Thirty Years of Research on the Level of Service Scales: A Meta-Analytic Examination of Predictive Accuracy and Sources of Variability

Article *in* Psychological Assessment · November 2013

DOI: 10.1037/a0035080 · Source: PubMed

3 authors, including:

J. Stephen Wormith
University of Saskatchewan
89 PUBLICATIONS **3,813** CITATIONS

SEE PROFILE

Psychological Assessment
2014, Vol. 26, No. 1, 156–176

1040-3590/14/$12.00 DOI: 10.1037/a0035080

Thirty Years of Research on the Level of Service Scales: A Meta-Analytic Examination of Predictive Accuracy and Sources of Variability

Mark E. Olver
University of Saskatchewan

Keira C. Stockdale
Saskatoon Police Service, Saskatoon, Saskatchewan, Canada, and University of Saskatchewan

J. Stephen Wormith
University of Saskatchewan

We conducted a comprehensive meta-analysis of the Level of Service (LS) scales, their predictive accuracy and group-based differences in risk/need, across 128 studies comprising 151 independent samples and a total of 137,931 offenders. Important potential moderators were examined including ethnicity, gender, LS scale variant, geographic region, and type of recidivism used to measure outcome. Results supported the predictive accuracy of the LS scales and their criminogenic need domains for general and violent recidivism overall, and among broad subgroups of interest, including females and ethnic minorities. Although results indicated that gender and ethnicity were not substantive sources of effect size variability, significant differences in effect size magnitude were found when analyses were conducted by geographic region. Canadian samples consistently demonstrated the largest effect sizes, followed by studies conducted outside North America, and then studies conducted in the United States. This pattern was observed irrespective of gender, ethnicity, LS domain, LS variant, or type of recidivism outcome, suggesting geographic region may be an important source of effect size variation. We discuss possible factors underlying this pattern of results and identify areas for future research.

Keywords: LSI, risk assessment, recidivism, meta-analysis

The Level of Service Inventory (LSI) was the first of a family of tools, broadly referred to as the Level of Service (LS) scales, designed to link offender assessment and intervention, that is, to appraise recidivism risk, to identify criminogenic needs (i.e., dynamic risk factors) for intervention, and to inform recommendations for treatment, case management, and community supervision. Intended for use by a range of criminal justice personnel, including mental health professionals and parole and probation officers, the LS scales have become the most frequently used risk assessment tools on the planet, recording 1,085,647 "officially declared administrations" in 2010 alone (Wormith, 2011, p. 80).

Initially developed in 1982 as the Level of Supervision Inventory (LSI; Andrews, 1982) and subsequently its companion, the Youth Level of Service Inventory (YLSI; Andrews, Robinson, & Hoge, 1984), the LS scales have undergone two substantial revisions. They include variants for youth and adult offender populations, self-report and screening versions, and adaptations for use in specific settings and jurisdictions. However, all LS scales are organized around a common structure of clusters of binary items featuring the "Big Four" covariates of criminal conduct (criminal history, antisocial attitudes, antisocial associates, and antisocial personality pattern) and what have become known as the "Central Eight" (adding the domains of education/employment, family/marital, leisure recreation, and substance abuse). Early versions also included segments devoted to financial, and accommodation domains, which were dropped following further validation research, and a personal/emotional domain, which was modified to antisocial pattern in keeping with the Central Eight.

The original LSI was followed by the Level of Service Inventory–Revised (LSI–R; Andrews & Bonta, 1995a), which remains the most widely used version, and its short form, the Level of Service Inventory–Revised: Screening Version (LSI–R:SV; Andrews & Bonta, 1995b). It was then followed by what the authors have referred to as "fourth generation risk assessment scales"

This article was published Online First November 25, 2013.

Mark E. Olver, Department of Psychology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada; Keira C. Stockdale, Saskatoon Police Service, Saskatoon, Saskatchewan, Canada, and Department of Psychology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada; J. Stephen Wormith, Department of Psychology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

The views, opinions, and assumptions expressed in this article are those of the authors and do not necessarily reflect the views or official positions of the University of Saskatchewan, Saskatoon Police Service, the Saskatoon Board of Police Commissioners, or the City of Saskatoon. J. Stephen Wormith receives royalties from sales of the Level of Service/Case Management Inventory from its publisher, Multi-Health Systems.

The authors would like to express their appreciation to James Bonta for contributing a number of unpublished Level of Service documents from his personal collection. Funding support for this research was provided by an internal grant awarded to the authors from The Centre for Forensic Behavioural Sciences and Justice Studies at the University of Saskatchewan.

Correspondence concerning this article should be addressed to Mark E. Olver, Department of Psychology, University of Saskatchewan, 9 Campus Drive, Arts Building Room 154, Saskatoon, SK S7N 5A5, Canada. E-mail: mark.olver@usask.ca

(Andrews, Bonta, & Wormith, 2006) that focus directly on the Central Eight, but also include supplementary scales and a case management component. They are the Level of Service/Case Management Inventory (LS/CMI; Andrews, Bonta, & Wormith, 2004) and Youth Level of Service Inventory/Case Management Inventory (YLS/CMI; Hoge & Andrews, 2003) and specific jurisdictional versions, the Level of Service Inventory–Ontario Revision (LSI–OR; Andrews, Bonta, & Wormith, 1995), which also served as a pilot version of the LS/CMI, and the Level of Service/Case Management Inventory: Saskatchewan Youth Edition (LSI–SK; Andrews, Bonta, & Wormith, 2001).

The Big Four and Central Eight underpin a general personality and cognitive social learning theory of criminal behavior that provides an explanatory model of the origin and continuation of criminal conduct, and informs methods for predicting, reducing, managing, and preventing criminal behavior (Andrews & Bonta, 1994, 2010). An application of this model, bridging the practices of assessment and intervention, are the principles of *risk* (match service intensity to the risk level of the client), *need* (target criminogenic needs, such as the Central Eight, for intervention), and *responsivity* (use of cognitive behaviorally based interventions, known as *general responsivity*, and tailoring service delivery to the idiosyncratic features of clientele such as motivation, culture, and cognitive ability, known as *specific responsivity*).

The LS scales fit well within the risk–need–responsivity (RNR; Andrews, Bonta, & Hoge, 1990) framework. Since the inception of these scales, there have been several evaluations of their psychometric properties, perhaps the most prominent of which has been the capacity of these instruments to accurately assess risk and predict subsequent recidivism. Certainly there is more to determining the worth of a risk assessment measure than mere prediction; however, strong predictive accuracy is a prerequisite in order for a tool to be useful for the many other applications that potentially follow from its use, as outlined by the RNR principles.

Meta-Analytic Findings: Clinical and Empirical Issues

There have been several meta-analyses of the family of LS scales, and these tools, in turn, have been situated in the midst of some important controversies in clinical forensic assessment research and practice. The results of meta-analyses that have included the LS scales are summarized in Table 1. Rice and Harris (2005) provided guidelines for the interpretation and conversion of effect sizes used in recidivism prediction, which can be used to interpret the values in Table 1, with point biserial correlations and equivalent Cohen's *d* values corresponding to small ($r = .10$, $d = .20$), medium ($r = .24$, $d = .50$), and large ($r = .37$, $d = .80$) effect sizes.

Comparative Predictive Accuracy to Other Tools

One potential area of debate concerns how well the LS scales predict various forms of recidivism and how this stands up to other tools. Gendreau, Goggin, and Smith (2002) found the LS tools to have high predictive accuracy for general recidivism, and moderate accuracy for the prediction of violence, concluding that the LS scales predicted general recidivism better than the Hare Psychopathy Checklist (PCL) scales and were at least as accurate for violence. Yang, Wong, and Coid (2010) subsequently used multilevel modeling procedures to draw direct comparisons among a collection of forensic assessment instruments, including the LS and PCL scales, in the prediction of violence. Limiting their

Table 1
Summary of Level of Service (LS) Predictive Accuracy Meta-Analyses

Meta-analysis	LS version	Sample composition	*n*	*k*	Recidivism criterion	Effect size
Gendreau et al. (2002)	LSI, LSI–R, LSI–OR, YO–LSI, Y–LSI	Both genders, all ages	7,367	33	General	$r = .39$
			3,297	16	Violent	$r = .28$
Schwalbe (2007)	YLS/CMI	Both genders, youth	3,265	11	General	AUC = .64, $r = .25$
Schwalbe (2008)	YLS/CMI	Female youth	204	3	General	$r = .32$
		Male youth	772	4	General	$r = .31$
Campbell et al. (2009)	LSI, LSI–R, LSI–OR, LS/CMI	Both genders, adult	4,361	19	Violent (community)	$r = .28$
			650	6	(institutional)	$r = .24$
Olver et al. (2009)	YLS/CMI (and SV and AA), LSI–SK, YO–LSI, Y–LSI, LSI–OR	Both genders, youth	5,722	19	General	$r = .32$
			1,995		Violent	$r = .25$
		Female youth	992	9	General	$r = .36$
			350	4	Violent	$r = .24$
		Male youth	2,968	9	General	$r = .33$
			974	4	Violent	$r = .23$
		Aboriginal youth	860	5	General	$r = .35$
		Non-Aboriginal youth	462	5		$r = .32$
Smith et al. (2009)	LSI, LSI–R	Female, adult	14,737	27	General	$r = .35$
		Within-study comparisons				
		Female adult	9,250	16	General	$r = .27$
		Male adult	33,616	16		$r = .26$
Yang et al. (2010)	LSI, LSI–R, LS/CMI	Both genders, adult	355	3	Violent	$r = .25$
Singh et al. (2011)	LSI–R, LS/CMI	Both genders, adult	4,005	8	Violent	*OR* = 1.75 (converted $r \approx .15$)

Note. LSI = Level of Service Inventory; LSI–R = Level of Service Inventory–Revised; LSI–OR = Level of Service Inventory–Ontario Revision; YO–LSI = Youth Offender Level of Service Inventory; Y–LSI = Youth Level of Service Inventory; YLS/CMI = Youth Level of Service Inventory/Case Management Inventory; LS/CMI = Level of Service/Case Management Inventory; SV = Screening Version; AA = Australian Adaptation; LSI–SK = Level of Service/Case Management Inventory: Saskatchewan Youth Edition; AUC = area under the curve; *OR* = odds ratio.

analyses to within-study comparisons (i.e., studies in which two or more instruments were directly examined on the same sample), Yang et al. found that all of the tools forecasted violence with comparable degrees of accuracy, and that effect size variability was accounted for by specific features of the study (e.g., region, sample, setting etc.) rather than any special property of the tools themselves. Singh, Grann, and Fazel (2011), by contrast, found the LS measures to have the weakest predictive accuracy for violence relative to other forensic assessment tools; however, this review did not draw within-study comparisons, use multilevel procedures, or obtain a comprehensive collection of LS studies from the period sampled, and they used binning procedures (i.e., risk bins from the tools were dichotomized and the data reanalyzed in 2 × 2 contingency tables), which would reduce risk scale variance, particularly for longer scales such as the LS tools, and hence, predictive accuracy.

Applications to Female Offenders

A second issue of controversy concerns the use of the LS scales with female offenders. Arguments have been advanced that female offenders constitute a unique group, with gendered pathways to crime, and thus unique circumstances and special service delivery needs—what has been referred to as a "gender informed" or "gender responsive" perspective. The LS scales have been criticized as not capturing or giving sufficient weight to the full range of needs unique to female offenders (Blanchette & Brown, 2006; Hannah-Moffat, 2009). Additional gender responsive needs have been identified such as parental stress, low self-esteem, childhood, domestic and sexual abuse, anger concerns, and poverty among other areas. Efforts have included developing gender-informed materials to supplement mainstream risk–need tools such as the LS scales (e.g., Van Voorhis, Wright, Salisbury, & Bauman, 2010). Whereas some researchers have found evidence for gender-informed supplements or specific gender-responsive needs to have incremental value beyond the LS scales in the prediction of recidivism (Van Voorhis et al., 2010), others have not (Rettinger & Andrews, 2010).

Gender neutral theory, on the other hand, contends that male and female offenders have similar criminogenic needs and can benefit from similar models of crime reduction (e.g., RNR); however, gender is viewed to be an important responsivity consideration with important implications for program design, intervention planning, and service delivery. Although male and female offenders are acknowledged to have important differences, the key covariates of criminal behavior and the methods for reducing it, such as RNR, are generally consistent, irrespective of gender (Andrews et al., 2011). As seen in Table 1, Smith, Cullen, and Latessa (2009) found quite strong predictive accuracy for the LS scales in adult female offenders, as did Schwalbe (2008) and Olver, Stockdale, and Wormith (2009) with female young offenders, with effect size magnitudes highly consistent with those found with male offenders.

Use With Ethnic Minorities

A further issue of potential controversy concerns the use of risk–needs measures, including the LS scales, with ethnic minorities. Ethnic minorities are overrepresented in correctional settings throughout North America and international jurisdictions (e.g., Brzozowski, Taylor-Butts, & Johnson, 2006; Calverley, 2007). Understandable apprehensions have been expressed about the appropriate use of risk assessment instruments with such populations given concerns voiced regarding limited research on the psychometric properties with ethnic minorities, lack of separate norms, ensuring proper training by administrators, and lack of attention to issues of diversity (Hannah-Moffat & Maurutto, 2003; Martel, Brassard, & Jaccoud, 2011). For instance, a review by Rugge (2006) concluded that Canadian Aboriginal offenders tended to score higher than non-Aboriginal offenders on forensic assessment tools, were more frequently classified as high risk, and demonstrated higher rates of recidivism. Aboriginal peoples, however, are also more likely to be victims of violent crime, to experience poverty and unemployment, and to have less formal education (Perreault, 2011; Scrim, 2010). Rugge noted that though Aboriginal ancestry may be a risk factor for crime, being of Aboriginal ancestry does not directly cause crime. Rather, some risk factors may be overrepresented among Aboriginal persons that increase their scores on such tools and may serve, in part, to increase their likelihood of coming in contact with the justice system.

Olver et al. (2009) found the LS scales to predict general recidivism across five samples of Aboriginal and non-Aboriginal youth, with comparable predictive accuracy among both broad ethnic groups. Recently, Gutierrez, Wilson, Rugge, and Bonta (2013) conducted a meta-analysis of the Central Eight risk factors gleaned from forensic assessment tools (including the LS scales) or operationalized through other means. Comparing Aboriginal and non-Aboriginal offenders, Gutierrez et al. found that all eight domains significantly predicted general and violent recidivism among both broad ethnic groups. Although slightly higher effect sizes emerged for the non-Aboriginal offenders for most domains in the prediction of general recidivism, these differences tended to be small in magnitude and were less frequent in the prediction of violence.

Present Study: The Need for Another LS Meta-Analysis

A solid foundation of research of the LS scales has been developed, by and large, supporting their criterion-related validity for important criminal justice outcomes. Issues persist, however, concerning the psychometric appropriateness, clinical utility, and theoretical relevance of the LS scales and individual domains with special offender populations. Important gaps in the literature also remain. For one, differences in the predictive accuracy of the LS tools, and their use in classification and case management decisions, have yet to be subjected to quantitative review among ethnic minority groups on a larger scale for both youth and adult populations. Moreover, only one of the aforementioned quantitative reviews examined the individual need domains of the LS scales, such as the Central Eight, as a function of gender, but did so with a limited number of studies ($k = 5$; Andrews et al., 2012). Finally, research has yet to draw comparisons among the many variants of the LS tools, or to examine the impact of other potentially important sources of effect size variation, such as geographic region of the study, incorporating all versions of the LS scales. Given that the LSI is the most frequently used risk assessment tool internationally, employed by legions of parole and probation offices, prisons

and hospitals, forensic examiners, and courts around the globe, these outstanding issues merit empirical attention in general, and an updated meta-analysis in particular. As such, in the present study we seek to redress these specific gaps in the literature and extend existing findings by way of a large-scale meta-analysis of the predictive accuracy of the family of LS scales and examination of potential sources of effect size variability.

