The four-factor model of the Psychopathy Checklist—Revised: Validation in a Dutch forensic inpatient sample

Almar J. Zwets a,b,⁎, Ruud H.J. Hornsveld c, Craig Neumann d, Peter Muris b,e, Hjalmar J.C. van Marle c,f

a Forensic Psychiatric Center De Kijvelanden, P.O. Box 900, 3160 AC Rhoon, The Netherlands
b Erasmus University Rotterdam, Faculty of Social Sciences, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
c Erasmus University Rotterdam, Medical Centre, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands
d University of North Texas, Department of Psychology, 1155 Union Circle # 311280, Denton, TX, United States
e Maastricht University, Clinical Psychological Science, Faculty of Psychology and Neuroscience, P.O. Box 616, 6200 MD Maastricht, The Netherlands
f Forensic Outpatient Clinic Het Dok Rotterdam, P.O. Box 363, 3000 AJ Rotterdam, The Netherlands

A R T I C L E   I N F O

Available online xxxx

Keywords:
Psychopathy Checklist—Revised
Four-factor structure
Forensic psychiatric inpatients
Measurement invariance

A B S T R A C T

In The Netherlands, the Ministry of Security and Justice requires the assessment of the Psychopathy Checklist—Revised (PCL-R; Hare, 1991; Hare, 2003) in all forensic psychiatric inpatients. To examine the four-factor structure of the Psychopathy Checklist—Revised, confirmatory factor analysis (CFA) was conducted using a Dutch sample of forensic psychiatric inpatients (N = 411) and the results indicated acceptable fit. Also, using multiple group CFA, the four-factor model provided an acceptable fit in both patients with a personality disorder and patients with a psychotic disorder, and there was reasonably good evidence of measurement invariance between these two subgroups. Furthermore, correlations with external measures of aggression and personality traits provided additional support for the validity of the four-factor model in patients with a personality disorder. In patients with a psychotic disorder fewer significant correlations with external measures were found. Taken together, the results support the use of the four-factor structure in Dutch offenders who are detained under hospital order.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

In Dutch forensic psychiatric settings, the Psychopathy Checklist—Revised (PCL-R; Hare, 1991, 2003) is required to be administered according to the Ministry of Security and Justice, given its ability to predict recidivism and disruptive institutional behavior (Hare & Neumann, 2009). A number of studies have indeed demonstrated that the PCL-R is a predictor of violent and non-violent recidivism (Douglas, Vincent, & Edens, 2006; Hare & Neumann, 2008; Hildebrand, Hesper, Spreen, & Nijman, 2005; Mokros, Vohs, & Habermeyer, 2013). For example, the study by Hildebrand et al. (2005) demonstrated that the PCL-R (Hare, 1991) may be a better predictor of recidivism than the Historical Clinical Risk management 20 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997) and Historisch Klinisch Toekomst–30 (HKT-30; Dienst Justitiële Inrichtingen, 2002). As a result of these studies, the PCL-R has important criminal justice implications in The Netherlands as it is often used as a tool in decision-making about leave or discharge.

The link between PCL-R scores and different forms of aggressive behavior has been the topic of multiple studies. Several authors have demonstrated that violent patients with a relatively low score on psychopathy mainly show reactive aggression, whereas those with a relatively high score tend to be both reactively and proactively aggressive (Cima & Raine, 2009; Cornell et al., 1996; Woodworth & Porter, 2002). These two forms of aggression seem to be related to different dynamic criminogenic needs and consequently require a different treatment approach (Andrews & Bonta, 2003). Therefore, whether a patient exhibits mainly reactive aggression or both reactive and proactive aggression requires a thorough assessment of the determinants of violent behavior, including the degree of psychopathy.

