

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/6774683>

Youth Psychopathy and Criminal Recidivism: A Meta-Analysis of the Psychopathy Checklist Measures

Article in *Law and Human Behavior* · March 2007

DOI: 10.1007/s10979-006-9019-y · Source: PubMed

CITATIONS

282

READS

2,318

3 authors, including:



John F Edens

Texas A&M University

188 PUBLICATIONS 8,995 CITATIONS

[SEE PROFILE](#)



Justin S Campbell

U.S. Navy

39 PUBLICATIONS 704 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Response Distortion in Personality Assessment [View project](#)



Psychopathic traits in Juveniles [View project](#)

Youth Psychopathy and Criminal Recidivism: A Meta-Analysis of the Psychopathy Checklist Measures

John F. Edens · Justin S. Campbell · John M. Weir

Published online: 4 October 2006

© American Psychology-Law Society/Division 41 of the American Psychological Association 2006

Abstract Although narrative reviews have suggested that “youth psychopathy” is a strong predictor of future crime and violence, to date no quantitative summaries of this literature have been conducted. We meta-analyzed recidivism data for the Psychopathy Checklist measures across 21 non-overlapping samples of male and female juvenile offenders. After removing outliers, psychopathy was significantly associated with general and violent recidivism (r_w 's of .24 and .25, respectively), but negligibly related to sexual recidivism in the few studies examining this low base rate outcome. Even after eliminating outliers, however, considerable heterogeneity was noted among the effects, with some of this variability being explained by the gender and ethnic composition of the samples. Effect sizes for the small number of female samples available for analysis were mostly small and nonsignificant, and psychopathy was a weaker predictor of violent recidivism among more ethnically heterogeneous samples. In relation to predicting both general and violent recidivism, psychopathy performed comparably to an instrument designed specifically to assess risk, the Youth Level of Service/Case Management Inventory (Hoge & Andrews, 2002).

Keywords Psychopathy · Juveniles · Recidivism · Gender · Ethnicity

Psychopathic personality disorder (psychopathy) represents a cluster of distinct affective, interpersonal, and behavioral characteristics (Cleckley, 1976; Hare, 2003) that demarcate a clinical construct of theoretical and practical significance. Of particular interest to both clinicians and policy makers is the demonstrated association between psychopathic traits and various forms of socially undesirable outcomes, such as community violence and criminal recidivism, which has been evident in several large-scale studies and meta-analytic reviews (e.g., Gendreau, Goggin,

J. F. Edens (✉)
Department of Psychology, Southern Methodist University,
310D Hyer Hall, Dallas, Texas, 75275-0442
e-mail: jedens@smu.edu

J. S. Campbell · J. M. Weir
University of South Florida,
Tampa, Florida

& Smith, 2002; Hemphill, Hare, & Wong, 1998; Salekin, Rogers, & Sewell, 1996). Moreover, this association is not limited to criminal populations, as research on civil psychiatric samples also suggests a link between psychopathic traits and violence (e.g., Skeem & Mulvey, 2001).

Considerable research efforts have been devoted to quantifying psychopathic traits (or at least their developmental precursors) in younger and younger samples. For example, two recent issues of the journal *Behavioral Sciences and the Law* have been devoted to the topic of “juvenile psychopathy” (Petrila & Skeem, 2003; Skeem & Petrila, 2004), with many of the articles focused on the relationship between psychopathy and various forms of socially deviant conduct (e.g., Corrado, Vincent, Hart, & Cohen, 2004; Spain, Douglas, Poythress, & Epstein, 2004). More broadly, youth violence is a topic of increasing public concern, perhaps inspired by tragic events such as the school shootings in Columbine, Colorado, as well as media accounts suggesting that youthful offenders of today are becoming increasingly predatory, violent, and untreatable (despite evidence suggesting that youth crime actually is on the decline; Baum, 2005).

Because of the potential explanatory power of psychopathy as a cause of violent conduct, it seems natural that researchers and clinicians would increasingly attend to psychopathy as a potential marker for identifying children and adolescents most likely to engage in serious violence (e.g., O’Shaughnessy, 2004). From a preventative intervention perspective, identification of “fledgling psychopaths” ostensibly might provide an important means of singling out those who may be most in need of intervention (Lynam, 2002). From a treatment matching perspective, it is conceivable that unique aspects of psychopathy or its developmental precursors require different methods of intervention than those used with more common “adolescent-onset” offenders. Regardless of the potential for psychopathy to inform clinical decision-making, the development and mass marketing of various measures to assess psychopathic traits among children and adolescents (e.g., Psychopathy Checklist: Youth Version [PCL:YV], Forth, Kosson, & Hare, 2003; Antisocial Process Screening Device [APSD], Frick & Hare, 2001; P-Scan, Hare & Hervé, 1999) portends an increased use in applied settings. Although systematic surveys have yet to be conducted, considerable anecdotal evidence suggests that the PCL:YV in particular is being used to inform clinical and legal decision-making for youths involved in the juvenile and criminal justice systems in the United States and Canada (Edens & Petrila, 2006; Forth, 2004; Skeem, Cauffman, & Nold, 2002; Zinger & Forth, 1998).

The putative role of psychopathy as a risk factor for violence is reflected in its inclusion in several risk assessment instruments and procedures, such as the HCR-20 (Webster, Douglas, Eaves, & Hart, 1997) and the Violence Risk Appraisal Guide (Quinsey, Harris, Rice, & Cormier, 1998). Risk measures developed specifically for youth populations, such as the first version of the Structured Assessment of Violence Risk in Youth (SAVRY; Borum, Bartel, & Forth, 2002), have incorporated this instrument as well. Such measures reflect the increasing use of psychopathy “outside the lab” to inform both clinical and legal decision-making across an array of contexts (Edens & Petrila, 2006; Ogloff & Lyon, 1998).

The rapidly expanding clinical and legal application of psychopathy has not progressed without numerous reservations being voiced. Many of these concerns revolve around the potential misuses of psychopathy scales, such as by poorly trained or uninformed examiners (DeMatteo & Edens, 2005; Edens, 2001, 2006; Edens & Petrila, 2006; Hare, 1998). While arguing in support of the construct validity of psychopathy measures, Zinger and Forth (1998) identified several additional qualms related to potential human rights abuses of these instruments when used to inform legal decision-making (see also Forth, 2004). Other criticisms have focused on the questionable applicability of an ostensibly stable, adult personality construct to youths (Edens, Skeem, Cruise, & Cauffman, 2001; Hart, Watt, & Vincent, 2002; Seagrave & Grisso, 2002). Still others have raised concerns about the stigmatizing nature of the traits attributed

to psychopathy (e.g., pathological lying, superficial charm, remorselessness), particularly when used in the context of legal decision-making where evidence is presented to nonmental health professionals (Edens, 2006; Edens, Guy, & Fernandez, 2003; cf. Murrie, Cornell, & McCoy, 2005).

Aside from the above-noted concerns, two other topics increasingly seem to be the focus of attention of researchers in this area. First, whether and/or how the construct of psychopathy can be applied to women is a growing area of research interest, with a plethora of recent reviews and empirical studies addressing this topic (see, e.g., Cale & Lilienfeld, 2002; Nicholls & Pettilä, 2005). In particular, the predictive validity of psychopathy with female adolescents has come under question, with effect sizes being reported that are statistically nonsignificant and considerably below what have typically been reported in the adult male literature (e.g., Campbell, 2003; Odgers, Repucci, & Moretti, 2005; Vincent, Odgers, Watkins, & Corrado, 2005).

A second area of considerable debate relates to the magnitude of the association between psychopathy and recidivism and the relative contribution of this construct in comparison to other risk factors. Some large-scale reviews (Gendreau et al., 2002; Walters, 2003b) have concluded that the PCL instruments are not the “unparalleled” measures of risk that was claimed in earlier meta-analyses (Salekin et al., 1996) and that their predictive validity is only comparable to or perhaps even less than other instruments that are not wedded to the potentially stigmatizing term “psychopath” (cf. Hemphill & Hare, 2004). Perhaps reflecting this line of thought, the recently developed MacArthur actuarial risk instrument (Banks et al., 2004) developed to predict violence among released psychiatric patients does not include psychopathy as a risk factor, despite its comparatively robust association with community violence at the bivariate level (Skeem & Mulvey, 2001).

Methodologically, prior reviews of the psychopathy/recidivism literature have suffered limitations in terms of informing our understanding of this relationship specifically among youths. First, many quantitative reviews (e.g., Gendreau et al., 2002; Hemphill et al., 1998; Walters, 2003a, b) either have not included studies of youths or have not differentiated between studies of adults versus studies of adolescents. Second, recent summaries focusing specifically on adolescents have been narrative reviews (e.g., Edens et al., 2001; Forth et al., 2003; Forth & Mailloux, 2000), which can be criticized for lacking statistical and methodological rigor and being susceptible to various forms of bias (for reviews, see Lipsey & Wilson, 2001; Rosenthal & DiMatteo, 2001). One especially troublesome problem for the narrative literature review is reliance upon significance testing. As Hunter and Schmidt (1996) aptly demonstrated, with a true population effect size of .30, the narrative literature reviewer who examines 50 studies with equal sample sizes and statistical power of .50 would observe 25 effects with nonsignificant *p*-values and another 25 effects with significant *p*-values. Despite the true population effect of .30, the reviewer would likely reach the erroneous conclusion that the literature was “unclear,” “mixed,” or indicative of no relationship. On the other hand, applying meta-analytic methods to the same database, the extreme effects on either side of the population effect size would cancel each other, the result being a mean effect size of .30 and a more accurate reflection of the true effect size. Partly because of its ability to reconcile disparate findings, meta-analysis is quickly becoming the state of the art method for establishing consensus regarding social science findings (Hunter, 1999). Consequently, professionals within forensic settings are increasingly calling for the use of meta-analysis to establish generally accepted scientific knowledge (Blumenthal, 2002).