Method

Selection of Studies

To identify studies that examined variants of LS instruments, we conducted computer searches of PsycINFO, ProQuest Dissertations and Theses, and Google Scholar using combinations of the search terms *LSI*, *Level of Supervision Inventory*, *Level of Service Inventory*, and *Level of Service Case Management Inventory*. Additional sources included articles in print and electronic format accumulated by the authors over several years, a review of well-known criminal justice journals (e.g., *Criminal Justice and Behavior*), and examination of reference lists from previous meta-analyses that included the LS tools (see Table 1).

Studies were examined for their suitability for inclusion with the following criteria. The studies must have included (a) one of the versions of the LS, such as the original LSI, LSI–R, LSI–R:SV, LSI Self-Report, YLS/CMI, YLS/CMI Screening Version (Hoge & Andrews, 2003) and its Australian Adaptation, LSI–SK, Youth Offender LSI, LSI–OR, and LS/CMI; (b) a measure of recidivism outcome (e.g., arrests, charges, convictions, etc.) in the institution or community after a period of follow-up, or the mean comparisons between gender or ethnic groups; and (c) sufficient information to code or compute a predictive validity effect size in terms of a common metric (Pearson *r* or a point biserial *r*). For some published work, the original thesis or dissertation was consulted to provide more detailed information either to obtain *r* or to compute it.

Procedure

A coding protocol was completed for each study in the analysis including author and source, geographic location, sample demographics, offender group, setting and facility, LS version, means and standard deviations of LS total scores and need domains, recidivism base rates, LS risk categories and recidivism percentages, and any predictive accuracy statistics for all community and institutional outcomes for LS total and criminogenic need scores. LS descriptives, predictive accuracy, and recidivism data were also coded for males and females and ethnic minority and nonminority groups. Studies were coded by the first and second authors. Nonredundant information was coded as much as possible to reduce the impact of a particular sample on aggregate findings, and care was taken to ensure that a given set of data was coded from one sample only. When a given finding from the same sample was reported across multiple publications, the effect size was coded from the largest or most representative and recent sample. Approximately 10% of the studies ($n = 13$) were randomly selected and independently recoded, and effect sizes were recomputed by either the first or second author, depending on who had coded the original study. An overall rate of agreement of 96.9% (444/458) was achieved for the study variables coded and effect sizes extracted. Discrepancies resulted from minor computational or coding errors or simple omissions and were resolved by consensus between raters.

Effect size coding. Predictive accuracy statistics were coded in terms of *r*, which in most cases was a point biserial correlation, or r_{pb} (i.e., a correlation between a continuous predictor, such as the score on a risk measure, and a binary criterion variable, such as dichotomous recidivism coded yes–no). When *r* was not reported, the appropriate formula was applied to convert the reported statistic or descriptive information (e.g., mean group differences between recidivists and nonrecidivists) into *r*, specifically a point biserial *r*, phi *r* (for 2 × 2 tables), or Cramer's *V* (for more than two predictor categories, such as LSI risk levels of low, medium, and high, and its association with binary outcome). Occasionally, a standard Pearson *r* was computed from a continuous outcome variable (e.g., number of new convictions). When only an area under the curve (AUC) statistic from receiver operating characteristic analyses was provided, the formulae provided in Rice and Harris (2005) were used. AUC values were first converted into the equivalent Cohen's *d*, in which $d = \sqrt{2} \times z(\text{AUC})$, and then into the equivalent r_{pb} using the formula $r = d/(\sqrt{d^2 + [1/pq]})$, where p = base rate and $q = 1 - p$. In some cases, multiple dependent measures had been coded on a single sample (e.g., separate correlations computed for binary charges and binary convictions). In such cases when more than one effect size represented a particular outcome measure within a study, a single effect size was created by averaging the two (see Lipsey & Wilson, 2001). Finally, for gender and ethnic group comparisons on LS score, the mean, standard deviation, and sample size were used to compute a variation on Cohen's *d* (Hedges's *g*) in which the mean difference between groups (e.g., male–female, ethnic minority–nonminority) is divided by the pooled standard deviation.

Data aggregation. All coded data were entered into a spreadsheet with SPSS for Windows 20.0. Mean weighted effect sizes, *r* and *d*, were then computed with Comprehensive Meta-Analysis 2.0 (Borenstein, Hedges, Higgins, & Rothstein, 2005). Both fixed- and random-effects models were used in the computation of *r*. In fixed-effects models, the correlation is simply weighted by the sample size of the study from which it is derived, with larger studies thus receiving greater weight in effect size aggregation. By contrast, in random-effects models, less importance is given to differences in sample size across studies due to the inclusion of a constant that represents unexplained variation across studies. As a result, relatively greater weight is given to smaller studies compared to the fixed-effects model, with the random-effects model approximating the unweighted average.

The studies demonstrated considerable variability in the magnitude of their effects. Homogeneity analyses were conducted to examine whether the effect sizes were dispersed around their mean, no greater than would be expected from sampling error alone, through computing the *Q* statistic, which is distributed as a chi-square and its significance is evaluated on $k - 1$ degrees of freedom. A significant *Q* indicates significant variability in effect sizes among studies (Lipsey & Wilson, 2001). We also computed I^2 to quantify the amount of effect size variability in which I^2 values of 25%, 50%, and 75% correspond to small, medium, and large variability, respectively (Higgins, Thompson, Deeks, & Altman, 2003). Given the large number of studies for some effect

sizes, it is not uncommon for substantial heterogeneity to be observed.

We screened for possible outliers based on the criteria from Hanson and Bussière (1998): (a) an effect size was very small or very large (e.g., ±2 standard deviations from the unweighted mean), (b) the Q statistic was significant, and (c) the outlier contributed to 50% or more of the variance in the value of the Q statistic. Given some of the very large samples and number of studies involved, we found that there were very few true outliers employing these criteria, that is, to the extent that a single study finding would contribute substantially to effect size heterogeneity, which would then be offset by its removal; however, where this was apparent, we reported the results with and without the outlier. Finally, we applied a fail-safe N procedure (Orwin, 1983) to estimate the number of missing studies with a predictive validity correlation of .00 that would be required to bring the observed fixed effect below Cohen's (1988) threshold for a small effect size (r = .10; cf. Poston & Hanson, 2010). We viewed this to be more practical than estimating the number of missing studies required to reduce the effect size to nonsignificance, a value that was markedly higher. We limited these calculations to LS total scores, which tended to feature a larger number of studies and were of particular salience given their focus on the aggregate tool.

Results

Study Search and Sample Description

The study search process followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009) guidelines as presented in Figure 1. The search identified 2,236 records, of which 128 studies met the inclusion criteria across 151 independent samples and 137,931 offenders. Overall, 126 usable documents were obtained (from the years 1981–2012) consisting of 72 published articles, 31 theses/dissertations, 17 government reports, and 6 conference presentations. Most studies were from Canada (k = 55) and the United States (k = 53), followed by Australia (k = 8), United Kingdom (k = 6), Singapore (k = 2), and, finally, Germany, Japan, New Zealand, and Pakistan (k = 1 each).

Overall, 80.5% of the sample was male and 19.5% female. The mean age (unweighted) across the samples was 26.67 years. For studies reporting ethnic composition of their samples (k = 88), approximately 63% of total participants were White, 18.9% Black, 9.8% Aboriginal, 5.5% Hispanic, 2.9% Asian, and 6.5% of other ethnic descent. Seventy-two percent of the samples featured adult populations, whereas 28% were youth. The mean length of follow-up (k = 103) was 26.4 months (SD = 23.8). Weighted mean rates of recidivism were 36% for general (i.e., any; k = 110), 35.2% nonviolent (k = 11), 13.7% violent (k = 34), and 6.5% sexual (k = 9) recidivism.

Comparisons in LS Scores as a Function of Ethnicity and Gender

Within-study group comparisons were conducted on mean LS total score and 11 need domains comparing ethnic minority and nonminority offenders (see top half of Table 2). Ethnic minorities scored significantly higher than nonminorities on most LS areas, including the total score, with a difference of approximately one quarter of a standard deviation (d = .24), a difference that may be classified as small in magnitude. An exception was the personal/emotional need in which nonminorities scored significantly higher. The magnitude of the differences varied, ranging from effect sizes that would be considered small in magnitude (d = −.07 to .30) to approximately moderate for group differences on education/employment (d = .40) and antisocial pattern (d = .50). In short, these results demonstrate that ethnic minorities have higher LS scores than nonminorities as compared within the same samples.

These analyses were repeated comparing males and females. As shown in the bottom half of Table 2, the results were more mixed. Most of the differences could be classified as small in magnitude, with males scoring significantly higher on LS total score, prior offenses, companions, leisure/recreation, substance abuse, antisocial pattern, and attitudes (d = .05 to .38). Female offenders, by contrast, scored significantly higher on education/employment, family/marital, financial, accommodations, and personal/emotional (d = −.08 to −.30).

Predictive Accuracy of the LS Tools: Community and Institutional Outcomes

As seen in Table 3, across 124 samples and 130,833 offenders, LS total scores significantly predicted general community recidivism, the most common outcome examined, with moderate accuracy overall (r_w = .30 and .29 for fixed- and random-effects models, respectively). An estimated 255 missing studies with a predictive validity correlation of 0 would be required to reduce the observed fixed effect below Cohen's (1988) threshold for a small effect size (r = .10).[1] LS total scores also demonstrated significant predictive validities for more specific community recidivism outcomes (e.g., violence), which tended to be examined in a smaller number of studies and were frequently smaller in magnitude than for general recidivism. The LS tools also demonstrated good prediction of institutional recidivism, including any misconduct and serious misconduct. The Q statistics and I^2 values were extremely large for general and violent recidivism, denoting substantial heterogeneity in effect size magnitude across the studies for these outcomes. I^2 values that were moderate in magnitude (43.89–72.52) were observed for the remaining outcomes.

Predictive Accuracy of LS Criminogenic Needs for Various Recidivism Outcomes

The predictive validity findings for the 11 need domains for general, violent, and (for sex offenders) sexual recidivism are presented in Table 4. The pattern of findings paralleled those for the aggregate scale total; that is, predictive validities tended to be higher on average for higher base rate outcomes. Given the large sample sizes involved, the confidence intervals (CIs) were quite

[1] This stands in contrast to the classic fail-safe N, in which case an estimated 171,535 missing studies with an effect size of 0 would be required to reduce this finding to nonsignificance.

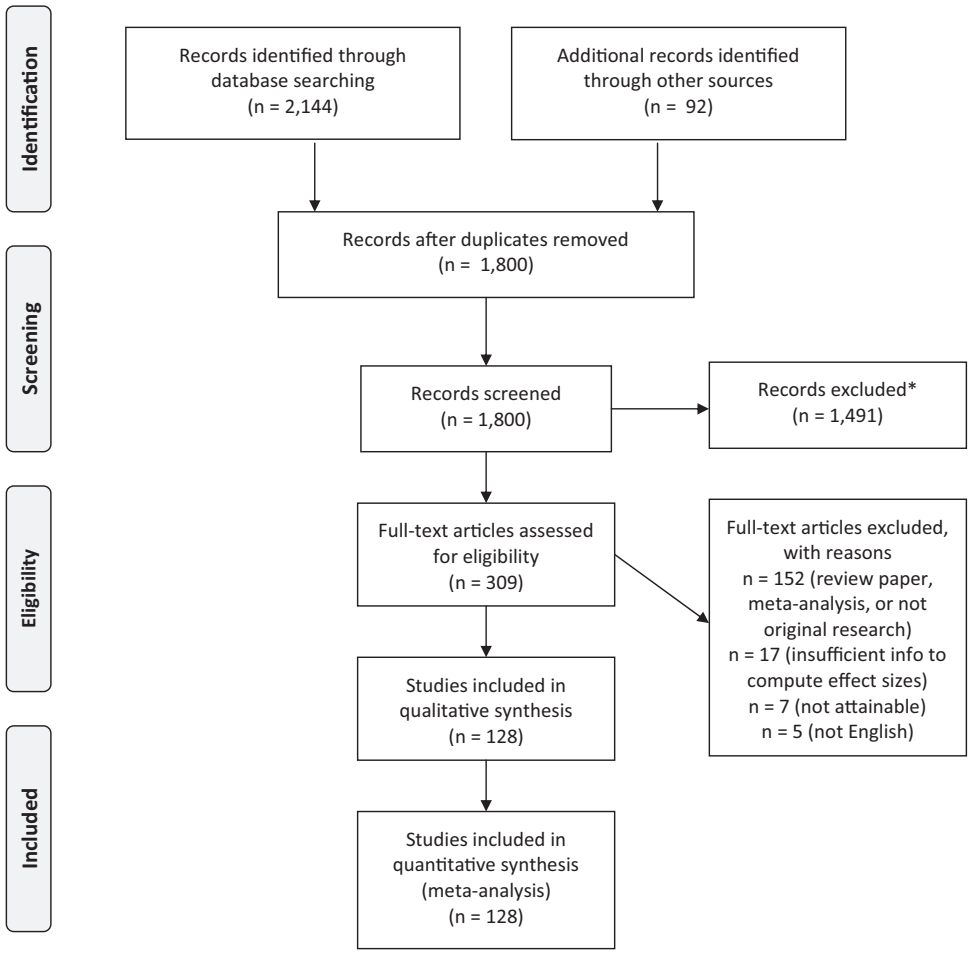

Figure 1. Level of Service meta-analysis PRISMA flow diagram. * Includes 36 records screened and excluded owing primarily to not being in English language, such that their eligibility could not be evaluated further. Adapted from "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement," by D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and The PRISMA Group, 2009, *PLoS Medicine, 6,* e1000097.

narrow, particularly for the fixed-effects analyses; nonoverlapping CIs are interpreted to mean that the effect sizes represent different population parameters. Though all of the needs significantly predicted general recidivism, prior offenses, education/employment, substance abuse, and companions appeared to be particularly strong predictors (r_w = .20 to .33); financial and personal/emotional were the weakest predictors, with their CIs (fixed effects) not overlapping with those of any of the remaining need categories. A similar trend was observed in the prediction of violence. Though all need areas significantly predicted this outcome, prior offenses, antisocial pattern, education/employment, companions, and attitudes had some of the highest predictive accuracies; family/marital, financial, accommodations, and personal/emotional demonstrated weaker prediction and quite small or inconsistent effect sizes between random- and fixed-effects models. Substantial heterogeneity continued to be observed for all effect sizes except for antisocial pattern. Finally, effect sizes for the prediction of sexual recidivism tended to be more similar in magnitude (fixed effects) across the need areas and smaller in magnitude compared to the prediction of other outcomes. Many of these were also not significant when random-effects models were computed. As in previous analyses, personal/emotional emerged as the weakest predictor of sexual recidivism and did not significantly predict this outcome.