The Dutch Ministry of Security and Justice broadly distinguishes two groups in forensic psychiatric inpatients: patients with a (chronic) psychotic disorder and patients with a personality disorder as their primary diagnosis. Although these two groups have unique features which might lead to criminal behavior, like threat/control-override symptoms in the case of patients with a chronic psychosis (Link & Stueve, 1994; Nederlof, Muris, & Hovens, 2011), they also seem to share common risk factors such as psychopathy (Hill, Neumann, & Rogers, 2004; Neumann, Hare, & Newman, 2007; Tengström, Grann, Langström, & Kullgren, 2000; Vitacco, Neumann, & Jackson, 2005). However, until now, no study that examined the factor structure of the PCL-R and its relation to external measures while distinguishing between chronic...
psychotic and personality disordered patients can be found. As discussed below, research has generally relied on studying combined subsamples of heterogeneous groups of forensic psychiatric patients.

1.1. Factor structure of the PCL-R

The underlying factor structure of the PCL-R has been a research topic for the last two decades. However, depending on the analytic approach that has been used (cf. Neumann, Kosson, & Salekin, 2007), studies about the factor structure have often resulted in a variety of somewhat divergent conclusions. Initial studies with a 22-item version and the definitive PCL-R with 20 items yielded evidence for a two-factor structure (Hare, 1991; Harpur, Hakstian, & Hare, 1988; Harpur, Hare, & Hakstian, 1989). Although this two-factor structure was confirmed in several studies (e.g., Hobson & Shine, 1998; Pham, 1998), other researchers could not always find an adequate fit in samples of North American minimum-security inmates (McDermott et al., 2000), sexual offenders (Weaver, Meyer, Van Nort, & Tristan, 2006), and Dutch violent forensic psychiatric inpatients (Hildebrand, De Ruiter, de Vogel, & van der Wolf, 2002).

In 2001, Cooke and Michie noted that the available research “does not provide compelling evidence for the adequacy of a two-factor model for psychopathy” (p. 172). Consequently, they proposed an alternative model that was more focused on psychopathy as a personality construct and less on criminality. Using item-response theory, confirmatory factor analysis, cluster analysis, and various rational proposals for their analysis of 1389 North American prisoners and forensic psychiatric inpatients, they suggested that a hierarchical three-factor model provided a better fit than the original two-factor model. In this three-factor model, the first factor of Hare’s two-factor model was divided into two separate factors, whereas the third factor consisted of only five items. Other remaining items which they believed only measured criminal behavior were discarded, because criminal behavior was in their opinion best viewed as a secondary feature of psychopathy (Cooke, Michie, Hart, & Clark, 2004). This three-factor model was disputed by Hare (2003) and colleagues (Hare & Neumann, 2008, 2010; Neumann, Vitacco, Hare, & Wupperman, 2005; Vitacco, Rogers et al., 2005). Based on factor analysis, item response theory and multidimensional scaling, Hare and Neumann (2005, 2006) proposed a model with four correlated factors, namely Interpersonal (glib/superficial charm, grandiose self-worth, pathological lying, conning/manipulative), Affective (lack of remorse or guilt, shallow affect, callous/lack of empathy, failure to accept responsibility for actions), Lifestyle (need for stimulation/proneness to boredom, impulsivity, irresponsibility, parasitic lifestyle, lack of realistic long-term goals), and Antisocial (poor behavior controls, early behavior problems, juvenile delinquency, revocation of conditional release, criminal versatility). This four-factor model is highly comparable to the traditional two-factor model (Hare & Neumann, 2008), given that each factor of this two-factor model is split up into two separate factors (factor 1 into an Interpersonal factor and an Affective factor; Factor 2 into a Lifestyle factor and an Antisocial factor). Based on an extensive review of the literature, Hare and Neumann (2008) proposed that “the presence of early and persistent antisocial behavior is an important feature of the psychopathy construct” (p. 62). Relatively, these authors suggested that psychopathy and its specific features could also be viewed in terms of extreme variants of normal personality traits and behaviors.

