconsin Division of Corrections.
- Barnoski, R. (2003). *Washington's Offender Accountability Act: An analysis of the Department of Corrections' risk assessment*. Olympia, WA: Washington State Institute for Public Policy.
- Barnoski, R. (2006). *Sex offender sentencing in Washington state: Predicting recidivism based on the LSI-R*. Olympia, WA: Washington State Institute for Public Policy.
- Blalock, H. M., Jr. (1979). *Social statistics* (2nd ed.). New York: McGraw-Hill.
- Bloom, B., Owen, B., & Covington, S. (2003). *Gender responsive strategies for female offenders: A summary of research, practice, and guiding principles for women offenders*. Washington, DC: National Institute of Corrections.
- Bonta, J. (1996). Risk-need assessment and treatment. In A. T. Harland (Ed.), *Choosing correctional options that work: Defining the demand and evaluating the supply* (pp.18-32). Newbury Park, CA: Sage.
- Bonta, J. (2002). Offender risk assessment: Guidelines for selection and use. *Criminal Justice and Behavior*, 29(4), 355-379.
- Bonta, J. (2007). Offender risk assessment and sentencing. *Canadian Journal of Criminology and Criminal Justice*, 49(4), 519-529.
- Bonta, J., Pang, B., & Wallace-Capretta, S. (1995). Predictors of recidivism among incarcerated female offenders. *The Prison Journal*, 75(3), 277-294.
- Brennan, T., Breitenbach, M., Dieterich, W., Salisbury, E. J., & Van Voorhis, P. (2012). Women's pathways to serious and habitual crime: A person-centered analysis incorporating gender responsive factors. *Criminal Justice and Behavior*, 39(11), 1481-1508.

- Brennan, T., Dieterich, W., & Ehret, B. (2009). Evaluating the predictive validity of the COMPAS risk and needs assessment system. *Criminal Justice and Behavior*, 36(1), 21-40.
- Brennan, T., & Oliver, W. L. (2000). *Evaluation of reliability and validity of COMPAS scales: National aggregate sample*. Traverse City, MI: Northpointe Institute for Public Management.
- Burgess, E. W. (1928). Factors determining success or failure on parole. In A. A. Bruce, A. J. Harno, E. W. Burgess, & J. Landesco (Eds.), *The workings of the indeterminate-sentence law and the parole system in Illinois* (pp. 221-234). Springfield, IL: State Board of Parole.
- Campbell, M.A., French, S., & Gendreau, P. (2009). The prediction of violence in adult offenders. *Criminal Justice and Behavior*, 36(6), 567-590.
- Chenane, J. L., Brennan, P. K., Steiner, B., & Ellison, J. M. (2015). Racial and ethnic differences in the predictive validity of the Level of Service Inventory-Revised among prison inmates. *Criminal Justice and Behavior*. Advanced online publication.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284-290.
- DeHart, D., Lynch, S., Belknap, J., Dass-Brailsford, P., & Green, B. (2014). Life history models of female offending: The roles of serious mental illness and trauma in women's pathways to jail. *Psychology of Women Quarterly*, 38(1), 138-151.
- Fass, T. L., Heilbrun, K., DeMatteo, D., & Fretz, R. (2008). The LSI-R and Compas: Validation data on two risk-needs tools. *Criminal Justice and Behavior*, 35(9), 1095-1108.
- Flores, A. W., Lowenkamp, C. T., Holsinger, A. M., & Latessa, E. J. (2006). Predicting outcome with the Level of Service inventory-Revised: The importance of implementation integrity. *Journal of Criminal Justice*, 34(5), 523-529.
- Flores, A. W., Lowenkamp, C. T., Smith, P., & Latessa, E. J. (2006). Validating the Level of Service Inventory-Revised on a sample of federal probationers. *Federal Probation*, 70(2), 44-48.
- Gendreau, P. (1996). The principles of effective intervention with offenders. In A. T. Harland (Ed.), *Choosing correctional options that work: Defining the demand and evaluating the supply* (pp. 117-130). Thousand Oaks, CA: Sage.
- Gendreau, P., Goggin, C., & Smith, P. (2002). Is the PCL-R really the "unparalleled" measure of offender risk? A lesson in knowledge cumulation. *Criminal Justice and Behavior*, 29(4), 397-426.
- Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology*, 34(4), 575-607.
- Gottfredson, D. M., & Snyder, H. N. (2005). *The mathematics of risk classification: Changing data into valid instruments for juvenile courts* (NCJ 209158). Washington, DC: U.S. Department of Justice, Office of Justice Programs.
- Grove, W., Zald, D., Lebow, B., Snitz, B. & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychology Assessment*, 12(1), 19-30.
- Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology*, 18(2), 135-164.
- Hannah-Moffat, K. (2005). Criminogenic needs and the transformative risk subject: Hybridization of risk/need in penalty. *Punishment and Society*, 7(1), 29-51.