Predictive Accuracy for General and Violent Recidivism as a Function of Ethnicity

The predictive validity of the LS total score and need areas was subsequently examined across broad ethnic minority and nonminority offender samples for general and violent recidivism outcomes (see Table 5). LS total scores significantly predicted both sets of recidivism outcomes for both ethnic subgroups. Across the ethnic minority samples, fixed-effects models for general and violent recidivism (r_w = .23 for both) generated significantly smaller effect sizes than among nonminorities (r_w = .32 and.29, respectively), as the CIs did not overlap. Under fail-safe procedures (general recidivism), the estimated number of missing stud-

Table 2
Mean Group Differences on Level of Service (LS) Total Score and Risk–Need Domains as a Function of Ethnicity and Gender

LS domain	*d*	95% CI	Direction	*Q*	I^2	*n*	*k*
			Ethnicity-based comparison				
Total score	.24	[.21, .26]	EMH	1518.82	98.68	74,892	21
Prior offenses	.26	[.23, .28]	EMH	530.02	98.68	33,443	8
Education/employment	.40	[.38, .43]	EMH	486.69	98.36	35,550	9
Family/marital	.29	[.26, .31]	EMH	545.87	98.53	35,550	9
Financial	.24	[.20, .28]	EMH	5.75	65.25	13,158	3
Accommodations	.25	[.22, .29]	EMH	40.03	95.00	13,158	3
Companions	.30	[.27, .32]	EMH	349.38	97.71	35,550	9
Leisure/recreation	.20	[.18, .23]	EMH	79.94	89.99	35,550	9
Substance abuse	.18	[.15, .20]	EMH	1497.70	99.47	35,550	9
Personal/emotional	−.07	[−.10, −.04]	NMH	78.81	93.66	16,200	6
Antisocial pattern	.50	[.46, .53]	EMH	284.18	99.30	19,350	3
Attitudes	.21	[.19, .24]	EMH	105.78	92.44	35,550	9
			Gender-based comparisons				
Total score	.12	[.10, .15]	MH	305.02	93.12	61,551	22
Prior offenses	.38	[.36, .41]	MH	162.15	90.75	47,140	16
Education/employment	−.08	[−.10, −.05]	FH	105.92	84.89	47,646	17
Family/marital	−.21	[−.24, −.19]	FH	127.69	87.47	47,646	17
Financial	−.30	[−.35, −.25]	FH	11.93	49.72	15,546	7
Accommodations	−.14	[−.19, −.09]	FH	42.98	86.04	15,546	7
Companions	.05	[.02, .07]	MH	161.14	90.07	47,646	17
Leisure/recreation	.06	[.04, .09]	MH	267.84	94.03	47,646	17
Substance abuse	.14	[.12, .17]	MH	168.62	90.51	47,646	17
Personal/emotional	−.29	−[.33, −.25]	FH	55.69	78.45	20,458	13
Antisocial pattern	.23	[.20, .27]	MH	16.56	81.89	27,188	4
Attitudes	.19	[.17, .22]	MH	122.80	86.97	47,646	17

Note. All *Q* statistics are significant at $p < .001$ except for financial ($p = .05$ for ethnicity-based comparisons, *ns* for gender-based comparisons). CI = confidence interval; EMH = ethnic minorities higher on a given domain; NMH = nonminorities higher; MH = males higher; FH = females higher.

ies with an effect size of 0 to reduce the observed fixed effects below the threshold of a small effect size was 49 for ethnic minorities and 55 for nonminorities. Substantial effect size heterogeneity (I^2) was observed for the prediction of three out of four outcomes by LS total score; an exception was the prediction of violence in ethnic minority samples.

Each criminogenic need domain significantly predicted general and violent recidivism for both broad ethnic groups, although there was a trend of greater effect size heterogeneity among nonminority samples. Evidence for the Big Four in general emerged, with prior offenses and antisocial pattern being particularly prominent predictors of both outcomes, fol-

Table 3
Prediction of Recidivism Outcomes by Level of Service Measures (Total Score)

	Random		Fixed					
Recidivism criterion	*r*	95% CI	*r*	95% CI	*Q*	I^2	*n*	*k*
				Community outcomes				
General	.29	[.27, .31]	.30	[.29, .30]	2015.40	93.90	130,833	124
Violent	.23	[.19, .27]	.21	[.20, .22]	450.13	91.56	60,997	39
Nonviolent	.25	[.18, .31]	.25	[.21, .29]	27.66	56.62	2,194	13
Sexual	.11	[.03, .18]	.14	[.11, .18]	17.39	65.50	3,163	7
Reincarceration	.32	[.28, .35]	.28	[.26, .29]	61.93	66.09	12,972	22
Technical violation	.27	[.23, .31]	.25	[.24, .27]	55.63	71.24	9,991	17
Halfway house failure	.41	[.30, .51]	.40	[.34, .45]	26.02	73.10	952	8
Offense severity	.27	[.20, .34]	.27	[.24, .30]	32.75	72.52	3,408	10
				Institutional outcomes				
Any misconduct	.24	[.19, .28]	.21	[.18, .24]	26.73	43.89	3,834	16
Serious misconduct	.21	[.14, .28]	.20	[.17, .23]	47.26	70.38	3,474	15

Note. Prediction of sexual recidivism is among sexual offenders only while the prediction of all other recidivism outcomes is across all offender groups. All weighted effect sizes are significant at $p < .001$. All *Q* statistics are significant at $p < .001$ except for nonviolent recidivism and sexual recidivism ($p < .01$) and any institutional misconduct ($p < .05$). CI = confidence interval.

Table 4
Predictive Validity of Level of Service (LS) Criminogenic Needs for General, Violent, and Sexual Recidivism

Recidivism criterion and LS domain	Random		Fixed		Q	I^2	n	k
	r	95% CI	r	95% CI				
General recidivism								
Prior offenses	.28	[.25, .32]	.29	[.28, .29]	1545.31	96.51	97,051	55
Education/employment	.24	[.21, .27]	.22	[.22, .23]	656.69	91.78	97,509	55
Family/marital	.14	[.12, .16]	.13	[.12, .13]	298.65	82.25	97,734	54
Financial	.12	[.09, .15]	.08	[.07, .09]	199.82	85.99	58,714	29
Accommodations	.14	[.11, .16]	.12	[.12, .13]	129.03	77.53	58,832	30
Companions	.22	[.19, .25]	.21	[.21, .22]	846.31	93.27	97,970	58
Leisure/recreation	.16	[.13, .19]	.16	[.16, .17]	623.40	91.66	97,352	53
Substance abuse	.20	[.16, .23]	.20	[.19, .20]	856.77	93.81	97,511	54
Personal/emotional	.14	[.10, .18]	.06	[.05, .07]	705.66	93.77	68,911	45
Antisocial pattern	.31	[.26, .35]	.33	[.32, .34]	17.28*	47.92	28,737	10
Attitudes	.19	[.16, .22]	.17	[.16, .18]	825.23	93.46	97,673	55
Violent recidivism								
Prior offenses	.21	[.16, .27]	.21	[.21, .22]	299.52	94.32	55,044	18
Education/employment	.20	[.15, .24]	.17	[.16, .17]	222.07	91.89	55,417	19
Family/marital	.11	[.09, .14]	.09	[.08, .09]	48.06	62.55	55,452	19
Financial	.09	[.01, .18]	.02	[.01, .04]	14.79**	72.96	23,471	5
Accommodations	.15	[.04, .25]	.07	[.05, .08]	24.29	83.54	23,499	5
Companions	.17	[.11, .22]	.16	[.15, .16]	336.73	94.65	55,440	19
Leisure/recreation	.12	[.08, .16]	.12	[.11, .13]	162.68	88.94	55,450	19
Substance abuse	.13	[.09, .18]	.11	[.10, .12]	221.27	91.87	55,447	19
Personal/emotional	.17	[.09, .25]	.04	[.03, .05]	161.85	93.20	27,503	12
Antisocial pattern	.23	[.22, .24]	.23	[.22, .24]	3.072[a]	0.00	27,944	7
Attitudes	.18	[.14, .21]	.13	[.12, .13]	126.89	85.82	55,433	19
Sexual recidivism								
Prior offenses	.11**	[.03, .20]	.14	[.10, .18]	7.24[a]	44.71	2,389	5
Education/employment	.07[a]	[−.04, .18]	.12	[.08, .16]	11.93*	66.48	2,389	5
Family/marital	.07*	[.01, .14]	.08	[.04, .12]	5.08[a]	21.18	2,389	5
Companions	.04[a]	[−.09, .16]	.12	[.08, .16]	15.62**	74.39	2,389	5
Leisure/recreation	.12	[.08, .16]	.12	[.08, .16]	3.47[a]	0.00	2,389	5
Substance abuse	.00[a]	[−.11, .11]	.06**	[.02, .10]	10.96*	63.49	2,389	5
Personal/emotional	−.02[a]	[−.21, .16]	−.03[a]	[−.12, .06]	11.49**	73.88	484	4
Attitudes	.09[a]	[.00, .18]	.10	[.06, .14]	8.24[a]	51.43	2,389	5

Note. Prediction of sexual recidivism is among sexual offenders only while the prediction of general and violent recidivism is across all offender groups. Unmarked weighted effect sizes (r) and measures of effect size heterogeneity (Q) are significant at $p < .001$. CI = confidence interval.
[a] Not significant.
* $p < .05$. ** $p < .01$.

lowed by companions and attitudes, as well as substance abuse and education/employment from the Central Eight. There were few disparities in the effect size magnitudes between ethnic groups in the much smaller number of studies that examined the criminogenic need domains; an exception may be education/employment, which was a stronger and more consistent predictor of violence in minority samples.

Given the tremendous breadth and complexity of ethnic group membership, as available data permitted, we computed effect size estimates among more specific ethnic groups for LS total score, with the LS scales significantly predicting general recidivism for each ethnic group (see Table 6). There was also generally good consistency between random- and fixed-effects models supporting the stability of the findings. Among Black samples, the removal of a single outlier (a study with a very large sample size and small effect) accounting for nearly two thirds of the effect size heterogeneity improved the fixed-effects estimate.

Predictive Accuracy for General and Violent Recidivism as a Function of Gender

These predictive accuracy analyses were repeated for male and female offender samples across the same set of LS scale components and recidivism outcomes (see Table 7). LS total scores demonstrated strong predictive accuracy, particularly for general recidivism, across male (r_w = .30 and .30) and female (r_w = .35 and .31) samples (fixed and random effects, respectively). The CIs for the fixed-effects model did not overlap, indicating that LS total scores actually predicted general recidivism better in the 45 female samples than in the 80 male samples. LS total scores predicted violent recidivism similarly in 12 female (r_w = .24 and .26) and 30 male (r_w = .28 and .24) samples (fixed and random effects, respectively). Fail-safe procedures estimated that 120 (female samples) and 166 (male samples) missing studies would be required to reduce the ob-

Table 5
Predictive Validity of Level of Service (LS) Total Score and Criminogenic Needs for Violent and General Recidivism as a Function of Ethnicity

	Ethnic minority								Nonminority							
	Random		Fixed						Random		Fixed					
LS domain	*r*	95% CI	*r*	95% CI	*Q*	I^2	*n*	*k*	*r*	95% CI	*r*	95% CI	*Q*	I^2	*n*	*k*
General recidivism																
Total score	.27	[.22, .32]	.23	[.22, .24]	513.00	93.18	25,780	36	.29	[.23, .34]	.32	[.31, .33]	603.79	96.19	40,989	24
Prior offenses	.29	[.21, .36]	.29	[.27, .31]	108.58	88.95	8,120	13	.32	[.24, .39]	.34	[.33, .35]	269.99	96.67	28,323	10
Education/employment	.22	[.18, .27]	.24	[.22, .26]	38.05	65.83	8,627	14	.26	[.20, .32]	.27	[.26, .28]	132.47	93.21	28,323	10
Family/marital	.14	[.11, .16]	.14	[.11, .16]	7.21[a]	0.00	8,294	13	.11	[.08, .14]	.14	[.12, .15]	25.19**	64.26	28,323	10
Financial	.12	[.09, .15]	.12	[.09, .15]	1.99[a]	0.00	3,503	5	.12	[.05, .19]	.16	[.14, .17]	28.85	86.14	10,919	5
Accommodations	.12	[.08, .15]	.12	[.08, .15]	3.67[a]	0.00	3,503	5	.14	[.10, .18]	.15	[.13, .17]	7.78[a]	48.59	10,919	5
Companions	.21	[.16, .27]	.21	[.19, .23]	61.14	80.37	8,295	13	.22	[.15, .28]	.25	[.24, .26]	151.87	94.07	28,323	10
Leisure/recreation	.16	[.10, .21]	.17	[.15, .20]	55.20	78.26	8,295	13	.16	[.10, .22]	.21	[.19, .22]	101.20	91.11	28,323	10
Substance abuse	.22	[.17, .27]	.23	[.20, .25]	52.35	77.08	8,295	13	.22	[.17, .27]	.25	[.23, .26]	80.75	88.85	28,323	10
Personal/emotional	.12	[.02, .22]	.08	[.05, .11]	50.65	84.21	4,146	9	.10	[−.01, .21]	.08	[.06, .09]	139.43	95.70	12,461	7
Antisocial pattern	.29	[.24, .34]	.30	[.28, .33]	6.63[a]	54.74	4,148	4	.27	[.16, .38]	.32	[.30, .33]	4.89[a]	59.06	15,862	3
Attitudes	.19	[.13, .25]	.18	[.16, .20]	62.90	80.92	8,295	13	.19	[.14, .23]	.21	[.19, .22]	54.12	83.37	28,323	10
Violent recidivism																
Total score	.24	[.17, .31]	.23	[.20, .26]	17.02**	70.62	4,178	6	.23	[.10, .35]	.29	[.27, .30]	51.06	92.17	17,416	5
Prior offenses	.23	[.16, .29]	.23	[.20, .26]	11.28*	65.55	4,149	5	.22	[.10, .33]	.27	[.26, .28]	36.12	91.69	17,342	4
Education/employment	.21	[.16, .27]	.21	[.19, .24]	9.78*	59.08	4,150	5	.16	[.10, .21]	.13	[.12, .15]	7.95*	62.42	17,342	4
Family/marital	.08	[.05, .11]	.08	[.05, .11]	3.70[a]	0.00	4,149	5	.09	[.07, .11]	.09	[.08, .11]	3.12[a]	3.90	17,342	4
Companions	.16	[.10, .22]	.17	[.14, .20]	11.18*	64.21	4,150	5	.18	[.07, .29]	.22	[.21, .24]	34.09	91.20	17,342	4
Leisure/recreation	.13	[.10, .16]	.13	[.10, .16]	1.56[a]	0.00	4,150	5	.14	[.07, .21]	.17	[.16, .19]	13.16**	77.19	17,342	4
Substance abuse	.15	[.08, .21]	.12	[.09, .15]	11.93*	66.46	4,150	5	.14	[.07, .21]	.17	[.15, .18]	11.21*	73.23	17,342	4
Personal/emotional	.15	[.07, .23]	.15	[.07, .23]	1.07[a]	0.00	533	3	.15	[−.07, .36]	.09	[.04, .14]	2.11[a]	52.60	1,542	2
Antisocial pattern	—	—	—	—	—	—	—	—	.23	[.22, .25]	.23	[.22, .25]	0.19[a]	0.00	15,800	3
Attitudes	.13	[.10, .16]	.13	[.10, .16]	3.26[a]	0.00	4,150	5	.14	[.07, .20]	.15	[.13, .16]	9.07*	66.94	17,342	4

Note. Unmarked effect sizes (*r*) and *Q* statistics are significant at $p < .001$. Insufficient *k* (<2) for financial and accommodations, as well as part of antisocial pattern (denoted by dash), to compute effect sizes for violence. CI = confidence interval.
[a] Not significant.
* $p < .05$. ** $p < .01$.

served fixed effects (general recidivism) below the threshold of a small effect size.