Hare’s four-factor model has been confirmed in several large PCL-based studies, including forensic psychiatric inpatients (Hill et al., 2004), a combined sample of offenders and forensic psychiatric inpatients, which included both males and females (Neumann, Hare, & Neumann, 2007), civil psychiatric patients (Vitacco, Neumann et al., 2005), mentally disordered offenders (Vitacco, Rogers et al., 2005), and adolescents (Kosson, Cytarski, Steuerwald, Neumann, & Walker-Matthews, 2002; Neumann, Kosson, Forth, & Hare, 2006). Recent research with Canadian (Oliver, Neumann, Wong, & Hare, 2012) and Swedish offenders (Neumann, Hare, & Johansson, 2012) has further confirmed the validity of the four-factor PCL-R model. Furthermore, the four-factor model has been examined for invariance of model parameters across a wide range of samples and methodologies, including male and female offenders and psychiatric patients (Bolt, Hare, Vitale, & Newman, 2004), North American and German offenders (Mokros et al., 2011), male civil psychiatric patients (Jackson, Neumann & Vitacco, 2007), and adolescents (Kosson et al., 2012; Neumann et al., 2006), as well as a mega-world general population sample using the Self-Report Psychopathy (SRP) scale (Neumann, Schmitt, Carter, Embley & Hare, 2012). In all these studies the evidence for invariance across diverse groups has generally been good, as well as providing further support for the four-factor model.

1.2. PCL-R factors in relation to external measures

To provide a better understanding of the PCL factors, a number of studies have addressed their relation to external correlates of psychopathy, including mental disorders (e.g., Hildebrand & De Ruiter, 2004), criminality (Blackburn & Coid, 1998), normal-range personality traits (Lynam & Dereffino, 2006), different forms of aggression (Cima & Raine, 2009; Cornell et al., 1996; Woodworth & Porter, 2002), violence in the community (Vitacco, Neumann et al., 2005), and institutional aggression (Guy, Edens, Anthony, & Douglas, 2005; Hildebrand, De Ruiter, & Nijman, 2004; Hill et al., 2004). The relation between the original two PCL-R factors (Hare, 1991) and “institutional adjustment” was examined by Walters (2003b) by means of a meta-analysis of 41 studies in different populations such as maximum adult security forensic psychiatric patients and juvenile security state school inmates. Institutional adjustment had been operationalized as “verbal infractions” or “physical aggression”. The original factor 2 of the PCL-R appeared to have a moderately well positive correlation with institutional adjustment, whereas the original factor 1 showed less robust associations. Guy et al. (2005) refined this analysis and found less evidence for divergent relationships between the two original PCL-R factors and various types of aggressive and violent behavior. In their study, the relation between PCL-R total, factor 1, and factor 2 scores on the one hand, and “General aggression” on the other hand, yielded low mean weighted effect sizes.

Most of the research has also indicated that in particular the original factor 2, which primarily refers to socially deviant behavior, is a good predictor of problem behaviors such as alcohol abuse (e.g., Readon, Lang, & Patrick, 2002), drug abuse (e.g., Lammers, 2009), aggressive behavior (e.g., Walters, 2003a), and even violent recidivism (e.g., Douglas et al., 2006; Hildebrand et al., 2005). Relations between the original factor 1 and these forms of problem behavior are often modest or even absent. However, given the emerging evidence that the four PCL-R factors may have differential links to various external correlates (Hare & Neumann, 2008), studies based on the older two-factor conception of the PCL may have missed the opportunity to uncover such a pattern of findings. Some studies employing the four-factor model of the PCL-R have shown similar results as the relation between the Lifestyle factor and the Antisocial factor with (violent) recidivism is often confirmed (e.g., Olver et al., 2012), while others have documented a more nuanced pattern of differential associations with violent (Vitacco, Neumann et al., 2005) or aggressive behavior (Hill et al., 2004).

1.3. PCL-R in forensic patients with a psychotic disorder

Several studies have specifically investigated the PCL-R in patients with a psychotic disorder. However, studies that focused on the applicability of the four-factor model are limited in this subgroup of patients. Hill et al. (2004) applied a confirmatory factor analysis to investigate the two-, three-, and four-factor model of the PCL-R Screening Version (PCL-R:SV; Hart, Cox, & Hare, 1995) in a sample of 149 male forensic psychiatric inpatients with mainly psychotic disorders. Results showed
that all models had a good fit, with the four-factor model displaying the best fit.