The purpose of the present study is to provide a quantitative overview of the psychopathy and criminal recidivism literature for male and female adolescent offenders using meta-analytic methodology to eliminate or minimize many limitations inherent in qualitative literature reviews. Consistent with outcome measures used in the adult recidivism literature

(e.g., Gendreau et al., 2002; Hemphill et al., 1998; Walters, 2003b), almost all of the juvenile studies of psychopathy have collapsed into one outcome essentially any form of recidivism (i.e., both nonviolent and violent crimes). As such, we examined a “general” outcome category in the present study that reflected this broadband inclusion of all forms of re-offending. Perhaps reflecting a heightened interest in more “severe” forms of crime, however, recidivism that specifically involved a violent act (e.g., assault, armed robbery, attempted murder) also was coded in most original studies as a separate and narrower outcome category, allowing us to address the relationship between psychopathy and this criterion. Additionally, in a smaller number of studies researchers also parsed out sexual re-offending as a distinct class of offenses. As such, the present meta-analysis will separately examine each of these outcomes (for a similar categorization of recidivism used in the adult literature, see Hemphill et al., 1998).

Aside from quantifying the predictive validity of psychopathy across the outcome measures mentioned above, another significant advantage of a meta-analytic approach over a narrative review is it allows for an examination of potential moderator variables that may help to explain heterogeneity noted among effect sizes. Prior meta-analyses of psychopathy research (e.g., Guy, Edens, Anthony, & Douglas, 2005; Hemphill et al., 1998; Walters, 2003b) have noted considerable variability among effect sizes, a finding that compels moderator analysis of study variables that might account for this heterogeneity. Aside from general concerns regarding the methodological rigor of studies included in meta-analyses (see, generally, Lipsey & Wilson, 2001, pp. 20–23), prior research suggests three more specific study characteristics that may impinge on the aggregate predictive validity of psychopathy, both among youths and adults. First, in the juvenile institutional adjustment literature (Edens & Compbell, 2005), some evidence of publication bias has been noted, with unpublished studies producing significantly smaller effects. Second, as noted earlier, potential gender differences in predictive validity (e.g., Odgers et al., 2005) raise questions as to whether psychopathy may be less useful as an indicator of recidivism risk among female adolescents. Third, the degree of ethnic diversity among samples and the country of origin of the sample have been suggested as potential moderators in the adult institutional adjustment literature (Edens, Petrila, & Buffington-Vollum, 2001; Guy et al., 2005; see also Hicks, Rogers, & Cashel, 2000). As such, we will examine the impact of these potential moderators for outcomes with significant effect size heterogeneity.

Although there are now several published scales ostensibly measuring psychopathic traits, we should clarify at the outset that we chose to focus exclusively on the Psychopathy Checklist measures because (a) they have been by far the most heavily researched in relation to criminal outcomes (Edens et al., 2001), with very few recidivism studies having been published examining other juvenile psychopathy measures, (b) other extant scales do not correlate particularly highly with the PCL measures (Forth et al., 2003), which raises serious questions as to whether the same underlying construct is being operationalized, and (c) pragmatically, at present the youth version of the PCL is the most likely to be used by clinicians in applied settings (Seagrave & Grisso, 2002). Although various debates continue regarding the PCL measures (e.g., whether their underlying structure is best represented by three- or four-factor models; Cooke, Michie, Hart, & Clark, 2005; Hare, 2003), they clearly seem to be the predominant measures of psychopathy used in clinical and forensic settings, with total scores being weighted most heavily in decision-making (Edens & Petrila, 2006; see also Marczyk, Heilbrun, Lander, & DeMatteo, 2005).

Finally, although our primary focus was on quantifying the predictive validity of psychopathy at the bivariate level, we also were interested in the performance of PCL measures in relation to other instruments used in risk assessments conducted on adolescents, given controversies about the relative contribution of psychopathy versus other risk measures in the adult literature (Gendreau et al., 2002; Hemphill & Hare, 2004). Although direct comparisons to other risk

instruments have been relatively sporadic, we were able to identify five studies with reasonably large sample sizes that compared the predictive utility of psychopathy to a widely used risk assessment instrument, the Youth Level of Service/Case Management Inventory (YLS/CMI; Hoge & Andrews, 2002). As such, we also report analyses of the performance of PCL-defined psychopathy relative to this measure.

Method

Sample identification

The search for effect sizes began with a review of abstracts from published peer reviewed academic journals and dissertations/theses circulating prior to July 1, 2005. Relevant books, book chapters, unpublished dissertations and theses also were reviewed. A web listing of psychopathy references developed by Stephen Hart, Adelle Forth, and Robert Hare (*Key References Related to the Study of Psychopathy*: <http://www.hare.org/references/>) also was examined for relevant publications. Computerized databases (PsychInfo, UMI Proquest Digital Dissertations) were searched to identify studies that employed a version of the PCL in a youthful sample, using the following keywords: “Psychopathy Checklist,” “Psychopathy Checklist Revised,” “18-item Psychopathy Checklist,” “18-item Psychopathy Checklist Revised,” “Psychopathy Checklist: Youth Version,” (Forth, Hart, & Hare, 1990), “PCL,” “PCL-R,” “18-item PCL,” “18-item PCL-R” and “PCL:YV.”

To be included in this meta-analysis, studies had to (1) use either the PCL:YV or the adult PCL measure with the ‘youth modifications’ recommended by Forth et al. (1990) with a primarily adolescent sample (i.e., mean age of less than 19 years); (2) employ a predictive research design in which the PCL measures were scored independent of the recidivism information (i.e., we did not include retrospective studies where PCL scores were used to “postdict” criminal history variables); and (3) present sufficient statistical information to code an effect size. In some instances, we were able to augment published accounts that lacked sufficient information to compute an effect size by obtaining the original dissertations or theses that provided greater detail than the journal article or submission (e.g., Brandt, 1993; McKinnon, 2004).

Procedure

The Hare scales (modified PCL, PCL-R, or PCL:YV) were coded for the total and (where available) two factor scores historically associated with the PCL measures (Hare, 1991). Several studies unfortunately did not report analyses related to the two factors. We were able to obtain this information from the study authors in most but not all instances (correspondence on file with the first author). Although the PCL:YV manual (Forth, Kosson, & Hare, 2003) reports analyses related to three and four-factor models, at the time of our literature review little published research was identified that examined the predictive validity of the three- and four-factor models.

The criterion measures were coded for the three forms of recidivism noted above—general, violent, and sexual, along with the base rate for each category. General recidivism referred to an outcome either described in the primary study as “general” or undifferentiated from violent or sexual recidivism. Effects coded as “violent” or “sexual” were expressly described as such in the primary study. Additional study characteristics were coded to provide a descriptive overview of the sample of studies as well as to investigate effect size heterogeneity. These characteristics included PCL descriptive statistics (means, standard deviations), scoring method

(i.e., file review only versus file review plus interview), basic demographic information (e.g., ethnic and gender composition), type of institution from which the data were collected (e.g., juvenile prison), country in which the data were collected, length of follow-up, and the source of the study (e.g., peer-reviewed journal, unpublished dissertation). Each study was coded by two raters, resulting in an overall rate of agreement of 94%. Discrepancies were resolved by consensus.

Effect size coding

A correlation coefficient was coded for all effects because r is widely understood, the statistical procedures for its aggregation are well established (Hedges & Olkin, 1985; Rosenthal, 1991), and the majority of original studies reported point biserial correlations. When correlations had to be calculated from other statistics (e.g., AUC values) formulas provided by Lipsey and Wilson (2001) were used. The correlations were transformed to Fisher's z scores (z_r) via SPSS syntax. The effects were aggregated using the SPSS macro "MeanES" (available online at <http://mason.gmu.edu/~dwilsonb/ma.html> and in Lipsey and Wilson (2001)). The macro computes basic central tendency statistics according to the procedures outlined by Hedges and Olkin (1985). Each z_r is weighted¹ by its inverse variance to create a summary statistic, z_+ . The macro also computes the mean weighted effect size and calculates 95% confidence intervals for both random and fixed effects models. Because of concerns about fixed effects models (Field, 2003), we report only the random effects results below.

Heterogeneity was examined using the Q statistic (Hedges & Olkin, 1985), which is distributed as a chi-square with $k - 1$ degrees of freedom (where k equals the number of studies). Given that Q is dependent on the number of studies included in a meta-analysis and can be severely distorted with small k values, we also report the I^2 statistic (Higgins & Thompson, 2002). I^2 is independent of k and estimates the percentage of total variation among studies due to heterogeneity rather than sampling error.