Some interesting patterns appeared when the individual needs were examined between gender groups. The Big Four were consistently significant and among the strongest predictors of general and violent recidivism across both genders. Education/employment was also consistently strong for both gender groups and across both recidivism outcomes, although there seemed to be a possible gender disparity with a stronger prediction of violence in male samples. In female offender samples, substance abuse and the personal/emotional domain were particularly strong predictors of general recidivism, and the CIs for fixed-effects models did not overlap with those from the male offender samples; this trend was not evident with the smaller number of studies that examined violent recidivism. Effect size heterogeneity also decreased across several of the need areas in the prediction of violence among gender groups.

Table 6
Ethnicity Moderator Analyses: Predictive Validity Effect Sizes of Level of Service Total Score for General Recidivism Among Specific Ethnic Minority Groups

	Random		Fixed					
Group	*r*	95% CI	*r*	95% CI	*Q*	I^2	*n*	*k*
Aboriginal	.30	[.27, .31]	.29	[.27, .31]	61.73	80.56	5,354	13
Asian	.32	[.25, .38]	.31	[.27, .34]	6.69	55.17	2,299	4
Black	.30	[.16, .42]	.18	[.16, .20]	246.11	96.75	10,314	9
Black[a]	.32	[.19, .44]	.33	[.30, .36]	96.58	92.75	3,790	8
Hispanic	.22	[.01, .41]	.26	[.23, .29]	114.85	95.65	3,288	6

Note. All effect sizes significant at $p < .001$, except for Hispanic random effects ($p = .037$). *Q* statistic not significant for Asian group; all remaining *Q* statistics significant at $p < .001$. CI = confidence interval.
[a] Outlier removed.

Within Study Comparisons as a Function of Ethnicity and Gender

Table 8 reports the results of within-study comparisons that examined the predictive accuracy of the LS scales among gender and ethnic subgroups. These are a subset of studies from the larger sample that specifically involve a direct comparison of male–female, ethnic minority–nonminority predictive accuracies from within the same sample and setting. The LS scales significantly predicted violent and general recidivism at magnitudes that were similar to the larger analyses reported above in Tables 5 and 7. Any preexisting disparities in predictive accuracy between demographic subgroups, though still evident, decreased somewhat when

Table 7
Predictive Validity of Level of Service (LS) Total Score and Criminogenic Needs for Violent and General Recidivism as a Function of Gender

	Female								Male							
	Random		Fixed						Random		Fixed					
LS domain	*r*	95% CI	*r*	95% CI	*Q*	I^2	*n*	*k*	*r*	95% CI	*r*	95% CI	*Q*	I^2	*n*	*k*
General recidivism																
Total score	.31	[.26, .35]	.35	[.34, .36]	385.41	88.58	17,802	45	.30	[.27, .34]	.30	[.29, .31]	1,427.11	94.46	77,920	80
Prior offenses	.30	[.24, .36]	.37	[.35, .39]	143.35	88.14	11,212	18	.27	[.22, .33]	.34	[.33, .34]	671.56	94.79	40,776	36
Education/employment	.24	[.19, .28]	.27	[.25, .29]	59.36	69.68	11,249	19	.26	[.22, .29]	.28	[.27, .29]	193.88	81.95	41,072	36
Family/marital	.15	[.11, .18]	.16	[.14, .18]	33.20**	45.78	11,249	19	.15	[.13, .18]	.16	[.15, .17]	87.21	59.87	41,341	36
Financial	.13	[.08, .17]	.12	[.09, .16]	13.15[a]	31.58	2,973	10	.13	[.09, .17]	.15	[.13, .16]	41.15	65.98	15,738	15
Accommodations	.14	[.06, .22]	.15	[.12, .19]	40.55	77.81	2,973	10	.14	[.11, .16]	.14	[.13, .16]	26.11[a]	38.71	15,956	17
Companions	.23	[.17, .28]	.27	[.25, .29]	119.39	84.09	11,317	20	.22	[.18, .26]	.26	[.25, .27]	320.17	88.76	41,326	37
Leisure/recreation	.16	[.11, .21]	.20	[.18, .22]	90.02	80.01	11,249	19	.16	[.13, .20]	.20	[.20, .21]	195.76	83.14	40,876	34
Substance abuse	.25	[.20, .30]	.30	[.28, .32]	92.14	80.46	11,249	19	.19	[.16, .22]	.24	[.23, .25]	177.23	80.25	41,311	36
Personal/emotional	.15[a]	[.04, .26]	.24	[.21, .26]	210.17	93.82	6,168	14	.12	[.08, .16]	.06	[.05, .08]	114.30	77.25	17,596	27
Antisocial pattern	.29	[.26, .31]	.29	[.26, .31]	0.47[a]	0.00	4,930	4	.30	[.25, .34]	.34	[.32, .35]	16.13*	50.40	23,614	9
Attitudes	.20	[.15, .25]	.23	[.21, .25]	96.62	80.34	11,316	20	.19	[.16, .22]	.21	[.20, .22]	124.39	72.67	35,160	35
Violent recidivism																
Total score	.26	[.20, .32]	.24	[.22, .25]	34.11	67.75	8,810	12	.24	[.20, .27]	.28	[.27, .29]	79.76	63.64	28,406	30
Prior offenses	.23	[.16, .30]	.23	[.21, .25]	28.42	82.41	8,269	6	.22	[.13, .31]	.29	[.27, .30]	26.96	77.74	22,654	7
Education/employment	.17	[.12, .22]	.15	[.13, .18]	12.26*	59.22	8,270	6	.24	[.19, .29]	.23	[.22, .25]	8.84[a]	32.14	22,654	7
Family/marital	.10	[.06, .15]	.10	[.08, .13]	10.23[a]	51.12	8,269	6	.12	[.07, .17]	.10	[.09, .11]	9.08[a]	33.91	22,654	7
Companions	.17	[.13, .22]	.16	[.14, .18]	10.97[a]	54.42	8,270	6	.17	[.09, .24]	.23	[.22, .25]	17.4**	65.54	22,654	7
Leisure/recreation	.14	[.09, .19]	.12	[.10, .15]	12.31*	59.37	8,270	6	.15	[.11, .20]	.18	[.17, .19]	7.58[a]	20.83	22,654	7
Substance abuse	.17	[.12, .22]	.15	[.13, .18]	12.21*	59.03	8,270	6	.14	[.08, .19]	.16	[.15, .18]	9.47[a]	36.61	22,654	7
Personal/emotional	.21	[.01, .40]	.17	[.14, .21]	37.78	92.06	3,396	4	.24	[.07, .40]	.22	[.15, .30]	8.68*	76.97	613	3
Antisocial pattern	.17	[.14, .20]	.17	[.14, .20]	0.07[a]	0.00	4,873	2	.22	[.17, .27]	.24	[.23, .25]	3.70[a]	18.95	22,041	4
Attitudes	.19	[.12, .26]	.14	[.12, .16]	25.65	80.51	8,269	6	.17	[.11, .23]	.16	[.15, .17]	12.51[a]	52.05	22,654	7

Note. Unmarked effect sizes (*r*) and *Q* statistics are significant at $p < .001$. Insufficient *k* (<2) for financial and accommodations to compute effect sizes for violence. CI = confidence interval.
[a] Not significant.
* $p < .05$. ** $p < .01$.

sample and setting were controlled in this manner. In light of these findings and for the sake of space and parsimony, we have limited the within-study analyses to the LS total score.

Predictive Accuracy Moderator Analyses by Country and Region

Effect sizes were subsequently aggregated across three broad geographic regions of the study, sample, and setting: Canada, United States, and outside North America. Table 9 presents the results for the prediction of general and violent recidivism by LS total score among these three countries and regions. The results were striking; although LS total score significantly predicted both outcomes in studies across each of the three regions, effect sizes were highest in Canadian samples, followed by studies conducted outside North America, and the smallest were from U.S. samples. This was found for both general and violent recidivism, with the CIs seldom overlapping for either fixed- or random-effects analyses. In short, LS total scores demonstrated significantly stronger prediction of general and violent recidivism in Canadian samples than the other two broad geographic regions, and studies from outside North America demonstrated significantly stronger prediction than studies from U.S. samples. The *Q* and I^2 statistics decreased considerably in magnitude, particularly in the prediction of violence, demonstrating the country or geographic region of origin for the study to be a potentially important source of effect size variability. We also applied fail-safe *N* procedures to estimate the number of missing studies with an effect size of 0 required to reduce these findings for the prediction of general recidivism by LS total score below the threshold of a small effect size as follows: Canadian samples ($n = 181$), United States ($n = 65$), and outside North America ($n = 39$).

These analyses were repeated for the prediction of general recidivism by the 10 LS domains. (Insufficient *k* existed among studies to conduct moderator analyses for the prediction of violence by the individual domains). As illustrated in Table 10, the same trends emerged as with prediction by LS total score demonstrating geographic region to be an important source of effect size variability. First, all need domains predicted general recidivism across each of the three geographic regions. Second, effect size magnitudes across all domains were highest for Canadian studies, followed by those outside North America, and lastly U.S. samples. Third, there was considerable range in effect size magnitude among the individual domains within samples from Canada ($r_w = .22$ to .41), United States ($r_w = .02$ to .20), and outside North America ($r_w = .09$ to .27). The strongest and most consistent predictors were prior offenses, education/employment, and companions across the three regions. Effect size heteroge-

Table 8
Within-Study Comparisons of Level of Service Predictive Accuracy for General and Violent Recidivism Among Gender and Ethnic Groups

Group	Random		Fixed		Q	I^2	n	k
	r	95% CI	r	95% CI				
General recidivism								
Ethnic minority	.29	[.23, .34]	.23	[.22, .24]	477.45	94.14	22,996	29
Nonminority	.28	[.23, .34]	.32	[.31, .33]	599.63	96.33	40,835	23
Female	.29	[.24, .34]	.32	[.31, .34]	186.75	83.94	11,805	31
Male	.29	[.24, .35]	.30	[.29, .31]	1294.51	97.68	58,472	31
Violent recidivism								
Ethnic minority	.24	[.17, .31]	.23	[.20, .26]	17.02**	70.62	4,178	6
Nonminority	.21	[.06, .34]	.29	[.27, .30]	50.96	94.11	17,262	4
Female	.25	[.22, .27]	.25	[.22, .27]	1.98[a]	0.00	5,257	6
Male	.29	[.21, .35]	.29	[.28, .31]	15.67**	68.10	22,760	6

Note. Unmarked weighted effect sizes (r) and measures of effect size heterogeneity (Q) are significant at $p < .001$ except as noted. The ks for ethnic minority and nonminority within-study comparisons are uneven, since frequently more than one ethnic minority subgroup was analyzed in a given study and individual effect sizes were computed for each (k representing the number of samples within a given study). CI = confidence interval.
[a] Not significant.
** $p < .01$.

neity also decreased among the criminogenic need domains when examined by region. This was most apparent among Canadian samples, which demonstrated small effect size heterogeneity ($I^2 = 0.00$ to 49.22) for six domains, in contrast to U.S. samples, which continued to demonstrate substantial heterogeneity across nine of the 10 domains ($I^2 = 75.18$ to 95.03).

Finally, we computed effect sizes for LS total score prediction of general recidivism among gender and ethnic groups among the three broad geographic regions (see Table 11). LS total scores continued to predict outcome irrespective of geographic region or demographic subgroup, but effect sizes were highest in Canadian samples and lowest in U.S. samples. The CIs demonstrated minimal overlap, and all weighted effect sizes were significantly different at $p < .001$ between the three geographic regions. The impact of aggregating effect sizes by region on effect size heterogeneity among the demographic subgroups was mixed; among U.S. samples, small to moderate effect size heterogeneity was observed across 22 female offender samples ($I^2 = 39.73$), whereas substantial heterogeneity was observed for effect sizes aggregated across the other three demographic groups. Among Canadian samples, the least heterogeneity was observed among ethnic minority offenders ($I^2 = 24.47$), with moderate to high heterogeneity among remaining demographic groups.

To further elucidate possible sources of variation, we correlated the unweighted effect size with author allegiance, coded dichotomously as to whether a study's authors included an LS scale developer or a student of an author versus no affiliation to the LS scales. Allegiance was significantly correlated with effect size magnitude for general recidivism ($r = .42$, $p < .001$); however, when examined exclusively among Canadian studies (all LS authors are Canadian), allegiance was not significantly correlated with effect size ($r = .04$, $p = .765$), suggesting that the allegiance effect may be an artifact of the regional differences observed earlier.

Table 9
Predictive Validity of Level of Service Total Score for General and Violent Recidivism by Country/Region

Country/region	Random		Fixed		Q	I^2	n	k
	r	95% CI	r	95% CI				
General recidivism								
Canada	.38	[.35, .41]	.43	[.42, .44]	186.77	73.23	39,688	51
United States	.20	[.18, .23]	.22	[.21, .23]	432.66	88.21	70,428	52
Outside North America	.30	[.28, .33]	.29	[.28, .31]	48.68	63.02	20,581	19
Violent recidivism								
Canada	.26	[.23, .29]	.27	[.26, .28]	77.90	65.34	35,338	28
United States	.12	[.11, .13]	.12	[.11, .13]	3.67[a]	0.00	24,644	7
Outside North America	.20	[.14, .26]	.20	[.14, .26]	1.63[a]	0.00	1,015	4

Note. Unmarked weighted effect sizes (r) and measures of effect size heterogeneity (Q) are significant at $p < .001$ except as noted. CI = confidence interval.
[a] Not significant.