- Hanson, R. K. (2005). Twenty years of progress in violence risk assessment. *Journal of Interpersonal Violence*, 20(2), 212-217.
- Hanson, R. K., & Thornton, D. (1999). *Static-99: Improving actuarial risk assessments for sex offenders*. Ottawa, Canada: Department of the Solicitor General of Canada.
- Hoffman, P. (1984). Screening for risk: A revised Salient Factor Score. *Journal of Criminal Justice*, 11(6), 539-547.
- Holsinger, A. M., Lowenkamp, C. T., & Latessa, E. J. (2003). Ethnicity, gender, and the Level of Service Inventory-Revised. *Journal of Criminal Justice*, 31(4), 309-320.
- Holtfreter, K., & Cupp, R. (2007). Gender and risk assessment: The empirical status of the LSI-R for women. *Journal of Contemporary Criminology*, 23(4), 363-382.
- Hunsley, J., & Di Giulio, G. (2001). Norms, norming, and clinical assessment. *Clinical Psychology: Science and Practice*, 8(3), 378-382.
- Jacobs, R., & Manese, W. (1977). An empirical keying approach to academic prediction with the Guilford-Zimmerman temperament survey. *Journal of Experimental Education*, 45(3), 18-23.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59(1), 12-19.
- Jones, N. J., Brown, S. L., Wanamaker, K. A., & Greiner, L. E. (2014). A quantitative exploration of gendered pathways to crime in a sample of male and female juvenile offenders. *Feminist Criminology*, 9(2), 113-136.
- Knight, R. A., Ronis, S. T., & Zakireh, B. (2009). Bootstrapping persistence risk indicators for juveniles who sexually offend. *Behavioral Sciences and the Law*, 27(6), 878-909.
- Labrecque, R. M., Smith, P., Lovins, B., & Latessa, E. J. (2014). The importance of reassessment: How changes in the LSI-R risk score can improve the prediction of recidivism. *Journal of Offender Rehabilitation*, 53(2), 116-128.
- Lowenkamp, C. T., & Bechtel, K. (2007). The predictive validity of the LSI-R on a sample of offenders drawn from the records of the Iowa department of corrections data management system. *Federal Probation*, 71(3), 25-29.
- Lowenkamp, C. T., Holsinger, A., & Latessa, E. J. (2001). Risk/need assessment, offender classification and the role of childhood abuse. *Criminal Justice and Behavior*, 28(5), 543-563.
- Lowenkamp, C. T., Latessa, E. J., & Holsinger, A. M. (2006). The risk principle in action: What have we learned from 13,676 offenders and 97 correctional programs? *Crime and Delinquency*, 52(1), 77-93.
- Nuffield, J. (1982). *Parole decision-making in Canada: Research towards decision guidelines*. Ottawa, Canada: Solicitor General of Canada.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.
- Raynor, P. (2007). Risk and need assessment in British probation: The contribution of the LSI-R. *Psychology, Crime, & Law*, 13(2), 125-138.
- Raynor, P., & Miles, H. (2007). Evidence-based probation in a microstate: The British Channel Island of Jersey. *European Journal of Criminology*, 4(3), 299-313.
- Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's *d*, and *r*. *Law and Human Behavior*, 29(5), 615-620.
- Ridgeway, G. (2013). The pitfalls of prediction. *National Institute of Justice Journal*, 271, 34-40.