Moderator analyses

Moderator analyses were performed using SPSS macros developed by Lipsey and Wilson (2001). Analyses of categorical moderator variables (e.g., publication status, gender) were conducted using a meta-analytic analog to an analysis of variance performed by the "MetaF" macro (Lipsey & Wilson, 2001). MetaF divides the total variability among the effects into the portion that can be explained by the categorical predictor variable and the residual portion (Q_B and Q_W , respectively). A significant Q_B indicates that the difference between the means of the categories is greater than what would be expected simply by sampling error (Lipsey & Wilson, 2001, pp. 135–136). For continuous moderator variables, we used the SPSS macro "MetaReg" (Lipsey & Wilson, 2001; pp. 138–140) to conduct a modified weighted least squares regression. MetaReg creates homogeneity statistics for the regression model (Q_R) and for the sum-of-squares residual (Q_E).

¹ The need to weight effect sizes is widely accepted in the meta-analytic literature, with the inverse variance weight being considered a superior method than simply weighting by sample size (see, e.g., Lipsey & Wilson, 2001, pp. 36–37; see also Hedges & Olkin, 1985). In the case of correlations, the formula to compute the inverse variance reduces to $n - 3$ (Lipsey & Wilson, 2001, p. 64). We should note, however, that the mean effect sizes reported later in this article were highly consistent regardless of whether they were weighted or unweighted.

Results

Study characteristics

As of July 1, 2005, our iterative search process led to the identification and coding of 21 published and unpublished studies (see Table 1) with independent samples that met the selection criteria described above. These studies represented a total nonredundant sample size of 2,867 cases ($M = 137$, $SD = 83$, range = 60–408) available for follow-up. Studies were completed between 1990 and 2005, with the median year being 2002. Ten studies were published or *in press* in peer-reviewed journals. Of particular note, 11 of the 21 studies were not included in the review of recidivism research in the PCL:YV manual (Forth, Kosson, & Hare, 2003); see Chapter 7), reflecting the large amount of recent work in this area. The studies were split relatively evenly between United States ($k = 9$) and Canadian samples ($k = 11$), with one additional study from Sweden. The average of the mean ages reported was 16 years ($SD = 1.10$; range = 14.60–18.40). Ethnic distributions varied considerably, with the percentage of participants identified as Caucasian ranging from 12–83% (two missing data). Not surprisingly, given the historical focus on psychopathy in male offenders, 15 of the samples were 100% male, whereas five were of mixed gender and only one was composed entirely of female offenders. Fortunately, we were able to obtain information to compute effect sizes separately for boys and girls for all of the studies with a mixed gender composition.

Sample descriptive statistics on the psychopathy measures also are reported in Table 1. The weighted average of the mean sample scores was 20.50, although it varied substantially across studies (range = 9–28). In the majority of studies (16) the PCL measures were scored based on file review only. Interrater reliability statistics were reported in all but two studies (Langstrom & Grann, 2002; Rowe, 2002) and were uniformly high (range = .83 to .99). In terms of criterion measures, we were able to code effects for general recidivism in all studies,² violent recidivism in 15 studies, and sexual recidivism in four studies. For the factor scores, we were able to obtain and code effects for general, violent, and sexual recidivism for 16, 13, and 3 studies, respectively. For the total sample of studies, the average length of follow-up was 31.75 months (range = 3–120 months), although two studies were missing this information (Jack, 2000; Marczyk, Heilbrun, Lander, & DeMatteo, 2003). For those studies reporting outcomes for violent recidivism, the average follow-up period was somewhat longer ($M = 40.69$ months, range = 12–120 months; data missing for one study). For the four studies examining sexual recidivism, the average number of months follow-up was even longer ($M = 65.62$ months, range = 34–120 months).

Because all but two studies (Jack, 2000; Ridenour, Marchant, & Dean, 2001) treated recidivism as a dichotomous outcome, the base rate of the criterion measures (here the percentage of the sample that recidivated) is a source of concern because of the potential impact this might have on the magnitude of the effect sizes. Although base rates below 20% or above 80% can result in distorted effects (Lipsey & Wilson, 2001), this does not appear to be a serious problem in relation to general and violent recidivism because the vast majority of base rates reported in the present study are within this range, although a few studies did not report these statistics. Despite the longer average follow-up periods reported, the base rates of sexual recidivism were very low for all four studies that examined this outcome (ranging from 8 to 15%), which significantly

² In three cases, effect sizes were available for violent and nonviolent recidivism, but not for general recidivism specifically. In these instances, we computed the mean effect size for these two categories to serve as a proxy for general recidivism. Dropping these three studies from the general recidivism analyses ($k = 18$) had virtually no impact on the mean effect size or the heterogeneity statistics that are reported in Table 2.

attenuates the magnitude that a point biserial correlation can achieve. Given these concerns, we performed the meta-analysis for sexual recidivism (described later) using both point biserial correlations and standardized mean difference effect sizes, which are in some regards a more appropriate metric for low base rate criterion measures because the two groups are contrasted by their respective mean scores (Lipsey & Wilson, 2001, pp. 48, 49, 62).

Meta-analytic results

Our initial meta-analytic results indicated that the mean weighted effect sizes (r_w) for the PCL Total score for general, violent, and sexual recidivism were .26 ($p < .001$), .23 ($p < .001$), and .07 ($p = ns$), respectively. The Q statistics, however, were very large both for general (44.77) and violent (44.01) recidivism. Similar to results from the adult literature (Guy et al., 2005; Walters et al., 2003b) and the juvenile institutional adjustment literature (Edens & Campbell, 2006), this indicates that there was significant variability among the effects. This interpretation is supported by the very large values obtained for I^2 (general = .55; violent = .68), which indicate that heterogeneity explained a considerable amount of variance in most of these effect sizes. Higgins and Thompson (2002) tentatively suggested that I^2 values $\geq .3$ should be considered indicative of “moderate” heterogeneity and those $\geq .5$ should be considered “severe.”

Given this variability, we examined the distribution of effect sizes to determine whether there might be significant outliers in the data set. For general recidivism, the effect size for the total score reported by Ridenour et al. (2001) was more than three standard deviations above the weighted mean effect size of the entire sample. This is not entirely surprising, as Ridenour et al. (2001) were the only researchers to combine juveniles involved in the justice system with community controls (13 emotionally handicapped and 39 students participating in social studies classes), an extreme groups design that may have artificially inflated variability in PCL scores and increased the likelihood of obtaining an exaggerated effect size relative to the other studies. Removal of this one study resulted in a substantial decrease in I^2 (from .55 to .45) for the total score, although considerable heterogeneity was still evident.

For violent recidivism, the total score effect size for Campbell (2003) was more than two standard deviations below the weighted mean effect size for the entire sample. Of note, the base rate of violent recidivism for this sample was the lowest of all studies as well. Removal of this effect size resulted in a very large decrease in I^2 (from .68 to .32), although heterogeneity continued to be in the moderate range. Given these results, each outlying effect size was removed from the subsequent analyses reported below for general and violent recidivism.³

Table 2 reports the mean weighted effect sizes (r_w) and 95% confidence intervals (CIs) for the PCL Total score for the three outcome categories. As can be seen, after removing the outliers the effect sizes for general and violent recidivism were relatively similar (.24 and .25, respectively), with both being significantly greater than chance. The analysis for sexual recidivism, noted above, fell just short of significance ($p = .07$), with an overall magnitude that would be considered small. When analyses for sexual recidivism were run using the standardized mean difference values, it resulted in a mean weighted effect size (ES_{sm}) of .21 (95% CI = .02 to .40) that also would be considered relatively modest, although it was statistically significant ($p = .03$). Table 2 also

³ We should note that there is debate in the field regarding how best to address outlier effect sizes in meta-analyses, with some commentators expressing concern over the exclusion of such studies (see Lipsey & Wilson, 2001, pp. 107–108 for a discussion of this issue). Given some of the methodological concerns noted above regarding both the Ridenour et al. (2001) and Campbell (2003) studies, however, we believed it was appropriate to remove these effect sizes from the subsequent analyses.

Table 1 Overview of studies included in the meta-analysis

Author (s)	<i>n</i>	Country	Status	Male (%)	White (%)	Mean age	Mean follow-up (months)	PCL <i>M</i> (<i>SD</i>)	Scoring method	Base rate	<i>r</i>
Forth, Hart, and Hare, 1990	71	Canada	P	100	77	16	27	24 (6.8)	F & I	(a) .79 (b) – (c) –	(a) .14 (b) .26 (c) –
Brandt, Kennedy, Patrick, and Curtin, 1997	127	US	P	100	28	16	21	25 (5.3)	F	(a) .88 (b) – (c) –	(a) .19 (b) .16 (c) –
Auslander, 1998	124	US	D	100	44	NR	34	28 (NR)	F	(a) – (b) .41 (c) .08	(a) .20 (b) .23 (c) –.03
Dodds, 1999	117	US	D	100	12	16	32	24 (6.0)	F	(a) .78 (b) – (c) –	(a) .22 (b) – (c) –
Cruise, 2000	105	US	D	100	67	15	3	14 (6.9)	F & I	(a) – (b) – (c) –	(a) .36 (b) – (c) –
Jack, 2000	150	Canada	D	100	69	NR	NR	22 (6.0)	F	(a) .73 (b) .15 (c) –	(a) .13 (b) .09 (c) –
Gretton, McBride, Hare, O'Shaughnessy, and Kumka, 2001	220	Canada	P	100	65	16	55	21 (7.3)	F	(a) .51	(a) .25
McEachran, 2001	108	Canada	T	100	85	15	120	19 (7.8)	F	(b) .30 (c) .15	(b) .19 (c) .09
										(a) .33 (b) .66 (c) –	(a) .43 (b) .46 (c) –