Table 10
Predictive Validity of Level of Service (LS) Criminogenic Needs for General Recidivism by Country/Region

LS domain	Country/region	Random		Fixed		Q	I^2	n	k
		r	95% CI	r	95% CI				
Prior offenses	Canada	.36	[.33, .40]	.41	[.40, .42]	101.11	74.29	34,090	27
	United States	.19	[.13, .24]	.20	[.20, .21]	301.59	95.03	47,420	16
	Outside NA	.27	[.22, .31]	.25	[.24, .27]	45.34	75.74	15,540	12
Education/employment	Canada	.30	[.27, .33]	.31	[.30, .32]	65.23	60.14	34,052	27
	United States	.18	[.15, .21]	.17	[.16, .18]	119.53	86.61	47,697	17
	Outside NA	.21	[.18, .24]	.22	[.20, .23]	20.89*	52.13	15,759	11
Family/marital	Canada	.18	[.16, .19]	.18	[.16, .19]	23.57[a]	0.00	34,428	27
	United States	.09	[.06, .12]	.08	[.07, .09]	68.32	78.04	47,547	16
	Outside NA	.15	[.11, .18]	.15	[.13, .16]	24.79**	59.66	15,759	11
Financial	Canada	.19	[.14, .23]	.19	[.14, .23]	3.27[a]	0.00	1,684	10
	United States	.08	[.05, .10]	.06	[.05, .07]	48.34	75.18	45,667	13
	Outside NA	.15	[.10, .19]	.17	[.15, .19]	14.93*	66.52	11,363	6
Accommodations	Canada	.22	[.16, .29]	.23	[.19, .28]	20.54*	46.45	1,902	12
	United States	.11	[.08, .14]	.11	[.10, .12]	58.87	81.31	45,567	12
	Outside NA	.13	[.09, .17]	.15	[.13, .17]	10.91[a]	54.18	11,363	6
Companions	Canada	.30	[.27, .32]	.32	[.31, .33]	43.14*	35.09	34,408	29
	United States	.15	[.11, .18]	.15	[.14, .16]	114.59	85.17	47,797	18
	Outside NA	.19	[.16, .23]	.19	[.17, .20]	27.50**	63.64	15,759	11
Leisure/recreation	Canada	.25	[.24, .26]	.25	[.24, .26]	21.11[a]	0.00	33,896	25
	United States	.09	[.06, .12]	.10	[.09, .11]	91.82	82.58	47,697	17
	Outside NA	.14	[.10, .18]	.15	[.13, .16]	30.45	67.16	15,759	11
Substance abuse	Canada	.25	[.21, .28]	.30	[.29, .30]	77.35	67.68	34,055	26
	United States	.15	[.11, .18]	.13	[.12, .14]	118.26	86.47	47,697	17
	Outside NA	.18	[.14, .23]	.18	[.17, .20]	41.58	75.95	15,759	11
Personal/emotional	Canada	.24	[.17, .31]	.31	[.29, .33]	88.96	82.01	5,824	17
	United States	.04	[.02, .06]	.02	[.01, .03]	48.41	69.01	47,547	16
	Outside NA	.13	[.07, .19]	.09	[.07, .10]	88.73	87.60	15,540	12
Attitudes	Canada	.27	[.24, .30]	.26	[.25, .27]	51.20**	49.22	34,217	27
	United States	.12	[.09, .16]	.12	[.11, .13]	141.92	88.73	47,697	17
	Outside NA	.16	[.13, .19]	.17	[.16, .19]	20.73*	51.77	15,759	11

Note. Unmarked weighted effect sizes (r) and Q statistics are significant at $p < .001$ except as noted. NA = North America; CI = confidence interval.
[a] Not significant.
* $p < .05$. ** $p < .01$.

Predictive Accuracy of LS Variants

The final set of analyses examined sources of effect size variability among different versions of the family of LS tools (see Table 12). Where sufficient k permitted, effect sizes were aggregated by geographic region. Strong predictive accuracy for general and violent recidivism was observed across the variants of the LS scales. The LSI–R and YLS/CMI each had the largest number of studies across the three geographic regions. Although the LSI–R demonstrated the smallest effect size overall for general recidivism, when this was aggregated by geographic region, the largest effect size was observed for Canadian samples at a magnitude consistent with other LS variants, followed by studies outside North America and the U.S. samples. Significant differences in LSI–R effect size magnitude were found between each of the three regions for general and violent recidivism; the lone exception was nonsignificant differences in the prediction of violence between Canadian samples and those from outside North America ($z = 1.67$, $p = .095$, random effects).

Similar trends were found for the YLS/CMI. The largest effect sizes in the prediction of general recidivism were found for Canadian samples and those outside North America, both of which had significantly higher effect sizes than U.S. samples ($p < .001$). For the prediction of violence, although Canadian samples again had the largest effect size magnitude, these were not significantly different from U.S. and outside North American regions. Effect size heterogeneity also decreased markedly among specific LS variants, particularly when examined by geographic region. This pattern seemed most evident for the YLS/CMI ($I^2 = 0.00$ to 64.41).

Discussion

We conducted the largest known meta-analysis to date of the family of LS risk assessment tools. Overall, 128 studies consisting of 151 independent samples from nine countries and 137,931 offenders were included in this review. This is approximately 3 times larger than one published by Smith et al. (2009), both in terms of number of samples and number of participants, and even more than the important contributions by Gendreau et al. (2002) and M. A. Campbell, French, and Gendreau (2009). The very large number of studies and the international nature of this study speak to substantial diversity of the samples included with respect to gender, culture/ethnicity, setting, and age among other factors that allowed for examination of important moderator variables informed by ongoing controversies in the extant literature.

Table 11
Predictive Validity of Level of Service Total Score for General Recidivism by Country/Region as a Function of Gender and Ethnicity

	Random		Fixed					
Country/region	*r*	95% CI	*r*	95% CI	*Q*	I^2	*n*	*k*
Male								
Canada	.36	[.33, .39]	.42	[.41, .43]	117.57	65.13	29,585	42
United States	.21	[.18, .25]	.18	[.17, .19]	211.49	88.65	31,473	24
Outside NA	.31	[.28, .34]	.30	[.29, .31]	27.52**	56.39	16,862	13
Female								
Canada	.44	[.39, .48]	.45	[.43, .46]	63.23	76.28	9,670	16
United States	.22	[.19, .26]	.21	[.19, .24]	34.84*	39.73	6,037	22
Outside NA	.29	[.22, .36]	.27	[.23, .31]	15.44*	61.15	2,095	7
Ethnic minority								
Canada	.41	[.36, .46]	.40	[.38, .43]	26.96**	59.20	5,101	12
Canada[a]	.40	[.38, .42]	.40	[.36, .43]	13.24[b]	24.47	5,061	11
United States	.18	[.11, .25]	.16	[.14, .17]	187.91	91.49	16,308	17
Outside NA	.26	[.21, .32]	.26	[.24, .28]	25.18	76.18	6,234	7
Nonminority								
Canada	.41	[.36, .47]	.42	[.41, .43]	34.68	74.05	1,844	10
United States	.21	[.17, .25]	.19	[.17, .20]	45.75	78.14	13,068	11
Outside NA	.28	[.24, .32]	.30	[.28, .31]	5.24[b]	61.86	9,751	3

Note. Unmarked weighted effect sizes (*r*) and *Q* statistics are significant at $p < .001$ except where noted. NA = North America; CI = confidence interval.
[a] Outlier removed. [b] Not significant.
* $p < .05$. ** $p < .01$.

LS Profile and Score Differences as a Function of Ethnicity and Gender

Mean comparisons demonstrated that ethnic minorities scored significantly higher than nonminorities on LS total score and all but one criminogenic need domain on the tool. However, the magnitude of these differences may be considered small in magnitude, with one exception being antisocial pattern, which was closer to medium. These meta-analytic results convincingly demonstrate what has been found in some (e.g., Holsinger, Lowenkamp, & Latessa, 2003), but not all (e.g., Bonta, 1989), studies of minority offenders and implied in systematic reviews (e.g., Rugge, 2006); the present results indicate that such conclusions also extend to the criminogenic need domains of the LS scales. It is important to bear in mind, however, that there are important social, historical, and contextual factors that may contribute to elevated risk scores and increase the possibility of ethnic minorities coming into contact with the justice system (Mann, 2010; Rugge, 2006).

Table 12
Predictive Validity of Level of Service (LS) Variants for General and Violent Recidivism: Overall and by Country/Region

	General recidivism								Violent recidivism							
	Random		Fixed						Random		Fixed					
LS variant	*r*	95% CI	*r*	95% CI	*Q*	I^2	*n*	*k*	*r*	95% CI	*r*	95% CI	*Q*	I^2	*n*	*k*
LSI	.32	[.27, .37]	.32	[.29, .36]	39.25**	51.59	2,934	20	.21	[.15, .28]	.21	[.15, .28]	0.57	0.00	833	2
LS/CMI/LSI–OR	.42	[.38, .47]	.44	[.43, .45]	55.99***	80.36	31,932	12	.27	[.22, .32]	.28	[.27, .29]	54.75***	81.74	31,427	11
LSI–SV	.27	[.20, .33]	.28	[.25, .32]	6.95	56.80	2,518	4	—	—	—	—	—	—	—	—
LSI–SR	.38	[.27, .48]	.42	[.37, .47]	14.27**	71.96	1,163	5	.28	[.15, .40]	.29	[.19, .38]	1.38	0.00	367	2
LSI–R (overall)	.25	[.22, .28]	.24	[.23, .25]	632.41***	91.46	78,505	55	.23	[.16, .28]	.13	[.12, .15]	78.50***	83.44	26,172	14
LSI–R (Canada)	.41	[.30, .52]	.41	[.38, .45]	55.02***	61.43	1,998	8	.31	[.23, .38]	.33	[.28, .37]	11.40	47.35	1,378	7
LSI–R (United States)	.20	[.17, .23]	.22	[.21, .23]	403.66***	91.33	60,998	36	.12	[.11, .13]	.12	[.11, .13]	1.28	0.00	24,279	5
LSI–R (Outside NA)	.29	[.26, .32]	.29	[.28, .31]	25.93**	61.43	15,509	11	.23	[.15, .31]	.23	[.15, .31]	0.33	0.00	515	2
YLS/CMI (overall)	.28	[.25, .31]	.25	[.24, .27]	82.72***	64.94	15,447	30	.23	[.18, .27]	.22	[.18, .25]	19.46	38.35	2,916	13
YLS/CMI (Canada)	.34	[.29, .38]	.33	[.29, .37]	16.15	31.87	2,514	12	.25	[.19, .32]	.24	[.20, .28]	15.96*	49.87	2,051	9
YLS/CMI (United States)	.22	[.19, .25]	.22	[.20, .24]	15.80	36.73	8,367	11	.19	[.09, .29]	.19	[.09, .29]	0.53	0.00	365	2
YLS/CMI (Outside NA)	.33	[.26, .40]	.28	[.25, .31]	16.86**	64.41	4,566	7	.16	[.07, .25]	.16	[.07, .25]	0.00	0.00	500	2

Note. All weighted effect sizes (*r*) are significant at $p < .001$. Dashes denote insufficient k (<2) to compute effect sizes. LSI = Level of Service Inventory; LS/CMI = Level of Service/Case Management Inventory; LSI–OR = Level of Service Inventory–Ontario Revision; LSI–SV = Level of Service Inventory–Screening Version; LSI–SR = Level of Service–Self-Report; LSI–R = Level of Service Inventory–Revised; YLS/CMI = Youth Level of Service Inventory/Case Management Inventory; NA = North America; CI = confidence interval.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Male and female comparisons on the tool demonstrated more mixed findings. Males, on the one hand, tended to have slightly higher LS total scores overall, more serious offense history, and pervasive patterns of antisocial behavior, followed by marginally higher scores on areas denoting concerns with antisocial peers, lack of prosocial leisure activities, and substance abuse concerns linked to crime. Females, by contrast, had markedly more serious personal/emotional concerns, financial problems, and family/marital difficulties and a significant but smaller set of effects indicating greater accommodation and education/employment concerns. The findings are consistent with assertions about salient areas of risk and need for female offenders, given their unique circumstances and possible gendered pathways to crime (e.g., Reisig, Holtfreter, & Morash, 2006). As current perspectives of female criminality focus on victimization and its psychological sequelae—domestic relationships, dependency, and social location (e.g., Bjerregaard & Smith, 1993; Bloom, Owen, Covington, & Raeder, 2002; Salisbury & Van Voorhis, 2009; Wright, Salisbury, & Van Voorhis, 2007)—it is not surprising to find female offenders scoring higher on the above-noted domains. We believe that these findings are instructive and may facilitate correctional planning and program development as they identify criminogenic needs that are particularly prevalent among women offenders (Blanchette & Brown, 2006; Hannah-Moffat, 2009; Holtfreter & Cupp, 2007). The finding that the women in these samples did not score as highly as men on the substance abuse domain may come as a surprise given both theoretical arguments (Bloom et al., 2002; Covington & Bloom, 2007) and empirical findings (McClellan, Farabee, & Crouch, 1997; Salisbury & Van Voorhis, 2009) indicating that substance abuse plays a critical role in the pathway to female criminality. However, the strength of the relationship between substance abuse and recidivism is quite another matter and must also be considered (see Gender and Predictive Accuracy).

In considering these findings, it is important to note that a difference on total risk score or on any criminogenic need domain by ethnicity or gender does not, in itself, bias the instrument against the high scoring minority group, as some have suggested or implied (LaPrairie, 1995; Martel et al., 2011). Rather, it is important to determine whether these differences in risk correspond to differences in outcome in that higher mean risk scores should correspond with higher recidivism rates and whether the linear relationship between risk and outcome remains comparable across these groups. Indeed, numerous studies have found that ethnic minorities have higher recidivism rates than White offenders (e.g., Wormith & Hogg, 2012), in which case they should, on average, score more highly on any valid risk assessment tool. From a prevention perspective, as discussed above, it is also important to determine what may have caused these differences in the first place.

Predictive Accuracy for General and Violent Recidivism

The LS tools significantly predicted all recidivism outcomes from a range of criterion variables; higher predictive accuracy tended to be observed for higher base rate outcomes, and this declined somewhat as criterion operationalizations narrowed. The direction and magnitude of the effect sizes were broadly consistent with past research, although the present study generated somewhat smaller effect sizes overall compared to other larger scale LS meta-analyses (e.g., M. A. Campbell et al., 2009; Gendreau et al., 2002). It is worth noting, however, that the disparity is quite small, but perhaps more importantly, these investigations contained a much higher proportion of Canadian studies; in Gendreau et al. (2002), for instance, all effect sizes for the prediction of violence came from Canadian studies, as did all but five effect sizes for the prediction of general recidivism. As we found in the present investigation, geographic region was a potent source of effect size variation; scrutiny of the effect sizes for the Canadian-based studies for general and violent recidivism bore a very high level of consistency with M. A. Campbell et al. (2009) and Gendreau et al. Comparisons to findings from Smith et al. (2009) are more complex. Their mean effect size across all of their samples of female offenders was higher than the effect size found in the current study for general recidivism; however, when considering only those studies that included both male and female offenders, effect sizes fell below those of the current study for both males and females, as noted in Table 1. Although Smith et al. did not analyze their data by region, it is quite possible that their decrement in effect size for female offenders was related to country, given that more than one third of their women offender samples were Canadian. Smith et al. also focused on LSI–R and included no male-only studies.

The predictive accuracy of LS criminogenic need areas varied considerably and raises questions about the appropriateness of including the 11 domains that are represented across the multiple versions of the LS. Although all domains were significant in the fixed-effects analyses, two in particular (financial and personal/emotional) were significantly less so than the others. This finding is consistent with Andrews and Bonta's (1995a, 1995b, 2010) characterization of the Central Eight risk–need domains and is reflected in more recent versions of the instrument, particularly the LS/CMI and the YLS/CMI. The addition of antisocial pattern to these versions of LS is supported by the large effect sizes it generated in all analyses. Our findings also provide support for the prominence of criminal history and antisocial pattern, two of the Big Four (Andrews & Bonta, 2010), but not criminal attitudes and criminal companions, which raises some question about the two tiers of risk–need domains as measured by the LS tools.

Ethnicity and Predictive Accuracy

In line with previous analyses across all samples, the family of LS tools and its individual need domains predicted general and violent recidivism among both broad and specific ethnic minority and nonminority groups. One notable difference was the lower predictive accuracy of LS total scores observed with the ethnic minority samples in fixed-effects models, although such differences decreased with random-effects models, particularly for violence. That is, the weighted effect sizes were significantly larger for nonminorities within the studies sampled (fixed effects), but were closer in magnitude when approximating the unweighted average and generalizing to the total population of studies (random effects). The results for the criminogenic need domains, particularly the Central Eight, are consistent with past research findings supporting the validity of these domains in international samples of ethnic minorities, as well as demonstrating considerable effect size heterogeneity (Gutierrez et al. 2013). As with Gutierrez et al.'s

(2013) comparisons of Aboriginal and non-Aboriginal offenders, the effect sizes for general recidivism were slightly larger for most of the domains for nonminority offenders, with less consistent discrepancies between ethnic groups in effect sizes for violence.