Furthermore, various studies have demonstrated that psychopathy is related to violent behavior among forensic inmates with a psychotic disorder (Fullam & Dolan, 2006; Fullam & Dolan, 2008; Rice & Harris, 1992; Tengström et al., 2000; Volavka & Citrome, 2008). Although most investigations did not explore the relation between psychopathy and violence on a facet-level, there is some evidence indicating that especially the Interpersonal and the Antisocial factor are related to violent behavior in forensic inmates with a psychotic disorder (Fullam & Dolan, 2008; Hill et al., 2004). Based on these and several other findings, the presence of psychopathy (or an antisocial personality disorder) has been proposed as one of the main trajectories for violent behavior in patients with schizophrenia.

1.4. The current study

We examined the PCL-R in a sample of Dutch forensic psychiatric patients with a psychotic disorder and a sample of patients with a personality disorder. Given the policy of the Dutch Ministry of Security and Justice, we choose to conduct multi-group confirmatory factor analysis based on the four-factor model of the PCL-R. In addition, the four factors were correlated with measures of prosocial behavior, aggressive behavior, and anger. Given the diversity of the patient sample and the role of the PCL-R in forensic psychiatry, it is important that the measured underlying trait is consistent among subgroups of patients. This issue is often addressed in measurement invariance studies which, in the case of the PCL-R, have focused on cultural differences, for instance between North American and German offenders (Mokros et al., 2011), African American and Caucasian inmates (Cook, Kossen, & Michie, 2001) or psychiatric patients (Jackson, Neumann, & Vitacco, 2007), or on sex differences within general population samples (Neumann, Schmitt, Carter, Embley, & Hare, 2012). Because measurement invariance is required to make meaningful comparisons, these studies play an important role in validating the use of the PCL-R within a variety of groups. Therefore, one of the goals of this study is to investigate if measurement invariance could be established between our samples of inmates with a personality disorder and patients with a psychotic disorder.

Finally, external correlates of the four-factor model were explored for each group separately by relating the factor scores to prosocial behavior and aggressive behavior, measured with an observation scale, and self-report questionnaires on anger and aggression.

We tested the following hypotheses:

1) The four-factor model is supported in the total group of forensic psychiatric patients as well as in each of the two subgroups (personality disorder, psychotic disorder) separately.

2) The model parameters of the four-factor model meet the criteria for measurement invariance between the two subgroups.

3) The Lifestyle factor and the Antisocial factor are significantly positively correlated with measures of aggression and anger in the total group and in each of the two subgroups.

2. Method

2.1. Setting

The current study was conducted at “FPC De Kijvelanden”, a forensic psychiatric center with 178 beds or places in Poortugaal, a village located in the vicinity of Rotterdam, The Netherlands. Patients are accommodated in nine wards at a rehabilitation unit department or in sheltered homes. During daytime, inmates do not stay on the ward, but follow educational, vocational, and treatment programs elsewhere in the hospital.

2.2. Patients

The study was conducted among a group of 411 patients detained under hospital order who were admitted in “De Kijvelanden” between 1996 and 2011. In The Netherlands, patients are detained under hospital order when the court has established a relation between a psychiatric disorder on the one hand and an offense on the other hand (e.g., Van Marle, 2000, 2002). These patients have committed an offense for which a maximum imprisonment of more than four years applies, for instance severe assault, manslaughter, or murder. Rulings are based on the evaluations of a psychiatrist and/or psychologist at a special assessment center of the Ministry of Security and Justice. Without care or treatment, recidivism should be deemed likely. In Dutch forensic psychiatric hospitals, a distinction is often made in patients with a personality disorder and patients with a chronic psychotic disorder. This dichotomy is applied not only to the wards, but also to the treatment programs that are provided.