Table 1 Continued

Author (s)	<i>n</i>	Country	Status	Male (%)	White (%)	Mean age	Mean follow-up (months)	PCL <i>M</i> (SD)	Scoring method	Base rate	<i>r</i>
Ridenour, Marchant, and Dean, 2001	80	US	P	100	81	15	12	14 (7.0)	F*	(a) .25 (b) – (c) –	(a) .62 (b) .48 (c) –
Langstrom and Grann, 2002	85	Sweden	P	100	NR	18	24	20 (7.1)	F	(a) .20 (b) .37 (c) –	(a) .20 (b) .22 (c) –
Rowe, 2002	408	Canada	P	80	NR	15	25	19 (8.7)	F	(a) .63 (b) .23 (c) –	(a) .41 (b) .27 (c) –
Catchpole and Gretton, 2003	67	Canada	P	85	55	18	12	24 (6.9)	F	(a) .58 (b) .23 (c) –	(a) .30 (b) .30 (c) –
Marczyk, Heilbrun, Lander, and DeMatteo, 2003	72	US	P	100	13	18	NR	9 (3.9)	F	(a) .77 (b) – (c) –	(a) .04 (b) – (c) –
O'Neil, Lidz, and Heilbrun, 2003	64	US	P	100	20	18	12	25 (5.6)	F	(a) – (b) – (c) –	(a) .33 (b) – (c) –
Campbell, 2003	226	Canada	D	83	83	16	19	15 (6.0)	F	(a) .71 (b) .12 (c) –	(a) .13 (b) –.11 (c) –
Gretton, Hare, and Catchpole, 2004	157	Canada	P	100	79	17	120	23 (7.1)	F	(a) .96 (b) .68 (c) .11	(a) .26 (b) .32 (c) .11
Parks, 2004	156	US	D	100	63	15	53	16 (9.5)	F	(a) .39 (b) – (c) .08	(a) .35 (b) – (c) .09

Table 1 Continued

Author (s)	<i>n</i>	Country	Status	Male (%)	White (%)	Mean age	Mean follow-up (months)	PCL <i>M</i> (<i>SD</i>)	Scoring method	Base rate	<i>r</i>
Douglas, 2005 ^a	84	US	R	100	79	16	29	23 (5.9)	F & I	(a) .44 (b) .24 (c) –	(a) .08 (b) .17 (c) –
Schmidt, McKinnon, Chattha, and Brownlee, 2005	130	Canada	M	63	73	15	36	22 (6.7)	F	(a) .47	(a) .40
Vincent, Odgers, Watkins, and Corrado, 2005 ^b	256	Canada	M	79	79	18	36	22 (6.7)	F & I	(b) .27 (c) –	(b) .35 (c) –
Odgers, Repucci, and Moretti, 2005	60	Canada	P	0	37	16	8	25 (5.2)	F & I	(a) .21 (b) – (c) –	(a) .07 (b) – (c) –

Note. (a) General; (b) Violent; (c) Sexual. D = Unpublished doctoral dissertation. M = Manuscript submitted for publication. P = Article published or accepted for publication. R = Unpublished raw data. T = Unpublished master's thesis. F = Psychopathy scored based on file information. F & I = Psychopathy scored based on file information and interview data. F* = File reviews plus interviews conducted for some but apparently not all cases. NR = Exact value not reported.

^aThis is community follow-up data for the sample of participants originally described in a study by Spain et al. (2001), where their institutional adjustment was examined.

^bThis is an expanded follow-up study to Corrado et al. (2004) that includes female participants.

reports the mean weighted effect sizes (r_w) and 95% CIs for the three criterion measures in relation to the two factors that historically have been associated with the PCL measures. It is notable that Factor 2 was a somewhat stronger correlate of general recidivism than was Factor 1, although the 95% CIs overlapped slightly. Results for violent recidivism were less disparate, however.

Notwithstanding the significant effects for general and violent recidivism in relation to psychopathy total and factor scores, the I^2 statistics reported in Table 2 continued to suggest moderate to severe heterogeneity. This indicates that there continued to be significant variability among the effects even after the removal of outliers and that an examination of potential moderator variables was warranted. Noted earlier, four categories of moderator variables (i.e., indicators of methodological rigor, publication status, gender, ethnicity/nationality) were hypothesized as possible causes of heterogeneity. We examine each of these factors in the following section, using the SPSS “metaF” and “metareg” macros described earlier to analyze their impact on effect sizes.

Moderator analyses

Methodological quality

Because of concerns regarding the potential impact of methodologically weaker studies on the results of meta-analyses (Lipsey & Wilson, 2001), as an initial step in identifying possible sources of heterogeneity we first examined whether a number of potentially important study design characteristics might explain a significant amount of variability in the effect sizes. The specific study design characteristics examined were (a) whether the PCL measures were scored using an interview in addition to file data, (b) whether internal consistency statistics were reported for the PCL measures, (c) the degree of variability evidenced on the predictor measures (operationalized by the total score standard deviation reported for each sample) and criterion measures (operationalized by the degree to which studies deviated from the optimal recidivism base rate of 50%), and

Table 2 Weighted mean effect sizes of psychopathy checklist scores as predictors of recidivism

Recidivism type	<i>k</i>	<i>n</i>	r_w	95% CI _w	<i>Q</i>	I^2
Total score						
General	20	2,787	.24	.19–.30***	34.82*	.45
Violent	14	2,067	.25	.20–.31***	19.08	.32
Sexual	4	654	.07	–.01–.15	1.59	.00
Factor 1						
General	15	2,157	.18	.13–.23***	19.84	.29
Violent	12	1,776	.19	.13–.26***	20.47*	.46
Sexual	3	437	.03	–.17–.22	8.39*	.76
Factor 2						
General	15	2,157	.29	.22–.36***	34.14**	.59
Violent	12	1,776	.26	.20–.33***	21.49*	.49
Sexual	3	437	.08	–.02–.17	.32	.00

Note. Analyses for general and violent recidivism were conducted after removing one outlier from each sample of studies.

* $p < .05$; ** $p < .01$; *** $p < .001$.

(d) the length of the follow-up period. Each of these factors conceivably could explain why the effect sizes were so variable across studies.⁴

Despite concerns regarding the potential impact of methodological rigor, no statistically significant results were obtained from these analyses. Somewhat counter-intuitively, results indicated that there were modest trends for effect sizes to be stronger in studies that did *not* include an interview in the scoring of the PCL measures. There was also a nonsignificant trend for effect sizes to be associated with the degree of variability on the PCL measures. There was no relationship between effect sizes and the base rate of general or violent recidivism, although there were trends for the length of the follow-up period to be associated with violent (but not general) recidivism effect sizes. (Further detail regarding these analyses is available from the first author upon request.)

Publication status

Noted earlier, file drawer effects are a common source of concern in relation to the interpretation of meta-analytic findings (Rosenthal & DiMatteo, 2001), and previous narrative (Edens et al., 2001, 2003) and some quantitative reviews (Edens & Compbell, 2005) have noted appreciably weaker effects in unpublished studies in the juvenile psychopathy literature. As such, to examine whether the variability in both the general and violent recidivism effect sizes could be explained by publication status, a meta-analytic analog to an analysis of variance was performed. This analysis failed to offer any support for a file drawer effect for the total score, however, with Q_B being small and nonsignificant for both general (1.30) and violent recidivism (.13). Similarly trivial and nonsignificant results were obtained for the factor scores as well.

Although we were as exhaustive as possible in our pursuit of unpublished research to include in this study, the file drawer effect is always a source of concern in relation to the identification of studies for inclusion in meta-analyses. To examine publication issues more broadly, we computed fail-safe N s using a formula described by Orwin (1983). A fail-safe N reflects the number of unpublished studies with an average effect size of .00 that would need to be located to reduce the obtained mean effect size to some predefined lower value. We computed the fail-safe N needed to reduce the significant weighted correlations reported in Table 2 to a value of .10, which is typically is considered the lower boundary of a “small” effect (Cohen, 1988). The fail-safe N s for the total score mean effects were 28 and 21 for general and violent recidivism, respectively. Also, the fail-safe N s for Factor 2 mean effects were 29 and 19 for general and violent recidivism respectively, whereas they were 12 and 11 for Factor 1.

Gender differences

Next, to examine whether heterogeneity could be explained by hypothesized gender differences in the predictive validity of the PCL measures, we performed a meta-analytic analog to an analysis of variance for the male and female samples. For general recidivism, mean effect sizes for the six female samples (combined $n = 288$) were relatively comparable to the male

⁴ For example, scoring the PCL measures without interviewing the juveniles might result in less valid scores that attenuate predictive validity (see Hare, 2003; for a discussion of concerns regarding file-only PCL scoring). Similarly, coefficient alpha might be lower among studies that neglected to report this information. Because the magnitude of validity coefficients is constrained by the reliability of a scale, this also might result in weaker effect sizes in these studies. Finally, severe range restriction on the PCL scores in a sample also has the potential to attenuate the magnitude of the obtained effect size, as can extreme deviations from an optimal recidivism base rate of 50%.