There were not sufficient samples to examine predictive accuracies for specific needs among ethnic and nonminorities among the three geographic regions or to do this for violence. When this was examined for LS total score and general recidivism, effect size variability decreased, as did the magnitude of differences in effect size between minority and nonminority groups. This was perhaps most evident for Canadian samples, which had very little effect size variability among ethnic minority samples, and for which the effect sizes magnitudes were negligible from nonminority samples. Substantial heterogeneity continued to exist in U.S. minority and nonminority samples for the prediction of general recidivism. One possibility may be that systemic bias within the justice system may distort the measurement of "true" recidivism, thus reducing the association between LS scores and outcome. The results of ethnicity moderator analyses, however, generally support the use of the LS tools for assessing recidivism risk among ethnic minority and nonminority samples, which was buttressed by the predictive validity demonstrated with specific ethnic groups.

Gender and Predictive Accuracy

The LS tools predicted general recidivism among female offenders at a broadly comparable magnitude to past research (Smith et al., 2009), and importantly, the predictive accuracy of the LS total score was very similar for males and females, particularly for random-effects models. Admittedly, there continued to be a substantial amount of heterogeneity among effect sizes for both gender groups, although this decreased somewhat as additional moderators were examined (e.g., geographic region). The LS domains each significantly predicted violent and general recidivism among both genders, and there tended to be few substantive differences in effect size magnitude; however, the domains of substance abuse and personal/emotional had significantly larger effect size magnitudes for females in the prediction of general recidivism. These results are consistent with assertions from proponents of gender-informed models of criminal behavior about the salience of certain risk–need domains for women, such as problems with substance abuse and personal/emotional well-being (e.g., Van Voorhis et al., 2010) and extend past findings by Andrews et al. (2012) pertaining to the significantly stronger impact of substance abuse on female recidivism. Although women as a whole may not score as highly as men on substance abuse, when they do it is particularly problematic. In short, the results support the predictive efficacy of the LS tools among female and male offenders for violent and general recidivism. There is little evidence to suggest that the instrument, overall, is better suited for, or performs better for, either gender group, at least in terms of recidivism prediction. In considering the individual need domains, our analyses suggest that some areas such as personal/emotional concerns and substance abuse difficulties may have special relevance for female offenders.

Regional Differences in Predictive Accuracy: An Important Source of Effect Size Variation

Previous meta-analytic research (Olver et al., 2009) has demonstrated risk assessment tools, many of which have Canadian origins, to have higher predictive validity in Canadian samples compared to other jurisdictions. We examined geographic region as a moderator and found that the largest effect sizes were observed, almost without exception, in Canadian samples, followed by those outside North America, and with U.S. samples demonstrating the lowest effect sizes. It is important to underscore that the LS scales and their risk–need domains still predicted all recidivism outcomes irrespective of geographic region; however, the consistent discrepancies observed in effect size magnitude should not be ignored, especially given that the confidence intervals seldom overlapped. Effect size heterogeneity also decreased noticeably in the regional analyses, particularly as other moderators were added, adding further weight to the importance of geographic region as an important source of variation. Interestingly, the U.S. studies also often demonstrated the highest effect size variability (I^2 values), and this often was not substantively lower than the values observed in broader aggregate analyses.

What might account for these regionally based discrepancies in effect sizes? First, predictive validity coefficients depend on three fundamental concepts: the nature of the true relationship and the precision of both the assessment and outcome measures. There are various sources of error in both measures, both of which could vary by region. The source documents in this meta-analysis did not routinely provide any reliability statistics besides alpha levels, and few offered any commentary about LS training, mean time to complete the LS, or quality assurance of scale administration in the field. Although one might speculate about the impact of very large caseloads as reported by some U.S. jurisdictions on the quality of LS assessment, more detailed data collection and analysis are required to determine whether there are systematic sources of error in the assessment protocol by country. Similarly, we and others (Andrews et al., 2011; Yang et al., 2010) have speculated about sources of systematic variation in measurement precision on the outcome measure, noting that Canadian researchers typically have access to a national database of offender criminal records designed to capture all offenders' offending anywhere in the country. More accurate measures of the outcome criterion will routinely generate higher estimates of predictive validity.

Second, the LS scales are Canadian developed and have been exported to other countries that have important cultural differences. Although cultural differences may account for some of the regional differences observed in predictive accuracy, we do not believe this is the primary source of such differences, partly because the Central Eight domains are found across cultures (Andrews & Bonta, 2010; Gutierrez et al., 2013) and partly because the LS authors have gone to considerable length working with international agencies translating items and operationalizing concepts to bring them in line with cultures and criminal codes around the globe. One possibility may be a difference in familiarity with risk assessment in Canada, compared to elsewhere, particularly when studies were conducted. For instance, this may include a longer history of use of the LS scales in Canada as well as ready access to the instrument developers and frequent training opportunities. A related possibility in non-Canadian jurisdictions may be rater drift; it is possible that when using the tool in the field, individuals using the tool may diverge from rating rules. Many of the U.S. studies were prospective examinations of the tool, rated by parole and probation

officers on hundreds or even thousands of youth or adult offenders; beset as these studies are with large caseloads and tight deadlines, it is possible that such circumstances may also serve to reduce rater accuracy. We were, however, able to determine that regional differences were not due to a possible confound of author affiliation with the LS, which was far more prevalent in Canadian studies. The number of plausible reasons for regional effect size variation and the fact that the LS scales continued to demonstrate significant predictive validity, despite realistic impediments to rater accuracy, would, in our view, support the continued use of the LS scales by U.S. forensic and correctional evaluators in legal proceedings.

Conclusions, Limitations, and Future Directions

The present study has some important strengths, limitations, and implications for future research avenues of the LS scales. Although the current investigation is the largest examination of the LS scales to date, we note that much of the interest in the LS is both international and agency (nonacademic) based. Therefore, it is quite possible that we have missed pertinent foreign language studies and other "gray literature." Perhaps the most pressing limitation of the present study is that relatively few investigations reported their results combining the domains of gender and ethnicity (e.g., female ethnic minorities); the norm, rather, was to report effect sizes as a function of one broad demographic group or another. We were surprised to see several studies with a sizable sample, male or female, that did not examine findings in light of ethnicity or even report the frequency of this characteristic in their samples. This may be redressed in future research as the volume of studies continues to grow, permitting more nuanced examinations of gender, ethnicity, and other possible effect size moderators.

As warned by Lipsey (2003), we are also cognizant of the possibility of confounded moderators. This includes the possibility of region being confounded by assessor training or caseload size, LS experience, agency quality assurance mechanisms, author affiliation, and precision of the outcome measure. It also includes ethnicity possibly being confounded by region, or gender being confounded by type of agency and the variations in practice that might be specific to women offenders, to mention only a few. Finally, the current examination was also limited to a rather straightforward examination of predictive validity and did not attempt to explore the application LS total and need scores to offender case management and treatment in accordance with third- and fourth-generation risk assessment principles (Andrews et al., 2006). Relatively few LS risk assessment studies have incorporated appropriate intervention into their analysis of outcome (e.g., Bonta et al., 2011; Luong & Wormith, 2011).

These limitations notwithstanding, the present meta-analysis is a large-scale examination of the LS scales across 30 years of published and unpublished research. The volume of studies permitted international comparisons and important, albeit broad-based, comparisons in LS scores and predictive accuracy among special subgroups, variants of the LS scales, and the criminogenic need domains, such as the Central Eight. These considerations would suggest that the present findings are representative of a key psychometric property for which this family of tools are most frequently applied—their criterion-related validity for future recidivism. The results also support the consolidation of the LS scales into the Central Eight domains as represented in the most recent versions of the instrument, the LS/CMI and the YLS/CMI. They do, however, raise some question about the primacy and universality of the Big Four as promoted by Andrews and Bonta (2010), at least as measured by the LS tools.

As with any tool, caution and discretion are recommended with professional applications of the LS scales, particularly with vulnerable populations for whom other circumstances exist that may have brought them into contact with the justice system and that may inform case management and service delivery to reduce risk and prevent recidivism. In turn, ongoing training and supervision in the use of the tool may help promote high-quality administrations in the field to ensure fair, valid, and effective applications of the LS scales.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Andrews, D. A. (1982). *The Level of Supervision Inventory: The first follow-up*. Toronto, Canada: Ontario Ministry of Correctional Services.

Andrews, D. A., & Bonta, J. (1994). *The psychology of criminal conduct*. Cincinnati, OH: Anderson.

Andrews, D. A., & Bonta, J. (1995a). *Level of Service Inventory–Revised (LSI–R): An offender assessment system: User's guide*. Toronto, Canada: Multi-Health Systems.

Andrews, D. A., & Bonta, J. (1995b). *Level of Service Inventory–Revised: Screening Version (LSI–R:SV): User's manual*. Toronto, Canada: Multi-Health Systems.

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). New Providence, NJ: LexisNexis.

Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior, 17,* 19–52. doi:10.1177/0093854890017001004

Andrews, D. A., Bonta, J., & Wormith, J. S. (1995). *Level of Service Inventory–Ontario Revision (LSI–OR): Interview and scoring guide*. Toronto, Canada: Ontario Ministry of the Solicitor General and Correctional Services.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2001). *Level of Service Inventory–Saskatchewan Youth Edition*. Toronto, Canada: Multi-Health Systems.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2004). *Level of Service/Case Management Inventory (LS/CMI): An offender assessment system: User's guide*. Toronto, Canada: Multi-Health Systems.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or needs assessment. *Crime & Delinquency, 52,* 7–27. doi:10.1177/0011128705281756

Andrews, D. A., Bonta, J., Wormith, S. J., Guzzo, L., Brews, A., Rettinger, J., & Rowe, R. (2011). Sources of variability in estimates of predictive validity: A specification with level of service risk and need. *Criminal Justice and Behavior, 38,* 413–432. doi:10.1177/0093854811401990

Andrews, D. A., Guzzo, L., Raynor, P., Rowe, R. C., Rettinger, J., Brews, A., & Wormith, J. S. (2012). Are the major risk/need factors predictive of both female and male reoffending? A test with the eight domains of the Level of Service/Case Management Inventory. *International Journal of Offender Therapy and Comparative Criminology, 56,* 113–133. doi:10.1177/0306624X10395716

*Andrews, D. A., Kiessling, J. J., Robinson, D., & Mickus, S. (1986). The risk principle of case classification: An outcome evaluation with young adult probationers. *Canadian Journal of Criminology, 28,* 377–384.

*Andrews, D. A., & Robinson, D. (1984). *The Level of Supervision Inventory: Second report* [Report to Research Services, Ontario Ministry of Correctional Services]. Unpublished manuscript.

Andrews, D. A., Robinson, D., & Hoge, R. D. (1984). *Manual for the Youth Level of Service Inventory*. Ottawa, Canada: Department of Psychology, Carleton University.

*Arnold, T. K. (2007). *Dynamic changes in Level of Service Inventory–Revised (LSI–R) scores and the effects on prediction accuracy* (Unpublished master's thesis). St. Cloud State University, St. Cloud, Minnesota.

*Austin, J., Colemen, D., Peyton, J., & Johnson, K. D. (2003). *Reliability and validity study of the LSI–R risk assessment instrument*. Unpublished manuscript, Institute on Crime, Justice and Corrections, George Washington University, Washington, D.C.

*Barnoski, R. (2006, February). *Sex offender sentencing in Washington State: Predicting recidivism based on the LSI–R* (Document No. 06-02-1201). Olympia, WA: Washington State Institute for Public Policy.

*Barnoski, R., & Aos, S. (2003). *Washington's Offender Accountability Act: An analysis of the Department of Corrections' risk assessment* (Document No. 03-12-1202). Olympia, WA: Washington State Institute for Public Policy.

*Bechtel, K., Lowenkamp, C. T., & Latessa, E. (2007). Assessing the risk of re-offending for juvenile offenders using the Youth Level of Service/Case Management Inventory. *Journal of Offender Rehabilitation, 45,* 85–108. doi:10.1300/J076v45n03_04

*Bhutta, M. H. (2013). *Risk and need assessment of offenders on probation in Lahore* (Unpublished doctoral dissertation). University of the Punjab, Lahore, Pakistan.

Bjerregaard, B., & Smith, C. (1993). Gender differences in gang participation, delinquency, and substance use. *Journal of Quantitative Criminology, 9,* 329–356. doi:10.1007/BF01064108

Blanchette, K., & Brown, S. L. (2006). *The assessment and treatment of women offenders: An integrative perspective*. Chichester, England: Wiley. doi:10.1002/9780470713013

Bloom, B., Owen, B., Covington, S., & Raeder, M. (2003). *Gender-responsive strategies: Research, Practice, and guiding principles for women offenders*. Washington, DC: U.S. Department of Justice, National Institute of Corrections.

*Bonta, J. (1981, February). *Prediction of success in community resource centres*. Paper presented at the 34th Annual Convention of the Ontario Psychological Association, Toronto, Canada.

*Bonta, J. (1989). Native inmates: Institutional response, risk, and needs. *Canadian Journal of Criminology, 31,* 49–62.

Bonta, J., Bourgon, G., Rugge, T., Scott, T., Yessine, A. K., Gutierrez, L., & Li, J. (2011). An experimental demonstration of training probation officers in evidence-based community supervision. *Criminal Justice and Behavior, 38,* 1127–1148. doi:10.1177/0093854811420678

*Bonta, J., & Higginbottom, S. (1991, August). *Parole risk prediction: A pilot project*. Poster presented at the 121st Congress of Correction of the American Correctional Association, Minneapolis, Minnesota.

*Bonta, J., & Motiuk, L. L. (1982). *Assessing incarcerated offenders for halfway houses*. Unpublished manuscript.

*Bonta, J., & Motiuk, L. L. (1985). Utilization of an interview-based classification instrument: A study of correctional halfway houses. *Criminal Justice and Behavior, 12,* 333–352. doi:10.1177/0093854885012003004

*Bonta, J., & Motiuk, L. L. (1986, August). *Use of the Level of Supervision Inventory for assessing incarcerates*. Paper presented at the 94th Annual Convention of the American Psychological Association, Washington, DC.

*Bonta, J., & Motiuk, L. L. (1987). The diversion of incarcerated offenders to correctional halfway houses. *Journal of Research in Crime and Delinquency, 24,* 302–323. doi:10.1177/0022427887024004006

*Bonta, J., Wallace-Capretta, S., & Rooney, J. (2000). A quasi-experimental evaluation of an intensive rehabilitation supervision program. *Criminal Justice and Behavior, 27,* 312–329. doi:10.1177/0093854800027003003

*Bonta, J., & Yessine, A. K. (2005). *The National Flagging System: Identifying and responding to high-risk, violent offenders* (User Report No. 2005-04). Ottawa, Canada: Public Safety and Emergency Preparedness Canada.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). Comprehensive Meta-Analysis (Version 2) [Computer software]. Englewood, NJ: Biostat.

*Bourgon, G., & Armstrong, B. (2005). Transferring the principles of effective treatment into a "real world" prison setting. *Criminal Justice and Behavior, 32,* 3–25. doi:10.1177/0093854804270618

*Brews, A. L. (2009). *The Level of Service Inventory and female offenders: Addressing issues of reliability and predictive validity* (Unpublished master's thesis). University of Saskatchewan, Saskatoon, Canada.

Brzozowski, J.-A., Taylor-Butts, A., & Johnson, S. (2006). Victimization and offending among the Aboriginal population in Canada. *Juristat, 26*(3).