Of the 411 patients, 269 had a personality disorder as their primary diagnosis on Axis II of the DSM-IV-TR (American Psychiatric Association, 2000), whereas 142 patients had a psychotic disorder as their main diagnosis. Most patients of the first group had a cluster B personality disorder (177 patients; 65.8%), followed by a personality disorder not otherwise specified (72 patients; 26.8%). Their mean age was 37.79 years (SD = 10.05; range: 19–66 years). Most patients of the second group had a schizophrenic disorder (89 patients; 62.7%), followed by a psychotic disorder not otherwise specified (19 patients; 13.4%). The patients with a psychotic disorder were on average 36.63 years old (SD = 10.46; range: 21–76 years). All of the patients were classified by experienced psychiatrists after an extensive psychiatric evaluation that included clinical and psychological evaluations.

2.3. Measures

The Psychopathy Checklist—Revised (PCL-R; Hare, 1991; Dutch version: Vertommen, Verheul, De Ruiter, & Hildebrand, 2002) is a checklist with 20 items, which have to be rated on a three-point scale with 0 = “does not apply,” 1 = “applies to some extent,” and 2 = “applies.” In a group of 1192 inmates, Cronbach’s α for the total score appeared to be .87 and the average inter-item correlation was .25 (Hare, 1991). Tentative evidence for the validity was found in a subgroup of 98 forensic psychiatric inpatients as there were modest, but meaningful correlations with self-report questionnaires such as the Minnesota Multiphasic Personality Inventory (MMPI; Dutch version: Derksen, De Mey, Sloore, & Hellenbosch, 1993).

For the current dataset, internal consistency of the PCL-R total score was good (George & Mallory, 2003) with Cronbach’s α of .83 for both groups of patients. For the group of patients with a personality disorder, the internal consistency of the Interpersonal, Lifestyle and Antisocial factors was acceptable (α’s of .71, .71, and .70 respectively) whereas the internal consistency of the Affective factor was somewhat lower (.67). The internal consistency of the group of patients with a psychotic disorder was acceptable for the Affective and Lifestyle factors (α’s of .76 and .77 respectively) and lower for the Interpersonal and Antisocial factors (α’s .60 and .69 respectively). However, the use of Cronbach’s alpha for the factor scores can be problematic as the number of items may influence the outcomes (Schmitt, 1996). Therefore, mean inter-item correlations (MICs) were also calculated as this descriptive statistic is a true indicator of item homogeneity (Simms & Watson, 2007). The mean inter-item correlations indicated acceptable homogeneity for all factors in patients with a personality disorder (Interpersonal = .39, Affective = .34, Lifestyle = .33, Antisocial = .31) and in patients with a psychotic disorder (Interpersonal = .29, Affective = .45, Lifestyle = .40, Antisocial = .30). For the total PCL-R scale, mean inter-item correlations were acceptable for both groups (.20 for both groups) as MICs of .20 or above indicate acceptable homogeneity (Nunnally & Bernstein, 1994).
For 370 patients, PCL-R scores were assessed on the basis of file information and frequent contacts with the patient, but without the structured interview. The PCL-R scores of the remaining 41 patients were assessed by means of a structured interview in combination with file study. However, there are indications that file information alone also yields reliable and valid PCL-R scores. According to Bolt et al. (2004), the psychometric properties, correlates, and predictive ability of the PCL-R scored from file-only reviews are in general much the same as those scored with the standard protocol, although PCL-R assessments from file reviews are, on average, several points lower than those arrived at through the standard protocol (e.g., Gran, Längström, Tengström, & Stålenheim, 1998).

For the current study, the files that were used as a source of information for scoring on the PCL-R comprised detailed information about life history, committed offenses and elaborated reports from psychiatrists and/or psychologists. These reports were often made in a special forensic assessment center (Pieter Baan Centrum), in which the patient had to stay for observation by order of the court. The PCL-R was administered by certified clinical psychologists who had completed a 3-day PCL-R workshop. The PCL-R scores for the 41 patients who were interviewed were assessed independently by two trained psychologists. There was a strong agreement between raters (ICC = .81, CI95: 67–89).