Table 3 Weighted mean effect sizes of PCL total scores as predictors of recidivism by gender

Gender	<i>k</i>	<i>n</i>	<i>r_w</i>	95% CI _w	<i>Q</i>	<i>I</i> ²
General recidivism						
Male	19	2,482	.25	.20–.30***	30.51*	.41
Female	5 ^a	207	.13	–.02–.29	.41	.00
Violent recidivism						
Male	14	1,870	.26	.20–.31***	20.24	.36
Female	5	228	.10	–.05–.24	2.62	.00

Note. PCL Total = Total scores from the Psychopathy Checklist measures.

^aThis analysis was conducted after removing one outlier study (Rowe, 2002).

* $p < .05$; ** $p < .01$; *** $p < .001$.

samples (.24 vs. .26), with Q_B being nonsignificant (.01). The results for the female samples were themselves exceedingly heterogeneous ($I^2 = .57$), however, primarily due to the inclusion of one outlier study (Rowe, 2002) that reported an exceptionally large r of .58. In contrast, the remaining five studies reported correlations ranging only between .07 and .20. After eliminating the outlier effect for female offenders, the resultant weighted mean effect size was much lower, and showed no evidence of heterogeneity (summary statistics are reported in Table 3). As can be seen by the values reported for Q and I^2 in Table 3, there continued to be a moderate degree of heterogeneity among the effects for male juveniles, however.

For violent recidivism, the moderator analysis for gender was significant, $Q_B = 4.37$, $p < .05$. The effect size for female offenders⁵ was not statistically significant, was quite homogeneous, and was appreciably lower than that of the male samples (see Table 3). Because the base rates of violent reoffending for female adolescents were relatively low overall (range = .14 to .27), we re-analyzed the data using standardized mean difference effect sizes. This also resulted in a nonsignificant weighted effect ($ES_{sm} = .29$), although the results were quite heterogeneous ($I^2 = .79$). In contrast, the weighted effect size for the male samples summarized in Table 3 was much larger, but also displayed moderate heterogeneity.

Ethnicity/nationality

We also hypothesized that country of origin might play a moderating role, given findings in the adult literature that effects in US studies are weaker than in Canadian and European samples when the criterion is institutional aggression (Guy et al., 2005). As such, moderator analyses were again conducted using MetaF (Lipsey & Wilson, 2001) to compare the magnitude of the effects for U.S. versus the remaining studies. These failed to offer any support for a nationality effect for the total score, however, with the Q_B values being small and nonsignificant for both general (.22) and violent recidivism (.00). In fact, the mean effect sizes were almost identical across countries. Virtually identical effects were noted for the factor scores as well.

Noted earlier, some studies have suggested that there may be important ethnic group differences in the predictive validity of psychopathy (e.g., Guy et al., 2005; Hicks et al., 2000). Therefore, to examine whether the heterogeneity of the observed effects could be attributed to ethnicity we conduct a modified weighted least squares regression, with the percentage of

⁵ When the samples were split by gender, it was clear that the outlying low effect (Campbell, 2003) resulted from the male participants in this sample, where the effect size was three standard deviations below the weighted mean and the base rate of violent reoffending was only .08. The effect size for the female sample was quite similar to the other female samples, however, so we retained this small sample ($n = 34$) in the moderator analyses for gender.

Table 4 Weighted mean effect sizes for psychopathy and the YLS/CMI as predictors of recidivism

Measure	<i>k</i>	<i>n</i>	<i>r_w</i>	95% CI _w	<i>Q</i>	<i>I</i> ²
General recidivism						
PCL Total	5	799	.27	.12–.42***	14.63**	.73
YLS/CMI	5	799	.25	.10–.39***	14.06**	.72
Violent recidivism						
PCL Total	4	727	.24	.14–.35***	5.07	.41
YLS/CMI	4	727	.21	.12–.31***	4.20	.29

Note. YLS/CMI = Youth Level of Service/Case Management Inventory (Hoge & Andrews, 2002). PCL Total = Total scores from the Psychopathy Checklist measures.

* $p < .05$; ** $p < .01$; *** $p < .001$.

white participants in each sample used as the predictor variable. For general recidivism, the percentage of white participants in each sample did not explain a significant amount of variance in the effect sizes for the total ($Q_R = .69$, $p = ns$, $R^2 = .04$, $\beta = .20$), Factor 1 ($Q_R = .04$, $p = ns$, $R^2 = .00$, $\beta = .06$), or Factor 2 ($Q_R = .55$, $p = ns$, $R^2 = .05$, $\beta = .21$) scores ($k = 18$, 13, and 13, respectively).

For violent recidivism, however, ethnicity moderated effect sizes to a significant degree for Factor 2 ($Q_R = 5.57$, $p < .05$, $R^2 = .40$, $\beta = .63$, $k = 10$) and approached statistical significance for the total score ($Q_R = 2.87$, $p = .09$, $R^2 = .22$, $\beta = .47$, $k = 12$). That is, the higher the proportion of non-White juveniles, the weaker the relationship between violent recidivism and psychopathy. Results for Factor 1 were not statistically significant ($Q_R = 1.41$, $p = ns$, $R^2 = .15$, $\beta = .39$, $k = 10$), although the amount of variance explained was relatively large. All of these analyses should be interpreted cautiously, given the relatively small number of studies.⁶

Comparisons with the YLS/CMI

Our literature search identified five studies (Catchpole & Gretton, 2003; Jack, 2000; Marczyk et al., 2003; McKinnon, 2004; Rowe, 2002⁷) that compared psychopathy to the YLS/SCM (Hoge & Andrews, 2002). Each study examined general recidivism (combined $n = 799$) and all except Marczyk et al. (2003) also reported analyses on violent recidivism (combined $n = 727$). Table 4 summarizes the results of these comparisons. As can be seen, both instruments performed similarly, although there was considerable heterogeneity among the general recidivism effects. This variability was primarily explained by the much lower correlations reported in the Marczyk

⁶ Even though methodological factors did not explain a significant amount of heterogeneity in the effect sizes, we were curious as to whether several of the nonsignificant trends noted for some of these factors might account for much of the same heterogeneity in violent recidivism that was explained by the ethnicity variable. We therefore conducted exploratory regression analyses to examine whether ethnicity would continue to explain a significant amount of heterogeneity after controlling for these factors. Only one analysis suggested that the moderating role of ethnicity was substantially reduced after controlling for a methodological variable. Noted earlier, there was a non-significant trend for greater variability on the PCL total score (i.e., larger *SDs*) to moderate effect sizes for violent recidivism ($Q_R = 2.66$, $p = .10$). When the regression analysis was re-run including the ethnicity variable along with the *SD* variable ($k = 11$), the overall model was significant ($Q_R = 6.50$, $p < .05$) and R^2 increased substantially to .46. Neither the beta for the *SD* variable nor the ethnicity variable was significant individually, however. Further detail regarding these analyses is available from the first author.

⁷ This is the same data set that is reported in Schmidt et al. (2005), except that the sample is slightly smaller ($n = 102$). Schmidt et al. did not include analyses related to the YLS/CMI but these results were included in the original McKinnon thesis.

et al. (2003) study for both instruments.⁸ Unfortunately, it was not possible to code for whether either measure demonstrated any incremental validity (see, generally, Hunsley, 2003) over the other in the prediction of recidivism, as analyses that would allow us to address systematically this question were not reported in the original studies. It seems unlikely that the measures were accounting for much if any unique variance in these outcomes, however, because they were highly intercorrelated ($r_w = .77$) in the three studies (combined $n = 625$) that reported their association with each other.

Discussion

The real-world implications of being labeled “psychopathic” are considerable and they appear to be growing in our society. Those identified as psychopaths increasingly may be subjected to serious deprivations of liberty (Edens & Petrila, 2006; Zinger & Forth, 1998) and denied treatment services (D’Silva, Duggan, & McCarthy, 2004), at least in part because of expectations that such individuals are more dangerous and intransigent than the typical offender. As such, the results of the current analysis are informative, both in terms of what we currently do and do *not* know about the relationship between psychopathy and future criminality among youths involved in the justice system. On the one hand, much like the adult literature (Gendreau, et al., 2002; Hemphill et al., 1998; Walters, 2003a,b) the aggregate results suggest that the relationship between psychopathy and both general and violent recidivism among male adolescents is statistically significant and of a magnitude that borders on what Cohen (1988, p. 82) conventionally would define as a “medium” effect.⁹ On the other hand, the moderate to severe heterogeneity observed among the obtained effects indicates a lack of consistent results across studies and compels an analysis of possible moderators that might account for this variability. Practically speaking, such heterogeneity also raises concerns about whether and how to interpret these “aggregate” effect sizes in relation to specific contexts and populations, as well as in relation to individual cases (Douglas, Vincent, & Edens, 2006; Edens, 2006). Moreover, the magnitude of these effects, despite being significant, indicates the vast majority of variability in recidivism remains to be explained by factors other than psychopathy.