- Ritter, N. (2013). Predicting recidivism risk: New tool in Philadelphia shows great promise. *National Institute of Justice Journal*, 271, 4-13.
- Sackett, P. R., & Wilk, S. L. (1994). Within-group norming and other forms of score adjustment in preemployment testing. *American Psychologist*, 49(11), 929-954.
- Salisbury, E. J., & Van Voorhis, P. (2009). Gendered pathways: A quantitative investigation of women probationers' paths to incarceration. *Criminal Justice and Behavior*, 36(6), 541-566.
- Salisbury, E. J., Van Voorhis, P., & Spiropoulos, G. V. (2009). The predictive validity of a gender-responsive needs assessment: An exploratory study. *Crime and Delinquency*, 55(4), 550-585.
- Schlager, M., & Simourd, D. (2007). Validity of the Level of Service Inventory-Revised (LSI-R) among African American and Hispanic male offenders. *Criminal Justice and Behavior*, 34(4), 545-554.
- Schwalbe, C. (2008). Strengthening the integration of actuarial risk assessment with clinical judgment in an evidence-based practice framework. *Children and Youth Services Review*, 30(12), 1458-1464.
- Singh, J. P., Fazel, S., Gueorguieva, R., & Buchanan, A. (2014). Rates of violence in patients classified as high risk by structured risk assessment instruments. *The British Journal of Psychiatry*, 204, 180-187.
- Smith, P., Cullen, F. T., & Latessa, E. J. (2009). Can 14,737 women be wrong? A meta-analysis of LSI-R and recidivism for female offenders. *Criminology and Public Policy*, 8(1), 183-208.
- Steadman, H. J., Silver, E., Monahan, J., Appelbaum, P. S., Clark-Robbins, P., Mulvey, E. P., Grisso, T., Roth, L. H., & Banks, S. (2000). A classification tree approach to the development of actuarial violence risk assessment tools. *Law and Human Behavior*, 24(1), 83-100.
- Sunil Rao, J. (2000). Bootstrapping to assess and improve atmospheric prediction models. *Data Mining and Knowledge Discovery*, 4(1), 29-41.
- Tollenaar, N., & van der Heijden, P. G. M. (2013). Which method predicts recidivism best?: A comparison of statistical, machine learning and data mining predictive models. *Journal of the Royal Statistical Society*, 176(2), 565-584.
- Van Voorhis, P. (2009). An overview of offender classification systems. In P. Van Voorhis, M. Braswell, & D. Lester (Eds.), *Correctional counseling and rehabilitation* (7th ed., pp. 133-161). Cincinnati, OH: Lexis/Nexis.
- Van Voorhis, P., Wright, E. M., Salisbury, E., & Bauman, A. (2010). Women's risk factors and their contributions to existing risk/needs assessment: The current status of a gender-responsive supplement. *Criminal Justice and Behavior*, 37(3), 261-288.
- Vose, B., Cullen, F. T., & Smith, P. (2008). The empirical status of the Level of Service Inventory. *Federal Probation*, 72(3), 22-29.
- Vose, B., Lowenkamp, C., Smith, P., & Cullen, F. T. (2009). Gender and the predictive validity of the LSI-R: A study of parolees and probationers. *Journal of Contemporary Criminal Justice*, 25(4), 459-471.
- Vose, B., Smith, P., & Cullen, F. T. (2013). Predictive validity and the impact of change in total LSI-R score on recidivism. *Criminal Justice and Behavior*, 40(12), 1383-1396.
- Whiteacre, K. W. (2006). Testing the Level of Service Inventory-Revised (LSI-R) for racial/ethnic bias. *Criminal Justice Policy Review*, 17(3), 330-342.