The Observation Scale for Aggressive Behavior (OSAB; Hornsveld, Nijman, Hollin, & Kraaimaat, 2007) measures behavior on the ward. The scale comprises 40 items that can be allocated to the subscales of irritation/anger, anxiety/gloominess, aggressive behavior, prosocial behavior, antecedent, and sanction. The staff of the ward rates frequency of the behavior of the patients in the preceding week on a four-point scale with 1 = never, 2 = occasionally, 3 = frequently, and 4 = “always.”

The Observation Scale for Aggressive Behavior (OSAB; Hornsveld, Nijman, Hollin, & Kraaimaat, 2007) measures behavior on the ward. The scale comprises 40 items that can be allocated to the subscales of irritation/anger, anxiety/gloominess, aggressive behavior, prosocial behavior, antecedent, and sanction. The staff of the ward rates frequency of the behavior of the patients in the preceding week on a four-point scale with 1 = never, 2 = occasionally, 3 = frequently, and 4 = “always.”

The Observation Scale for Aggressive Behavior (OSAB; Hornsveld, Nijman, Hollin, & Kraaimaat, 2007) measures behavior on the ward. The scale comprises 40 items that can be allocated to the subscales of irritation/anger, anxiety/gloominess, aggressive behavior, prosocial behavior, antecedent, and sanction. The staff of the ward rates frequency of the behavior of the patients in the preceding week on a four-point scale with 1 = never, 2 = occasionally, 3 = frequently, and 4 = “always.”

The Observation Scale for Aggressive Behavior (OSAB; Hornsveld, Nijman, Hollin, & Kraaimaat, 2007) measures behavior on the ward. The scale comprises 40 items that can be allocated to the subscales of irritation/anger, anxiety/gloominess, aggressive behavior, prosocial behavior, antecedent, and sanction. The staff of the ward rates frequency of the behavior of the patients in the preceding week on a four-point scale with 1 = never, 2 = occasionally, 3 = frequently, and 4 = “always.”

The Observation Scale for Aggressive Behavior (OSAB; Hornsveld, Nijman, Hollin, & Kraaimaat, 2007) measures behavior on the ward. The scale comprises 40 items that can be allocated to the subscales of irritation/anger, anxiety/gloominess, aggressive behavior, prosocial behavior, antecedent, and sanction. The staff of the ward rates frequency of the behavior of the patients in the preceding week on a four-point scale with 1 = never, 2 = occasionally, 3 = frequently, and 4 = “always.”

The Observation Scale for Aggressive Behavior (OSAB; Hornsveld, Nijman, Hollin, & Kraaimaat, 2007) measures behavior on the ward. The scale comprises 40 items that can be allocated to the subscales of irritation/anger, anxiety/gloominess, aggressive behavior, prosocial behavior, antecedent, and sanction. The staff of the ward rates frequency of the behavior of the patients in the preceding week on a four-point scale with 1 = never, 2 = occasionally, 3 = frequently, and 4 = “always.”

The Observation Scale for Aggressive Behavior (OSAB; Hornsveld, Nijman, Hollin, & Kraaimaat, 2007) measures behavior on the ward. The scale comprises 40 items that can be allocated to the subscales of irritation/anger, anxiety/gloominess, aggressive behavior, prosocial behavior, antecedent, and sanction. The staff of the ward rates frequency of the behavior of the patients in the preceding week on a four-point scale with 1 = never, 2 = occasionally, 3 = frequently, and 4 = “always.”

The Observation Scale for Aggressive Behavior (OSAB; Hornsveld, Nijman, Hollin, & Kraaimaat, 2007) measures behavior on the ward. The scale comprises 40 items that can be allocated to the subscales of irritation/anger, anxiety/gloominess, aggressive behavior, prosocial behavior, antecedent, and sanction. The staff of the ward rates frequency of the behavior of the patients in the preceding week on a four-point scale with 1 = never, 2 = occasionally, 3 = frequently, and 4 = “always.”