Analysis of potential moderators indicated that the heterogeneity among the effects could be attributed to some extent to the gender and ethnic composition of the samples, with appreciably weaker effects being noted for female juveniles and samples with higher percentages of non-Caucasian participants. Results for female adolescent offenders in particular were rather discouraging, although they should be considered preliminary due to the small number of studies. Only one study (Rowe, 2002) produced a large effect size, and only for general recidivism. The remaining results raise questions about the relevance of psychopathy in relation to decision-making for female adolescents, at least in relation to risk assessment issues. Although it is possible that future studies will provide greater support for the utility of psychopathy in this regard, at present it would be difficult to defend ethically the use of the PCL:YV with female adolescents for risk assessment purposes. As indicated in principle 9.02 of the Ethical Principles of Psychologists and Code of Conduct, psychologists use assessment instruments “whose validity and reliability have been established for use with members of the population tested” (American Psychological Association, 2002, p. 1071). Examiners would be better served by

⁸ Although we were unable to code an exact effect size, Marczyk (2002) noted that neither psychopathy nor the YLS/CMI was significantly correlated with violent offenses.

⁹ The conventional values suggested for *point biserial correlations* such as those that predominate this study are .10 (small), .24 (medium), and .37 (large).

incorporating other known risk factors with a stronger empirical relationship with recidivism (cf. [Nicholls & Petrila, 2005](#)).

It is interesting that the effects for adult women have been somewhat stronger regarding criminal behavior (e.g., [Nicholls, Ogloff, & Douglas, 2004](#)), although there is not a large corpus of research that has addressed this issue to date. If ultimately reflective of a genuine and substantial difference in predictive validity across adolescent and adult females, it would raise numerous pragmatic and theoretical questions regarding the measurement of psychopathic traits in female juvenile offenders. For example, it might be that the putative developmental instability of such traits ([Edens et al., 2001](#); [Seagrave & Grisso, 2002](#)) makes it more difficult to identify during adolescence those female offenders who are more likely to continue to engage in serious criminal conduct. The measurement of an ostensibly more stable personality disposition among adult women, however, may make it easier to identify those women who are relatively more likely to commit future crimes.

In the case of ethnicity, it is possible that the moderating effect noted here for violent recidivism is a spurious finding related to the variability of the psychopathy scores across samples (i.e., more ethnically diverse samples had less variability in their PCL scores). Although these moderator analyses must be viewed as preliminary due to the small number of studies and the relatively simplistic means by which ethnicity was operationalized, the amount of variability explained among the effect sizes was rather large (R^2 ranging from .15 to .40 for the total and factor scores). These results also suggest that examiners should be exceedingly leery of making generalizations about the relationship between psychopathy and criminality among minority youth until more sophisticated research designs can be employed to better ascertain the nature of this relationship. If there is a clear “slope bias” (i.e., weaker validity coefficients among non-Caucasian juvenile samples) then our field will have to reconsider the appropriateness of using psychopathy to inform legal decision-making for ethnic minorities, in that many if not most uses of this construct in the legal arena seem to revolve around its predictive validity ([DeMatteo & Edens, 2005](#); [Edens & Petrila, 2006](#)).

The ethnicity effect reported here is interesting to contrast with what is known about group mean differences in psychopathy across ethnic groups. Although research in this area is in a nascent state, existing studies generally have failed to find meaningful group differences in PCL scores, at least among Black and White adult and adolescent offender samples (for meta-analytic overviews of this literature, see [McCoy & Edens, in press](#); [Skeem, Edens, Camp, & Colwell, 2004](#)). The studies included in the present meta-analysis included a much wider array of ethnic groups, however, which may explain these differing results. As well, the absence of a significant “intercept bias” on psychopathy measures in no way precludes the possible existence of a slope bias in relation to predictive validity ([Skeem et al., 2004](#)).

Aside from gender and ethnicity, analyses examining other potential moderator variables generally failed to explain a statistically significant degree of heterogeneity in the effect sizes for general and violent recidivism. All of the moderator analyses need to be viewed with caution, however, given the small k and generally underpowered nature of the findings. For example, although we failed to note significant effects for the use or absence of interviews in the scoring of the PCL measures, the analysis was based on a very small sample of studies that included interviews. Item response theory (IRT) analyses have suggested that there may be some minor item-level differences between scores based on file-only data versus those that also included interview data ([Hare, 2003](#)). Such results in principle might lead one to expect that file-only PCL ratings would be less valid, although we found no evidence to support that assertion here, at least in relation to predictive validity.

Although based on a small number of studies, our results related to sexual recidivism are informative in relation to clinical and legal decision-making. It is worth noting that one additional

sample with youthful offenders (Langstrom & Grann, 2000) reported a weak and nonsignificant relationship between this outcome and psychopathy, although we were unable to code the exact effect size from this study for the present meta-analysis.¹⁰ To the extent that sexual re-offending among juveniles is a low base rate event, it appears that psychopathy is not likely to be a practically useful indicator of the likelihood of such behavior occurring in the future (along with most other predictor variables, it would seem). The results of our analyses using standardized mean differences rather than point biserial correlations, however, suggest that psychopathy *might* have some modest utility in samples where the base rate of this type of recidivism is closer to 50%, such as adolescent offenders who are repetitively sexually aggressive. The utility of psychopathy in comparison to other risk factors for sexual reoffending is unclear, although some studies have suggested other variables that may be more relevant to forecasting this particular type of crime (Langstrom & Grann, 2000; Parks, 2004). Along these lines it is noteworthy that in the adult literature there is some consistent evidence supporting the utility of the interaction between psychopathy and indicators of deviant sexual arousal in the prediction of sexual recidivism (for an overview, see Hare, 2003). In relation to juveniles, however, the lone study that has examined this issue (Gretton, McBride, Hare, O'Shaughnessy, & Kumka, 2001) failed to replicate these results. The interaction of these two variables did significantly predict violent recidivism more generally, however.

Although limited somewhat by missing data, our findings suggest that Factor 2 may be somewhat more relevant than Factor 1 to the prediction of general recidivism for juveniles. This interpretation is qualified by the slight overlap between the 95% CIs for these variables, however, as well as the moderate to high degree of heterogeneity identified in these analyses. Nevertheless, the magnitude of the difference (r_w .18 vs .29) in relation to general recidivism is fairly consistent with other meta-analytic findings (e.g., Gendreau et al., 2002; Hemphill et al., 1998; Guy, Edens, Anthony, & Douglas, 2005; Walters, 2003) and, interestingly, was evident despite the reliance on file-only scoring of PCL items in most studies. IRT analyses have suggested that the distinction between Factor 1 and 2 items may be more likely to blur in the absence of interview information (see Hare, 2003, p. 75, for a more complete discussion).

An ongoing point of debate regarding the factor scores is the extent to which—given their moderate intercorrelation—Factor 1 can explain any unique variance in recidivism once the variance attributable to Factor 2 is statistically controlled. Because of the intrinsic limitations of meta-analytic techniques, there was no way to partial out the effects of each factor on the other in order to address this question directly here. At least some prior research has, however, suggested that Factor 1 accounts for little variance in violent behavior once Factor 2 is statistically controlled (Skeem & Mulvey, 2001).

In terms of the relative utility of the PCL measures, the results of the direct comparison of psychopathy to the YLS/CMI reflect something of a double-edged sword. On the one hand, the fact that a lone personality construct can perform similarly to a more comprehensive risk assessment measure is remarkable, and not without precedent in the adult literature (e.g., Edens, Skeem, & Douglas, *in press*; Hemphill et al., 1998). Moreover, the effect sizes reported here for general recidivism are actually quite a bit higher for the PCL measures than the aggregate effect sizes reported by Cottle, Lee, and Heilbrun (2001) for other actuarial risk scales used with adolescent offenders. That said, however, the comparability of the YLS/CMI and the PCL in this study also raises questions about the need to use a construct or instrument with the potential to be highly prejudicial with adolescents (Edens et al., 2003; cf. Murrie et al., 2005) when other,

¹⁰ This was a subset of sex offenders who were included in the subsequent Langstrom and Grann (2002) paper, the results of which we were able to code for our general and violent recidivism analyses.

ostensibly less stigmatizing risk factors may work as well. What is needed in this regard is research examining the incremental validity (Hunsley, 2003) of psychopathy in relation to more elaborate risk factors to determine whether any unique variance is explained by this construct once the predictive utility of other measures is statistically controlled (Douglas et al., 2006). If psychopathy truly can “give us something more” than instruments such as the YLS/CMI then its use would be more defensible with white male adolescent offenders. Absent such data, however, one could argue that it is difficult to justify its introduction into juvenile risk assessment contexts at this time. We should note that this would not preclude its applied use for other purposes (e.g., treatment matching) if there were stronger empirical justification for such applications.

Aside from direct comparisons with other risk assessment instruments, it is also worth noting that further research is needed on the extent to which the PCL measures can improve on other variables that have a demonstrated relationship with recidivism among youths. For example, Cottle et al. (2001) identified a number of background and criminal history variables (e.g., age at first commitment, age at first contact with the law, family problems, conduct problems) with effect sizes that were comparable to or greater than those obtained here for the PCL measures and general recidivism (analyses for violent recidivism were not reported in their meta-analysis). Many of these factors are themselves significantly correlated with PCL scores (see Forth et al., 2003, Chapter 7)—as would be expected, given that they are historical variables that go into the scoring of some of the items comprising the scale. We are aware of very little research that has examined whether juvenile psychopathy measures can predict recidivism beyond these historical factors, which again raises concerns as to the incremental validity of the PCL measures in this regard. We should note, however, one large-scale study of adult psychiatric patients (Skeem & Mulvey, 2001) has suggested that Factor 2 continues to correlate with community violence even after controlling for a host of other risk factors.