*Caldwell, M. F., & Dickinson, C. (2009). Sex offender registration and recidivism risk assessment in juvenile sexual offenders. *Behavioral Sciences and the Law, 27,* 941–956. doi:10.1002/bsl.907

Calverley, D. (2007). Youth custody and community services in Canada, 2004–2005. *Juristat, 27*(2).

*Campbell, C. A. (2009). *The exploration of a mixed method and empirical short form using the Youth Level of Service/Case Management Inventory (YLS/CMI)* (Unpublished master's thesis). Michigan State University, East Lansing.

Campbell, M. A., French, S., & Gendreau, P. (2009). The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior, 36,* 567–590. doi:10.1177/0093854809333610

*Canales, D. D. (2010). *The utility of the Level of Service/Risk-Need-Responsivity (LS/RNR) instrument in predicting recidivism for adult mentally ill offenders involved in a mental health court* (Unpublished master's thesis). University of New Brunswick, Saint John, Canada.

*Catchpole, R. E., & Gretton, H. M. (2003). The predictive validity of risk assessment with violent young offenders. *Criminal Justice and Behavior, 30,* 688–708. doi:10.1177/0093854803256455

*Chng, J., Hong, N. L., & Misir, C. (2002). Security classification revisited: Predicting institutional violence using LSI–R and SAQ. In J. Wong, T. T. Fang, L. K. Sem, & T. Gob (Eds.), *Correctional research compendium (2006)* (pp. 82–87). Singapore: Singapore Prison Service.

*Chu, C. M., Ng, K., Fong, J., & Teoh, J. (2012). Assessing youth who sexually offended: The predictive validity of the ERASOR, J-SOAP-II, and YLS/CMI in a non-Western context. *Sexual Abuse: Journal of Research and Treatment, 24,* 153–174. doi:10.1177/1079063211404250

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

*Coulson, G., Ilaqua, G., Nutbrown, V., Giulekas, D., & Cudjoe, F. (1996). Predictive utility of the LSI for incarcerated female offenders. *Criminal Justice and Behavior, 23,* 427–439. doi:10.1177/0093854896023003001

Covington, S. S., & Bloom, B. E. (2007). Gender-responsive treatment and services in correctional settings. In E. Leeder (Ed.), *Inside and out: Women, prison, and therapy* (pp. 9–34). Binghamton, NY: Haworth.

*Daffern, M., Ogloff, J. R. P., Ferguson, M., & Thomson, L. (2005). Assessing risk for aggression in a forensic psychiatric hospital using the Level of Service Inventory–Revised: Screening Version. *International Journal of Forensic Mental Health, 4,* 201–206. doi:10.1080/14999013.2005.10471224

*Dahle, K.-P. (2006). Strengths and limitations of actuarial prediction of criminal reoffence in a German prison sample: A comparative study of LSI–R, HCR-20, and PCL–R. *International Journal of Law and Psychiatry, 29,* 431–442. doi:10.1016/j.ijlp.2006.03.001

*Davidson, J. T. (2007). *Risky business: What standard assessments mean for female offenders* (Unpublished doctoral dissertation). University of Hawaii, Manoa.

*Dowdy, E. R., Lacy, M. G., & Unnithan, P. (2002). Correctional prediction and the Level of Service Inventory. *Journal of Criminal Justice, 30,* 29–39. doi:10.1016/S0047-2352(01)00120-9

*Evans, S. A. (2009). *Gender disparity in the prediction of recidivism: The accuracy of LSI–R Modified* (Unpublished doctoral dissertation). University of Alabama, Tuscaloosa.

*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,* 1095–1108. doi:10.1177/0093854808320497

*Ferguson, A. M., Ogloff, J. R. P., & Thomson, L. (2009). Predicting recidivism by mentally disordered using the LSI–R:SV. *Criminal Justice and Behavior, 36,* 5–20. doi:10.1177/0093854808326525

*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,* 523–529. doi:10.1016/j.jcrimjus.2006.09.007

*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.

*Flores, A. W., Travis, L. F., & Latessa, E. J. (2004). *Case classification for juvenile corrections: An assessment of the Youth Level of Service/Case Management Inventory (YLS/CMI): Final report*. Washington, DC: National Institute of Justice.

*Folsom, J., & Atkinson, J. L. (2007). The generalizability of the LSI and the CAT to the prediction of recidivism in female offenders. *Criminal Justice and Behavior, 34,* 1044–1056. doi:10.1177/0093854807300097

*Frize, M., Kenny, D., & Lennings, C. (2008). The relationship between intellectual disability, Indigenous status and risk of reoffending in juvenile offenders on community orders. *Journal of Intellectual Disability Research, 52,* 510–519. doi:10.1111/j.1365-2788.2008.01058.x

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,* 397–426. doi:10.1177/0093854802029004004

*Girard, L., & Wormith, J. S. (2004). The predictive validity of the Level of Service Inventory–Ontario Revision on general and violent recidivism among various offender groups. *Criminal Justice and Behavior, 31,* 150–181. doi:10.1177/0093854803261335

*Gossner, D., & Wormith, J. S. (2007). The prediction of recidivism among young offenders in Saskatchewan. *Canadian Journal of Police & Security Services, 5,* 70–82.

*Guay, J.-P. (2012). *Predicting recidivism with street gang members*. (Corrections Research User Report No. 2012-02). Ottawa, Canada: Public Safety Canada.

Gutierrez, L., Wilson, H., Rugge, T., & Bonta, J. (2013). The prediction of recidivism with Aboriginal offenders: A theoretically informed meta-analysis. *Canadian Journal of Criminology and Criminal Justice, 55,* 55–99. doi:10.3138/cjccj.2011.E.51

Hannah-Moffat, K. (2009). Gridlock or mutability: Reconsidering "gender" and risk assessment. *Criminology & Public Policy, 8,* 209–219. doi:10.1111/j.1745-9133.2009.00549.x

Hannah-Moffat, K., & Maurutto, P. (2003). *Youth risk/need assessment: An overview of issues and practices*. Ottawa, Canada: Department of Justice Canada.

Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66,* 348–362. doi:10.1037/0022-006X.66.2.348

*Hanson, R. K., & Wallace-Capretta, S. (2004). Predictors of criminal recidivism among male batterers. *Psychology, Crime & Law, 10,* 413–427. doi:10.1080/10683160310001629283

*Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior, 20,* 315–335. doi:10.1177/0093854893020004001

*Hendricks, B., Werner, T., Shipway, L., & Turinetti, G. J. (2006). Recidivism among spousal abusers: Predictions and program evaluation. *Journal of Interpersonal Violence, 21,* 703–716. doi:10.1177/0886260506287310

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327,* 557–560. doi:10.1136/bmj.327.7414.557

*Hilton, N. Z., Harris, G. T., Popham, S., & Lang, C. (2010). Risk assessment among incarcerated male domestic violence offenders. *Criminal Justice and Behavior, 37,* 815–832. doi:10.1177/0093854810368937

Hoge, R. D., & Andrews, D. A. (2003). *Youth Level of Service/Case Management Inventory (YLS/CMI) user's manual*. Toronto, Canada: Multi-Health Systems.

*Hoge, R. D., Andrews, D. A., & Leschied, A. W. (1994). Tests of three hypotheses regarding the predictors of delinquency. *Journal of Abnormal Child Psychology, 22,* 547–559. doi:10.1007/BF02168937

*Hoge, R. D., Andrews, D. A., & Leschied, A. W. (1996). An investigation of risk factors and protective factors in a sample of youthful offenders. *Journal of Child Psychology and Psychiatry, 37,* 419–424. doi:10.1111/j.1469-7610.1996.tb01422.x

*Hogg, S. M. (2011). *The Level of Service Inventory (Ontario Revision) scale validation for gender and ethnicity: Addressing reliability and predictive validity* (Unpublished master's thesis). University of Saskatchewan, Saskatoon, Canada.

*Hollin, C. R., & Palmer, E. J. (2006). The Level of Service Inventory–Revised profile of English prisoners: Risk and reconviction analysis. *Criminal Justice and Behavior, 33,* 347–366. doi:10.1177/0093854806286195

Holsinger, A. M., Lowenkamp, C. T., & Latessa, E. J. (2003). Ethnicity, gender, and the Level of Service Inventory–Revised. *Journal of Criminal Justice, 31,* 309–320. doi:10.1016/S0047-2352(03)00025-4

*Holsinger, A. M., Lowenkamp, C. T., & Latessa, E. J. (2004, Winter/Spring). Validating the LSI–R on a sample of jail inmates. *Journal of Offender Monitoring, 2004*, 8–9.

*Holsinger, A. M., Lowenkamp, C. T., & Latessa, E. J. (2006a). Exploring the validity of the Level of Service Inventory–Revised with Native American offenders. *Journal of Criminal Justice, 34,* 331–337. doi:10.1016/j.jcrimjus.2006.03.009

*Holsinger, A. M., Lowenkamp, C. T., & Latessa, E. J. (2006b). Predicting institutional misconduct using the Youth Level of Service/Case Management Inventory. *American Journal of Criminal Justice, 30,* 267–284. doi:10.1007/BF02885895

Holtfreter, K., & Cupp, R. (2007). Gender and risk assessment: The empirical status of the LSI–R for women. *Journal of Contemporary Criminal Justice, 23,* 363–382. doi:10.1177/1043986207309436

*Holtfreter, K., Reisig, M. D., & Morash, M. (2004). Poverty, state capital, and recidivism among women offenders. *Criminology & Public Policy, 3,* 185–208. doi:10.1111/j.1745-9133.2004.tb00035.x

*Hong, N. L., Misir, C., & Chng, J. (2002). LSI–R:SV: A quick risk assessment tool. In J. Wong, T. T. Fang, L. K. Sem, & T. Gob (Eds.), *Correctional research compendium (2006)* (pp. 88–93). Singapore: Singapore Prison Service.

*Ilacqua, G. E., Coulson, G. E., Lombardo, D., & Nutbrown, V. (1999). Predictive validity of the Young Offender Level of Service Inventory for criminal recidivism of male and female young offenders. *Psychological Reports, 84,* 1214–1218.

*Jack, L. A. (2000). *Psychopathy, risk/need factors, and psychiatric symptoms in high-risk youth: Relationships between variables and their link*

to recidivism (Unpublished doctoral dissertation). Simon Fraser University, Burnaby, Canada.

*Jones Hubbard, D. (2002). *Cognitive-behavioral treatment: An analysis of gender and other responsivity characteristics and their effects on success in offender rehabilitation* (Unpublished doctoral dissertation). University of Cincinnati, Cincinnati, Ohio.

*Jung, S., & Rawana, E. P. (1999). Risk and need assessment of juvenile offenders. *Criminal Justice and Behavior, 26,* 69–89. doi:10.1177/0093854899026001004

*Kelly, C. E., & Welsh, W. N. (2008). The predictive validity of the Level of Service Inventory–Revised for drug involved offenders. *Criminal Justice and Behavior, 35,* 819–831. doi:10.1177/0093854808316642

*Kim, H. (2010). *Prisoner classification re-visited: A further test of the Level of Service Inventory–Revised (LSI–R) intake assessment* (Unpublished doctoral dissertation). Indiana University of Pennsylvania, Indiana, Pennsylvania.

*Kirkpatrick, B. (1998). Field test of the LSI–R: A study of offenders under intensive community supervision. *Perspectives, 22*(4), 24–28.

*Kirkpatrick, B. (1999). Exploratory research of female risk prediction and the LSI–R. *Corrections Compendium, 24*(5), 1–3, 14–17.

LaPrairie, C. (1995). *Seen but not heard: Native people in the inner city*. Ottawa, Canada: Department of Justice Canada.

*Latessa, E. J., Lowenkamp, C. T., Myer, A., & Ndreka, M. (2009). *MIDAS outcome evaluation: Final report*. Unpublished manuscript, University of Cincinnati, Cincinnati, Ohio.

*Latessa, E., Smith, P., Lemke, R., Makarios, M., & Lowenkamp, C. (2009). *Creation and validation of the Ohio Risk Assessment System: Final report*. Unpublished manuscript, University of Cincinnati, Cincinnati, Ohio.

Lipsey, M. W. (2003). Those confounded moderators in meta-analysis: Good, bad and ugly. *Annals of the American Academy of Political and Social Science, 587,* 69–81. doi:10.1177/0002716202250791

Lipsey, M. W., & Wilson, D. B. (2001). *Applied Social Research Methods Series: Vol. 49. Practical meta-analysis*. Thousand Oaks, CA: Sage.

*Livsey, S. E. (2005). *Is the YLS a valid predictor of juvenile recidivism?* (Unpublished doctoral dissertation). Michigan State University, East Lansing.

*Lovell, D., Gagliardi, G. J., & Phipps, P. (2005). *Washington's Dangerous Mentally Ill Offender Law: Was community safety increased?* (Document No. 05-03-1901). Olympia, WA: Washington State Institute for Public Policy.

*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. M., & Latessa, E. J. (2001). Risk/need assessment, offender classification, and the role of childhood abuse. *Criminal Justice and Behavior, 28,* 543–563. doi:10.1177/009385480102800501

*Lowenkamp, C. T., & Latessa, E. J. (2002). *Level of Service Inventory Revised (LSI–R) validation study: North Dakota*. Unpublished manuscript, University of Cincinnati, Cincinnati, Ohio.

*Lowenkamp, C. T., & Latessa, E. J. (2006). *Norming and validation study of the Level of Service Inventory Revised (LSI–R) and the Illinois Pre-Screen (IL Pre-Screen)*. Unpublished manuscript, University of Cincinnati, Cincinnati, Ohio.

*Lowenkamp, C. T., Lovins, B., & Latessa, E. (2009). Validating the Level of Service Inventory–Revised and the Level of Service Inventory: Screening Version with a sample of probationers. *Prison Journal, 89,* 192–204. doi:10.1177/0032885509334755

*Loza, W., & Loza-Fanous, A. (2001). Effectiveness of the Self-Appraisal Questionnaire in predicting offenders' postrelease outcome: A comparison study. *Criminal Justice and Behavior, 28,* 105–121. doi:10.1177/0093854801028001005

*Luong, D. (2007). *Risk assessment and community management: The relationship between implementation quality and recidivism* (Unpublished master's thesis). University of Saskatchewan, Saskatoon, Canada.

Luong, D., & Wormith, J. S. (2011). Applying risk/need assessment to probation practice and its impact on the recidivism of young offenders. *Criminal Justice and Behavior, 38,* 1177–1199. doi:10.1177/0093854811421596

*Manchak, S. M., Skeem, J. L., & Douglas, K. S. (2008). Utility of the Revised Level of Service Inventory in predicting recidivism after long-term incarceration. *Law and Human Behavior, 32,* 477–488. doi:10.1007/s10979-007-9118-4

*Manchak, S. M., Skeem, J. L., Douglas, K. S., & Siranosian, M. (2009). Does gender moderate the predictive utility of the Level of Service Inventory–Revised (LSI–R) for serious violent offenders? *Criminal Justice and Behavior, 36,* 425–442. doi:10.1177/0093854809333058

Mann, M. M. (2010). *Good intentions, disappointing results: A progress report on federal Aboriginal corrections*. Ottawa, Canada: Office of the Correctional Investigator. Retrieved from http://www.oci-bec.gc.ca/cnt/rpt/pdf/oth-aut/oth-aut20091113-eng.pdf

*Marczyk, G. (2002). *Predicting juvenile recidivism and validating juvenile risk factors in an urban environment* (Unpublished doctoral dissertation). Medical College of Pennsylvania and Hahnemann University, Philadelphia.