Given the ordinal data of the PCL-R items, the robust weighted least squares estimation procedure was used for these analyses. To compare the fit of the four-factor model in the two subgroups (personality disorder versus psychotic disorder) a standard approach for testing for measurement invariance was used (Olver et al., 2012). The basic assumption of measurement invariance is that the scale items provide the same information (e.g., discrimination) across different groups. To test for measurement invariance between the two groups, we conducted a series of multiple group CFAs: an unconstrained four-factor model which allowed item parameters to be freely estimated across the two groups (configural invariance), the same four-factor model in which the factor loadings were constrained but the thresholds were allowed to freely vary across groups (metric invariance), and a model in which both the factor loadings and thresholds were constrained to be equal across the two groups (scalar invariance). The unconstrained model served as a baseline model for the comparison with subsequent analyses. For this model, the syntax in Mplus was used to specify the fit statistics for the four-factor model and the measurement invariance constraints. The syntax in Mplus was used to specify the fit statistics for the four-factor model and the measurement invariance constraints. The syntax in Mplus was used to specify the fit statistics for the four-factor model and the measurement invariance constraints. The syntax in Mplus was used to specify the fit statistics for the four-factor model and the measurement invariance constraints. The syntax in Mplus was used to specify the fit statistics for the four-factor model and the measurement invariance constraints.

3. Results

3.1. Four factor model in the current samples

Table A.1 shows the mean PCL-R item scores, mean item scores, mean factor scores and mean total scores of the patients with a
personality disorder and the patients with a psychotic disorder. The mean PCL-R total scores did not differ significantly from scores previously obtained in comparable forensic psychiatric samples with (Tengström et al., 2000) and without psychotic disorders (Hare, 2003). In the current study, patients with a personality disorder had a significantly higher mean total score on the PCL-R (M = 21.82) than patients with a psychotic disorder (M = 17.32; see Table B.1). This significant difference between the two groups was found for three factors of the four-factor model (χ^2 ≥ 6.03, p < .05), with the biggest difference emerging on the Interpersonal factor (χ^2 = 60.40, p < .01). The mean total scores on the Antisocial factors were not significantly different between both subgroups (χ^2 = 2.72, n.s.).

A four-factor model CFA was carried out for the total sample, and for both groups of patients. Results indicate that the four-factor model provided a moderate to acceptable fit to the data for the total sample (CFI = .90, RMSEA = .076, CI90 = .068–.084; see Table B.1). Similar results were found for the two subgroups. Patients with personality disorder had an acceptable fit on the absolute index (RMSEA = .07, CI90 = .064–.084), whereas the incremental index was just under the acceptable range (CFI = .89). The four-factor model had an acceptable fit for both the absolute index (RMSEA = .07, CI90 = .056–.087) and the incremental index (CFI = .90) in patients with a psychotic disorder. Item-to-factor loadings and correlations (see Figs. A.1 and A.2) were all significant (p < .001) for both groups.

Threshold values for the PCL-R items were also examined for the total group and both subgroups. These ‘extremity’ values provide information about to what extent a certain behavior has to be present before an item-rating (a PCL-R rating of 2 in this case) is likely to be endorsed (Reise, 1999). Results (Fig. B.1) indicate that items representing the Interpersonal factor tend to have the highest threshold values for both subgroups. This might indicate that these items are rated when they are explicitly present. On the other hand, items of the Affective factor in particular had low threshold values, which would indicate that these items tended to be rated a 2 when this behavior was already present in limited extent.