APPENDIX

Table 1

Offender Subgroup Category Information (N = 1,549)

Subgroup	<i>n</i>	% Recidivists
White Males	717	39.2
Non-White Males	496	46.8
White Females	227	33.9
Non-White Females	109	33.9

Table 2

Descriptive Characteristics of Offenders, by Group Type

Characteristic	Construction (<i>n</i> = 794)		Validation (<i>n</i> = 755)	
	<i>n</i>	%	<i>n</i>	%
Male	615	77.5	598	79.2
White	484	61.0	460	60.9
Age ^a	32.0	10.9	32.8	12.0
Mean LSI-R Score ^b (SD)	18.5	8.2	17.9	8.7
Mean Criminal History Score ^c (SD)	3.3	2.2	3.2	2.2
Mean Selected LSI-R Items Score (SD)	2.2	1.5	2.1	1.5
Mean Subgroup Weighted LSI-R Score ^b (SD)	18.6	8.5	18.1	8.9
Recidivism	322	40.6	305	40.4

Note: ^a Construction group *n* = 786 and validation group *n* = 751 due to missing information.

^b Range of scores = 0 to 54; ^c Range of scores = 0 to 10; ^d Range of scores = 0 to 7.

Table 3
LSI-R Empirical Key by Subgroup

Domain (n of items)	White Males			Non White Males			White Females			Non White Females		
	r	item weight	domain weight	r	item weight	domain weight	r	item weight	domain weight	r	item weight	domain weight
Criminal History (10)	0.17	1.39	13.89	0.33	1.67	16.67	0.13	1.52	15.25	0.27	1.09	10.93
Education/Employment (10)	0.15	1.22	12.24	0.19	0.95	9.5	0.08	0.93	9.26	0.4	1.61	16.19
Financial (2)	0.15	1.25	2.5	0.12	0.62	1.24	0.02	0.28	0.55	0.23	0.93	1.86
Family/Marital (4)	0.1	0.84	3.34	0.22	1.1	4.39	0.12	1.35	5.41	0.2	0.83	3.33
Accommodation (3)	0.09	0.78	2.33	0.12	0.58	1.75	0.01	0.06	0.17	0.24	0.96	2.87
Leisure/Recreation (2)	0.12	0.98	1.95	0.16	0.81	1.62	0.03	0.39	0.78	0.11	0.46	0.92
Companions (5)	0.13	1.06	5.29	0.25	1.24	6.22	0.13	1.52	7.62	0.14	0.58	2.92
Alcohol/Drug Problem (9)	0.11	0.91	8.19	0.15	0.78	7.04	0.06	0.73	6.6	0.11	0.45	4.07
Emotional/Personal (5)	0.04	0.35	1.74	0.12	0.58	2.92	0.01	0.13	0.63	0.18	0.73	3.63
Attitudes/Orientation (4)	0.08	0.63	2.51	0.13	0.66	2.64	0.17	1.93	7.7	0.45	1.82	7.29

Table 4

Comparative Predictive Validity of Four LSI-R Scoring Methods, by Group Type

Group	Standard LSI-R		Criminal History		Selected LSI-R Items		Subgroup Weighted LSI-R	
	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI
Males	.68	[.64, .72]	.62	[.57, .66]	.67	[.62, .71]	.68	[.64, .73]
Females	.65	[.56, .74]	.63	[.54, .73]	.64	[.55, .73]	.66	[.57, .75]
Whites	.68	[.63, .73]	.64	[.59, .69]	.69	[.64, .74]	.69	[.64, .74]
Non-Whites	.67	[.60, .73]	.61	[.54, .67]	.62	[.56, .68]	.66	[.60, .72]
White Males	.69	[.64, .75]	.64	[.58, .70]	.69	[.64, .75]	.70	[.64, .76]
Non-White Males	.67	[.60, .74]	.59	[.52, .66]	.64	[.57, .71]	.67	[.60, .73]
White Females	.62	[.50, .74]	.61	[.49, .73]	.68	[.58, .79]	.64	[.53, .76]
Non-White Females	.68	[.53, .83]	.67	[.53, .82]	.56	[.40, .72]	.69	[.54, .84]
Total Sample	.67	[.63, .71]	.62	[.58, .66]	.66	[.62, .70]	.68	[.64, .72]