Finally, limitations of this research should be noted, which hopefully can inform further research in this area. First, there was not a large number of studies available to include in the present meta-analysis, particularly in relation to samples of adolescent female offenders and effect sizes for sexual recidivism. Second, we did not examine the more recent three and four factor models of psychopathy that have come to the fore in recent years. More recent studies are beginning to include these factors in recidivism analyses and future meta-analyses hopefully will be able to address the relative contribution of these different models. A third possible limitation is that we made no attempt to control for measurement error in this study. This was because we were interested primarily in the performance of these instruments in error-filled “real-world” settings (i.e., “warts and all,” Lipsey & Wilson, 2001, p. 109) rather than the abstract relationship between hypothetical constructs. Although measurement error may result in effects that underestimate the “true” (i.e., error free) relationship between psychopathy and recidivism, in practice one can never eliminate measurement error in applied settings. Therefore, we would argue that clinicians using assessments of psychopathy should not base their opinions and risk statements about individual examinees on effects that attempt to control for measurement error, because such error is part and parcel of any clinical evaluation. A final limitation worth noting is that official criminal recidivism statistics such as those used in this study can be criticized because they almost certainly do not represent a complete index of criminal behavior. In some regards these indicators probably underestimate such conduct because of undetected crimes, whereas they may also overestimate such behavior in cases of wrongful arrests/convictions.

Despite these caveats, we believe the present findings are informative on several fronts. In the relatively small number of studies currently available for analysis, psychopathy was significantly associated with both general and violent recidivism among male youths. These general conclusions are substantially tempered, however, by the moderate to severe degree of heterogeneity

noted among the effect sizes, the very modest effects for female offenders and for sexual re-offending, and the possibility that psychopathy may be less predictive among ethnically diverse samples of juvenile offenders, at least in relation to violent recidivism. It is our hope that this quantitative summary of the available literature will spur further research into these important topics.

References

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

- American Psychological Association (2002). Ethical principles of psychologists and code of conduct. *American Psychologist*, 57, 1060–1073.
- *Auslander, B. A. (1998). *An exploratory study investigating variables in relation to juvenile sexual reoffending*. Unpublished doctoral dissertation, Florida State University.
- Banks, S., Robbins, P., Silver, E., Vesselinov, R., Steadman, H., Monahan, J., et al. (2004). A multiple-models approach to violence risk assessment among people with mental disorder. *Criminal Justice & Behavior*, 31, 324–340.
- Baum, K. (2005). *Juvenile victimization and offending, 1993–2003*. Bureau of Justice Statistics, Office of Justice Programs.
- Blumenthal, J. A. (2002). Law and social science in the twenty-first century. *Southern California Interdisciplinary Law Journal*, 12, 1–53.
- Borum, R., Bartel, P., & Forth, A. (2002). Manual for the Structured Assessment of Violence Risk in Youth (SAVRY). Tampa, FL: University of South Florida.
- *Brandt, J. R. (1993). *Assessment of psychopathy in a population of incarcerated adolescent offenders*. Unpublished doctoral dissertation, Department of Psychology, Florida State University, Tallahassee, FL.
- Brandt, J. R., Kennedy, W. A., Patrick, C. J., & Curtin, J. J. (1997). Assessment of psychopathy in a population of incarcerated adolescent offenders. *Psychological Assessment*, 9, 429–435.
- Cale, E., & Lilienfeld, S. (2002). Sex differences in psychopathy and antisocial personality disorder: A review and integration. *Clinical Psychology Review*, 22, 1179–1207.
- *Campbell, M. A. (2003). *The relative contribution of psychopathy and traditional risk factors in predicting young offender recidivism*. Unpublished doctoral dissertation, Department of Psychology, Dalhousie University, Halifax, Nova Scotia, Canada.
- Catchpole, R., & Gretton, H. (2003). The predictive validity of risk assessment with violent young offenders. *Criminal Justice and Behavior*, 30, 688–708.
- Cleckley, H. (1976). *The mask of sanity* (5th ed.). St. Louis, MO: Mosby.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cooke, D., Michie, C., Hart, S., & Clark, D. (2005). Searching for the pan-cultural core of psychopathic personality disorder. *Personality and Individual Differences*, 39, 283–295.
- Corrado, R., Vincent, G., Hart, S., & Cohen, I. (2004). Predictive validity of the psychopathy checklist: Youth version for general and violent recidivism. *Behavioral Sciences and the Law*, 22, 5–22.
- Cottle, C. C., Lee, R. J., & Heilbrun, K. (2001). The prediction of criminal recidivism in juveniles: A meta-analysis. *Criminal Justice and Behavior*, 28, 367–394.
- *Cruise, K. (2000). *Measurement of adolescent psychopathy: construct and predictive validity in two samples of juvenile offenders*. Unpublished doctoral dissertation, Department of Psychology, University of North Texas, Denton, Texas.
- DeMatteo, D., & Edens, J. F. (2005). *The role and relevance of the Psychopathy Checklist-Revised in court: A case law survey of United States courts (1991–2004)*. Manuscript submitted for publication.
- *Dodds, L. D. (1999). *Legislative exclusion statutes and psychopathy: Balancing rehabilitation ideals and retribution goals with juvenile offenders*. Unpublished doctoral dissertation, Allegheny University of Health Sciences.
- *Douglas, K. S. (2005) [Examining the predictive validity of the PCL:YV in a young offender sample]. Unpublished raw data.
- Douglas, K. S., Vincent, G., & Edens, J. F. (2006). Risk for criminal recidivism: The role of psychopathy. In C. Patrick (Ed.), *Handbook of psychopathy* (pp. 533–554). New York: Guilford.
- D'Silva, K., Duggan, C., & McCarthy, L. (2004). Does treatment really make psychopaths worse? A review of the evidence. *Journal of Personality Disorders*, 18, 163–177.

- Edens, J. F. (2001). Misuses of the Hare Psychopathy Checklist-Revised in court: Two case examples. *Journal of Interpersonal Violence*, 16, 1082–1093.
- Edens, J. F. (2006). Unresolved controversies concerning psychopathy: Implications for clinical and forensic decision-making. *Professional Psychology: Research and Practice*, 37, 59–65.
- Edens, J. F. & Campbell, J. S. (2005). *Identifying youths at risk for institutional misconduct: A meta-analytic investigation of the Psychopathy Checklist measures*. Submitted for publication.
- Edens, J. F., Guy, L. S., & Fernandez, K. (2003). Psychopathic traits predict attitudes toward a juvenile capital murderer. *Behavioral Sciences and the Law*, 21, 807–828.
- Edens, J. F., & Petrila, J. (2006). Legal and ethical issues in the assessment and treatment of psychopathy. In C. Patrick (Ed.), *Handbook of psychopathy* (pp. 573–588). New York: Guilford.
- Edens, J. F., Petrila, J., & Buffington-Vollum, J. K. (2001). Psychopathy and the death penalty: Can the Psychopathy Checklist-Revised identify offenders who represent “a continuing threat to society”? *The Journal of Psychiatry and Law*, 29, 433–481.
- Edens, J. F., Skeem, J. L., Cruise, K. R., & Cauffman, E. (2001). Assessment of ‘juvenile psychopathy’ and its association with violence: A critical review. *Behavioral Sciences and the Law*, 19, 53–80.
- Edens, J. F., Skeem, J. L., & Douglas, K. S. (in press). Incremental validity analyses of the violence risk appraisal guide and the psychopathy checklist: Screening version in a civil psychiatric sample. *Assessment*.
- Field, A. P. (2003). The problems in using fixed-effects models of meta-analysis on real-world data. *Understanding Statistics*, 2, 77–96.
- *Forth, A. E. (2004). *Psychopathy across the lifespan: Controversies and applications*. Plenary session presented at the annual meeting of the Association for the Treatment of Sexual Abusers, Albuquerque, NM.
- Forth, A. E., Hart, S. D., & Hare, R. D. (1990). Assessment of psychopathy in male young offenders. *Psychological Assessment*, 2, 342–344.
- Forth, A., Kosson, D., & Hare, R. (2003). *Psychopathy Checklist: Youth Version*. Toronto: MHS.
- Forth, A. E., & Mailloux, D. L. (2000). Psychopathy in youth: What do we know? In C.B. Gacono (Ed.), *The clinical and forensic assessment of psychopathy: A practitioner's guide* (pp. 25–54). Mahwah, NJ: Erlbaum.
- Frick, P. J., & Hare, R. D. (2001). *Antisocial Process Screening Device*. Toronto: MHS.
- Gendreau, P., Goggin, C., & Smith, P. (2002). Is the PCL-R really the “unparalleled” measure of offender risk?. *Criminal Justice and Behavior*, 29, 397–426.
- *Gretton, H., McBride, M., Hare, R., O'Shaughnessy, R., & Kumka, G. (2001). Psychopathy and recidivism in adolescent sex offenders. *Criminal Justice and Behavior*, 28, 427–449.
- *Gretton, H., Hare, R., & Catchpole, R. (2004). Psychopathy and offending from adolescence to adulthood. *Journal of Consulting and Clinical Psychology*, 72, 636–645.
- Guy, L. S., Edens, J. F., Anthony, C., & Douglas, K. S. (2005). Does psychopathy predict institutional misconduct among adults? A meta-analytic investigation. *Journal of Consulting and Clinical Psychology*, 73, 1056–1064.
- Hare, R. D. (1998). The Hare PCL-R: Some issues concerning its use and misuse. *Legal and Criminological Psychology*, 3, 99–119.
- Hare, R. D. (1991). *The Hare Psychopathy Checklist—Revised manual*. Toronto: MHS.
- Hare, R. D. (2003). *The Hare Psychopathy Checklist—Revised manual* (2nd ed.). Toronto: MHS.
- Hare, R. D., & Hervé, H. F. (1999). *Hare P-Scan: Research Version manual*. Toronto: MHS.
- Hart, S. D., Watt, K. A., & Vincent, G. M. (2002). Commentary on Seagrave and Grisso: Impressions of the state of the art. *Law and Human Behavior*, 26, 241–245.
- Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego: Academic Press.
- Hemphill, J. F., & Hare, R. D. (2004). Some misconceptions about the Hare PCL-R and risk assessment: A reply to Gendreau et al. *Criminal Justice and Behavior*, 31, 203–243.
- Hemphill, J. F., Hare, R. D., & Wong, S. (1998). Psychopathy and recidivism: A review. *Legal and Criminological Psychology*, 3, 139–170.
- Hicks, M., Rogers, R., & Cashel, M. (2000). Predictions of violent and total infractions among institutionalized male juvenile offenders. *Journal of the American Academy of Psychiatry and the Law*, 28, 183–190.
- Higgins, J., & Thompson, S. (2002). Quantifying heterogeneity in meta-analysis. *Statistics in Medicine*, 21, 1539–1558.
- Hoge, R., & Andrews, D. (2002). *Youth Level of Service/Case Management Inventory*. Toronto: MHS.
- Hunsley, J. (2003). Introduction to the special section on incremental validity and utility in clinical assessment. *Psychological Assessment*, 15, 443–445.
- Hunter, M. (1999). *How science takes stock*. Russell Sage Foundation.
- Hunter, J., & Schmidt, F. (1996). Cumulative research knowledge and social policy formulation: The critical role of meta-analysis. *Psychology, Public Policy, and Law*, 2, 324–347.
- *Jack, L. (2000). *Psychopathy, risk/need factors, and psychiatric symptoms in high-risk youth: Relationships between variables and their link to recidivism*. Unpublished doctoral dissertation, Department of Psychology, Simon Fraser University, British Columbia.