*Marshall, J., Egan, V., English, M., & Jones, R. M. (2006). The relative validity of psychopathy versus risk/needs-based assessments in the prediction of adolescent offending behavior. *Legal and Criminological Psychology, 11,* 197–210. doi:10.1348/135532505X68719

Martel, J., Brassard, R., & Jaccoud, M. (2011). When two worlds collide: Aboriginal risk management in Canadian corrections. *British Journal of Criminology, 51,* 235–255. doi:10.1093/bjc/azr003

McClellan, D. S., Farabee, D., & Crouch, B. M. (1997). Early victimization, drug use, and criminality: A comparison of male and female offenders. *Criminal Justice and Behavior, 24,* 455–476.

*McConnell, B. (1996). *The prediction of female federal offender recidivism with the Level of Supervision Inventory* (Unpublished honors thesis). Queen's University, Kingston, Canada.

*McKinnon, L. J. (2004). *Predicting risk of violence in a young offender population: The predictive validity of the PCL:YV and the YLS/CMI* (Unpublished master's thesis). Lakehead University, Thunder Bay, Canada.

*Mihailides, S., Jude, B., & Van den Bossche, E. (2005). The LSI–R in an Australian setting: Implications for risk/needs decision-making in forensic contexts. *Psychiatry, Psychology and Law, 12,* 207–217. doi:10.1375/pplt.2005.12.1.207

*Mills, J. F., & Kroner, D. G. (2006). The effect of discordance among violence and general recidivism risk estimates on predictive accuracy. *Criminal Behaviour and Mental Health, 16,* 155–166. doi:10.1002/cbm.623

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine, 6,* e1000097. doi:10.1371/journal.pmed.1000097

*Mooney, N. P. (2010). *Predicting reoffending within in the New Zealand Youth Justice System: Evaluating measures of risk, need, and psychopathy* (Unpublished doctoral dissertation). Massey University, Wellington, New Zealand.

*Morton, K. E. (2003). *Psychometric properties of four risk assessment measures with male adolescent sexual offenders* (Unpublished master's thesis). Carleton University, Ottawa, Canada.

*Motiuk, L. L. (1991). *Antecedents and consequences of prison adjustment: A systematic assessment and reassessment approach* (Unpublished doctoral dissertation). Carleton University, Ottawa, Canada.

*Motiuk, L. L. (1993). Using the LSI and other classification systems to better predict halfway-house outcome. *Journal of Community Corrections, 5*(4), 8–9.

*Motiuk, L. L., Bonta, J., & Andrews, D. A. (1990, June). *Dynamic predictive criterion validity in offender assessment*. Paper presented at the annual meeting of the Canadian Psychological Association, Ottawa, Canada.

*Motiuk, M. S., Motiuk, L. L., & Bonta, J. (1992). A comparison between self-report and interview-based inventories in offender classification. *Criminal Justice and Behavior, 19,* 143–159. doi:10.1177/0093854892019002003

*Newman, T. R. (2004). *Utilizing the LSI–R specific to the population of the Community Correctional Center* (Unpublished master's thesis). Northern Kentucky University, Highland Heights.

*Nowicka-Sroga, M. (2004). *The LSI–OR: A recidivism follow-up study within a sample of male young offenders* (Unpublished doctoral dissertation). University of Ottawa, Ottawa, Canada.

*Nugent, P. (2000). *The use of detention legislation: Factors affecting detention decisions and recidivism among high-risk federal offenders in Ontario* (Unpublished doctoral dissertation). Queen's University, Kingston, Canada.

*O'Keefe, M. L., & Wensuc, G. E. (1998). *The validation of the LSI on community corrections populations* [Report to the Article 11.5 Advisory Committee, Colorado]. Unpublished manuscript.

*Olver, M. E., Stockdale, K. C., & Wong, S. C. P. (2012). Short and long-term prediction of recidivism using the YLS/CMI in a sample of serious young offenders. *Law and Human Behavior, 36,* 331–344. doi:10.1037/h0093927

Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2009). Risk assessment with young offenders: A meta-analysis of three assessment measures. *Criminal Justice and Behavior, 36,* 329–353. doi:10.1177/0093854809331457

*Onifade, E., Davidson, W., Campbell, C., Turke, G., Malinowski, J., & Turner, K. (2008). Predicting recidivism in probationers with the Youth Level of Service Case Management Inventory (YLS/CMI). *Criminal Justice and Behavior, 35,* 474–483. doi:10.1177/0093854807313427

*Onifade, E., Smith Nyandoro, A., Davidson, W. S., & Campbell, C. (2010). Truancy and patterns of criminogenic risk in a young offender population. *Youth Violence and Juvenile Justice, 8,* 3–18. doi:10.1177/1541204009338251

Orwin, R. G. (1983). A fail-safe *N* for effect size in meta-analysis. *Journal of Educational Statistics, 8,* 157–159. doi:10.2307/1164923

*Palmer, E. J., & Hollin, C. R. (2007). The Level of Service Inventory–Revised with English women prisoners: A needs and reconviction analysis. *Criminal Justice and Behavior, 34,* 971–984. doi:10.1177/0093854807300819

Perreault, S. (2011). *Violent victimization of Aboriginal people in the Canadian provinces, 2009* (Statistics Canada Catalogue No. 85-002-X). Retrieved from http://www.statcan.gc.ca/pub/85-002-x/2011001/article/11415-eng.pdf

Poston, J. M., & Hanson, W. E. (2010). Meta-analysis of psychological assessment as a therapeutic intervention. *Psychological Assessment, 22,* 203–212. doi:10.1037/a0018679

*Raynor, P. (2007). Risk and need assessment in British probation: The contribution of LSI–R. *Psychology, Crime & Law, 13,* 125–138. doi:10.1080/10683160500337592

*Rector, B., Wormith, J. S., & Banka, D. (2007, June). Predictive validity of the LSI–SK Youth Edition. In D. A. Andrews (Chair), *Level of Service Inventory: Risk/need assessment of female, young and Aboriginal offenders*. Symposium presented at the First North American Correctional and Criminal Justice Psychology Conference, Ottawa, Canada.

*Reisig, M. D., Holtfreter, K., & Morash, M. (2006). Assessing recidivism risk across female pathways to crime. *Justice Quarterly, 23,* 384–405. doi:10.1080/07418820600869152

*Rennie, C., & Dolan, M. (2010). Predictive validity of the Youth Level of Service/Case Management Inventory in custody sample in England. *Journal of Forensic Psychiatry & Psychology, 21,* 407–425. doi:10.1080/14789940903452311

*Rettinger, L. J., & Andrews, D. A. (2010). General risk and need, gender specificity, and the recidivism of female offenders. *Criminal Justice and Behavior, 37,* 29–46. doi:10.1177/0093854809349438

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's *d,* and *r*. *Law and Human Behavior, 29,* 615–620. doi:10.1007/s10979-005-6832-7

*Rowe, R. C. (1995). *The utilization of an interview-based classification instrument for parole board decision making in Ontario* (Unpublished master's thesis). Carleton University, Ottawa, Canada.

*Rowe, R. C. (2002). *Predictors of criminal offending: Evaluating measures of risk/needs, psychopathy, and disruptive behavior disorders* (Unpublished doctoral dissertation). Carleton University, Ottawa, Canada.

Rugge, T. (2006). *Risk assessment of male aboriginal offenders* (Corrections Research Report No. 2006-01). Ottawa, Canada: Public Safety Canada.

Salisbury, E. J., & Van Voorhis, P. (2009). Gendered pathways: A quantitative investigation of women probationers' paths to incarceration. *Criminal Justice and Behavior, 36,* 541–566. doi:10.1177/0093854809334076

*Salisbury, E. J., Van Voorhis, P., & Spiropoulos, G. V. (2009). The predictive validity of a gender-responsive needs assessment: An exploratory study. *Crime & Delinquency, 55,* 550–585. doi:10.1177/0011128707308102

*Schlager, M. D. (2005). *Assessing the reliability and validity of the Level of Service Inventory–Revised (LSI–R) on a community correction sample: Implications for corrections and parole policy* (Unpublished doctoral dissertation). Rutgers, The State University of New Jersey, Newark.

*Schlager, M. D., & Simourd, D. J. (2007). Validity of the Level of Service Inventory–Revised (LSI–R) among African American and Hispanic male offenders. *Criminal Justice and Behavior, 34,* 545–554. doi:10.1177/0093854806296039

*Schmidt, F., Hoge, R. D., & Gomes, L. (2005). Reliability and validity analyses of the Youth Level of Service/Case Management Inventory. *Criminal Justice and Behavior, 32,* 329–344. doi:10.1177/0093854804274373

Schwalbe, C. S. (2007). Risk assessment for juvenile offenders: A meta-analysis. *Law and Human Behavior, 31,* 449–462. doi:10.1007/s10979-006-9071-7

Schwalbe, C. S. (2008). A meta-analysis of juvenile justice risk assessment instruments: Predictive validity by gender. *Criminal Justice and Behavior, 35,* 1367–1381. doi:10.1177/0093854808324377

Scrim, K. (2010). Aboriginal victimization in Canada: A summary of the literature. *Victims of Crime Research Digest, 2010*(3), 15–19.

*Shields, I. (1991, June). *The Young Offender Level of Service Inventory as a predictor of recidivism: A one-year follow-up study*. Paper presented at the annual meeting of the Canadian Psychological Association, Calgary, Canada.

*Simourd, D. J. (2004). Use of dynamic risk/need assessment instruments among long-term incarcerated offenders. *Criminal Justice and Behavior, 31,* 306–323. doi:10.1177/0093854803262507

*Simourd, D. J. (2006). *Validation of risk/needs assessments in the Pennsylvania Department of Corrections: Final report*. Unpublished manuscript.

Singh, J. P., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review, 31,* 499–513. doi:10.1016/j.cpr.2010.11.009

*Skowron, C. (2004). *Differentiation and predictive factors in adolescent sexual offending* (Unpublished doctoral dissertation). Carleton University, Ottawa, Canada.

Smith, P., Cullen, F. T., & Latessa, E. J. (2009). Can 14,737 women be wrong? A meta-analysis of the LSI–R and recidivism for female offenders. *Criminology & Public Policy, 8,* 183–208. doi:10.1111/j.1745-9133.2009.00551.x

*Stewart, C. A. (2011). *Risk assessment of federal female offenders* (Unpublished doctoral dissertation). University of Saskatchewan, Saskatoon, Canada.

*Swoboda, D. G. (2006). *Predicting misconducts among young adult male inmates using the Level of Service Inventory–Revised and the Personality Assessment Inventory* (Unpublished doctoral dissertation). Indiana University of Pennsylvania, Indiana, Pennsylvania.

*Takahashi, M. (2010). *Predictive validity of the Youth Level of Service/Case Management Inventory among Japanese juvenile offenders* (Unpublished master's thesis). Southern Illinois University, Carbondale.

*Taylor Kindrick, C. Y. (2011). *Girls and boys, apples and oranges? A theoretically informed analysis of gender-specific predictors of delinquency* (Unpublished doctoral dissertation). University of Cincinnati, Cincinnati, Ohio.

*Thompson, A. P., & McGrath, A. (2012). Subgroup differences and implications for contemporary risk-need assessment with juvenile offenders. *Law and Human Behavior, 36,* 345–355. doi:10.1037/h0093930

*Thompson, A. P., & Pope, Z. (2005). Assessing juvenile sex offenders: Preliminary data for the Australian adaptation of the Youth Level of Service/Case Management Inventory (Hoge & Andrews, 1995). *Australian Psychologist, 40,* 207–214. doi:10.1080/00050060500243491

*Upperton, R. A., & Thompson, A. P. (2007). Predicting juvenile offender recidivism: Risk–need assessment and juvenile justice officers. *Psychiatry, Psychology and Law, 14,* 138–146. doi:10.1375/pplt.14.1.138

*van de Ven, J. T. C. (2004). *Assessment of risk and need factors and service use in diverted youth (Unpublished doctoral dissertation*). Carleton University, Ottawa, Canada.

*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,* 261–288. doi:10.1177/0093854809357442

*Vieira, T. A., Skilling, T. A., & Peterson-Badali, M. (2009). Matching court-ordered treatment services with treatment needs: Predicting treatment success with young offenders. *Criminal Justice and Behavior, 36,* 385–401. doi:10.1177/0093854808331249

*Viljoen, J. L., Elkovitch, N., Scalora, M. J., & Ullman, D. (2009). Assessment of re-offense risk in adolescents who have committed sexual offenses. *Criminal Justice and Behavior, 36,* 981–1000. doi:10.1177/0093854809340991

*Vose, B., Lowenkamp, C. T., 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,* 459–471. doi:10.1177/1043986209344797

*Walters, G. D. (2011). Predicting recidivism with the Psychological Inventory of Criminal Thinking Styles and Level of Service Inventory–Revised: Screening Version. *Law and Human Behavior, 35,* 211–220. doi:10.1007/s10979-010-9231-7

*Walters, G. D., & Schlauch, C. (2008). The Psychological Inventory of Criminal Thinking Styles and Level of Service Inventory–Revised: Screening Version as predictors of official and self-reported disciplinary infractions. *Law and Human Behavior, 32,* 454–462. doi:10.1007/s10979-007-9117-5

*Watkins, I. (2011). *The utility of Level of Service Inventory–Revised (LSI–R) assessments within NSW correctional environments* (Research Bulletin No. 29). Sydney, Australia: Corrective Services NSW.

*Whiteacre, K. W. (2006). Testing the Level Service Inventory–Revised (LSI–R) for racial/ethnic bias. *Criminal Justice Policy Review, 17,* 330–342. doi:10.1177/0887403405284766

Wormith, J. S. (2011). The legacy of D. A. Andrews in the field of criminal justice: How theory and research can change policy and practice. *International Journal of Forensic Mental Health, 10,* 78–82. doi:10.1080/14999013.2011.577138

Wormith, J. S., & Hogg, S. (2012). *The predictive validity of Aboriginal offender recidivism with a general risk/needs assessment inventory*. Unpublished manuscript, Department of Psychology, University of Saskatchewan, Saskatoon, Canada.

*Wormith, J. S., Hogg, S., & Guzzo, L. (2012). The predictive validity of a general risk/needs assessment inventory on sexual offender recidivism and an exploration of the professional override. *Criminal Justice and Behavior, 39,* 1511–1538. doi:10.1177/0093854812455741

*Wormith, J. S., Olver, M. E., Stevenson, H. E., & Girard, L. (2007). The long-term prediction of offender recidivism using diagnostic, personality, and risk/need approaches to offender assessment. *Psychological Services, 4,* 287–305. doi:10.1037/1541-1559.4.4.287

Wright, E. M., Salisbury, E. J., & Van Voorhis, P. (2007). Predicting the prison misconducts of women offenders: The importance of gender-responsive needs. *Journal of Contemporary Criminal Justice, 23,* 310–340. doi:10.1177/1043986207309595

*Wright, E. M., Van Voorhis, P., Bauman, A., & Salisbury, E. (2007). *Gender-responsive risk/needs assessment: Final report* [Prepared for the Minnesota Department of Corrections and the Advisory Task Force on the Woman and Juvenile Female Offender in Corrections]. Unpublished manuscript.

Yang, M., Wong, S. C. P., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment instruments. *Psychological Bulletin, 136,* 740–767. doi:10.1037/a0020473

Received March 10, 2013
Revision received July 22, 2013
Accepted September 30, 2013 ■