- Langstrom, N., & Grann, M. (2000). Risk for criminal recidivism among young sexual offenders. *Journal of Interpersonal Violence*, 15, 855–871.
- *Langstrom, N., & Grann, M. (2002). Psychopathy and violent criminal recidivism among young criminal offenders. *Acta Psychiatrica Scandinavica*, 106, 86–92.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Applied Social Research Methods Series (Vol. 49). Thousand Oaks, CA: Sage.
- Lynam, D. (2002). Fledgling psychopathy: A view from personality theory. *Law & Human Behavior*, 26, 255–259.
- Marczyk, G. R. (2002). *Predicting juvenile recidivism and validating juvenile risk factors in an urban environment*. Unpublished doctoral dissertation, Department of Clinical and Health Psychology, Hahnemann University, Philadelphia, PA.
- Marczyk, G. R., Heilbrun, K., Lander, T., & DeMatteo, D. (2003). Predicting juvenile recidivism with the PCL:YV, MAYSI, and YLS/CMI. *International Journal of Forensic Mental Health*, 2, 7–18.
- *Marczyk, G. R., Heilbrun, K., Lander, T., & DeMatteo, D. (2005). Juvenile decertification: Developing a model for classification and prediction. *Criminal Justice and Behavior*, 32, 278–301.
- McCoy, W. K., & Edens, J. F. (in press). Do black and white youths differ in levels of psychopathic traits? A meta-analysis of the Psychopathy Checklist measures. *Journal of Consulting and Clinical Psychology*.
- McEachran, A. (2001). The predictive validity of the PCL:YV and the SAVRY in a population of adolescent offenders. Unpublished master's thesis, Department of Psychology, Simon Fraser University, Burnaby, British Columbia, 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, School of Social Work, Lakehead University, Thunder Bay, Canada.
- Murrie, D. C., Cornell, D., & McCoy, W. K. (2005). Psychopathy, conduct disorder, and stigma: Does diagnostic labeling influence juvenile probation officer recommendations?. *Law and Human Behavior*, 29, 323–342.
- Nicholls, T. L., Ogloff, J. R. P., & Douglas, K. S. (2004). Assessing risk for violence among male and female civil psychiatric patients: The HCR-20, PCL:SV, and VSC. *Behavioral Sciences and the Law*, 22, 127–158.
- Nicholls, T. L., & Petrila, J. (2005). Psychopathy and gender: An overview of important issues and introduction to the special issue. *Behavioral Sciences and the Law*, 23, 729–741.
- *Oggers, C. L., Repucci, N. D., & Moretti, M. M. (2005). Nipping psychopathy in the bud: An examination of the convergent, predictive, and theoretical utility of the PCL:YV among adolescent girls. *Behavioral Sciences and the Law*, 23, 743–763.
- Ogloff, J., & Lyon, D. (1998). Legal issues associated with the concept of psychopathy. In D. J. Cooke, A. E. Forth, & R. D. Hare (Eds.) *Psychopathy: Theory, research, and implications for society* (pp. 401–422). Kluwer.
- *O'Neil, M., Lidz, V., & Heilbrun, K. (2003). Adolescents with psychopathic characteristics in a substance abusing cohort. *Law and Human Behavior*, 27, 299–313.
- Orwin, R. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, 8, 157–159.
- O'Shaughnessy, R. J. (2004). Violent adolescents: Psychiatry, philosophy, and politics. *Journal of the American Academy of Psychiatry and the Law*, 32, 12–20.
- Parks, G. (2004). *Juvenile sex offender recidivism: Typological differences in risk assessment*. Unpublished doctoral dissertation, Department of Psychology, Walden University.
- Petrila, J., & Skeem, J. L. (2003). An introduction to the special issues on juvenile psychopathy and some reflections on the current debate. *Behavioral Sciences and the Law*, 21, 689–694.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.
- *Ridenour, T. A., Marchant, G. J., & Dean, R. S. (2001). Is the revised Psychopathy Checklist clinically useful for adolescents?. *Journal of Psychoeducational Assessment*, 19, 227–238.
- Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, 52, 59–82.
- *Rowe, R. (2002). *Predictors of criminal offending: Evaluating measures of risk/needs, psychopathy, and disruptive behaviour disorders*. Unpublished doctoral dissertation, Department of Psychology, Carleton University, Ottawa, CA.
- Salekin, R. T., Rogers, R., & Sewell, K. W. (1996). A review and meta-analysis of the Psychopathy Checklist and Psychopathy Checklist-Revised: Predictive validity of dangerousness. *Clinical Psychology: Science and Practice*, 3, 203–215.
- *Schmidt, F., McKinnon, L., Chattha, H., & Brownlee, K. (2005). *Validity of the Psychopathy Checklist: youth version across gender and ethnicity*. Manuscript submitted for publication.
- Seagrave, D., & Grisso, T. (2002). Adolescent development and the measurement of juvenile psychopathy. *Law and Human Behavior*, 26, 219–239.

- Skeem, J., Cauffman, E., & Nold, J. (2002). *Recent research and legal developments on juvenile psychopathy*. Paper presented at the annual American Psychology-Law Society conference, Austin, Texas.
- Skeem, J. L., Edens, J. F., Camp, J., & Colwell, L. H. (2004). Are there racial differences in levels of psychopathy? A meta-analysis. *Law and Human Behavior*, 28, 505–527.
- Skeem, J. L., & Mulvey, E. P. (2001). Psychopathy and community violence among civil psychiatric patients: Results from the MacArthur violence risk assessment study. *Journal of Consulting and Clinical Psychology*, 69, 358–374.
- Skeem, J. L., & Petrila, J. (2004). Juvenile psychopathy: Informing the debate. *Behavioral Sciences and the Law*, 22, 1–4.
- Spain, S. E., Douglas, K. S., Poythress, N. G., & Epstein, M. (2004). The relation between psychopathic features, violence and treatment outcomes: The comparison of three measures of psychopathic features. *Behavioral Sciences and the Law*, 22, 85–102.
- *Vincent, G. M., Odgers, C. L., Watkins, A., & Corrado, R.R. (2005). *Predictive validity of the psychopathy checklist: Youth version for recidivism with female adolescent offenders*. Manuscript submitted for publication.
- Walters, G. (2003). Predicting criminal justice outcomes with the Psychopathy Checklist and Lifestyle Criminality Screening Form. *Behavioral Sciences and the Law*, 21, 89–102.
- Walters, G. (2003). Predicting institutional adjustment and recidivism with the Psychopathy Checklist factor scores: A meta-analysis. *Law and Human Behavior*, 27, 541–558.
- Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20: Assessing risk for violence (version 2)*. Vancouver, Canada: Simon Fraser University.
- Zinger, I., & Forth, A. E. (1998). Psychopathy and Canadian criminal proceedings: The potential for human rights abuses. *Canadian Journal of Criminology*, 40, 237–277.