3.2. Differences between patients with a personality disorder and patients with a psychotic disorder

To determine measurement invariance, three multi-group CFDAs of the four-factor model (configural invariance, metric invariance and scalar invariance) were tested. The unconstrained model (configural invariance) had acceptable model fit (CFI = .90, RMSEA = .07, CI90 = .06–.08). Model fit of the model with constrained factor loadings and free thresholds (metric invariance) was also acceptable (CFI = .91, RMSEA = .07, CI90% = .06–.07), whereas the model fit of the model with constrained factor loadings and constrained thresholds (scalar invariance) was just under acceptable (CFI = .89, RMSEA = .07, CI90% = .06–.08). Scalar measurement invariance could be confirmed as the change of CFI value in the two models with constrained factor loadings and/or thresholds, compared to the unconstrained model, was equal or under .01. However, according to the more traditional chi-square

### Table A.1

<table>
<thead>
<tr>
<th>PCL-R item</th>
<th>Personality disorder</th>
<th>Psychotic disorder</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Glibness/superficial charm</td>
<td>0.83 (0.80)</td>
<td>0.30 (0.58)</td>
<td>1.08 25 9 0 43.92**</td>
</tr>
<tr>
<td>2. Grandiose sense of self worth</td>
<td>0.94 (0.80)</td>
<td>0.61 (0.76)</td>
<td>79 39 24 0 15.47**</td>
</tr>
<tr>
<td>3. Need for stimulation/proveness to boredom</td>
<td>1.14 (0.83)</td>
<td>0.84 (0.82)</td>
<td>61 43 38 0 11.89**</td>
</tr>
<tr>
<td>4. Pathological deception</td>
<td>0.71 (0.77)</td>
<td>0.33 (0.61)</td>
<td>104 27 10 1 23.50**</td>
</tr>
<tr>
<td>5. Conning/ manipulative</td>
<td>1.12 (0.83)</td>
<td>0.42 (0.67)</td>
<td>96 31 14 1 63.18**</td>
</tr>
<tr>
<td>6. Lack of remorse or guilt</td>
<td>1.65 (0.54)</td>
<td>1.62 (0.56)</td>
<td>5 44 93 0 0.30</td>
</tr>
<tr>
<td>7. Shallow affect</td>
<td>1.36 (0.63)</td>
<td>1.23 (0.62)</td>
<td>15 80 47 0 4.36*</td>
</tr>
<tr>
<td>8. Callous/lack of empathy</td>
<td>1.37 (0.67)</td>
<td>1.37 (0.74)</td>
<td>22 79 47 0 31.52**</td>
</tr>
<tr>
<td>9. Parasitic lifestyle</td>
<td>0.92 (0.80)</td>
<td>0.69 (0.71)</td>
<td>64 58 20 0 8.13**</td>
</tr>
<tr>
<td>10. Poor behavior controls</td>
<td>1.44 (0.74)</td>
<td>1.33 (0.75)</td>
<td>24 71 24 1 1.87</td>
</tr>
<tr>
<td>11. Promiscuous sexual behavior</td>
<td>1.01 (0.91)</td>
<td>0.48 (0.76)</td>
<td>96 22 23 1 31.52**</td>
</tr>
<tr>
<td>12. Early behavioral problems</td>
<td>0.82 (0.84)</td>
<td>0.56 (0.80)</td>
<td>78 21 24 1 19.81**</td>
</tr>
<tr>
<td>13. Lack of realistic long-term goals</td>
<td>0.97 (0.76)</td>
<td>1.15 (0.77)</td>
<td>32 54 53 3 4.90*</td>
</tr>
<tr>
<td>14. Impulsivity</td>
<td>1.37 (0.77)</td>
<td>1.19 (0.74)</td>
<td>27 60 54 1 4.83*</td>
</tr>
<tr>
<td>15. Irresponsibility</td>
<td>1.29 (0.75)</td>
<td>1.06 (0.75)</td>
<td>36 61 44 1 8.90**</td>
</tr>
<tr>
<td>16. Failure to accept responsibility for actions</td>
<td>1.59 (0.61)</td>
<td>1.54 (0.65)</td>
<td>12 41 88 1 0.75</td>
</tr>
<tr>
<td>17. Many short-term marital relationships</td>
<td>0.48 (0.74)</td>
<td>0.20 (0.50)</td>
<td>117 15 6 4 15.96**</td>
</tr>
<tr>
<td>18. Juvenile delinquency</td>
<td>0.74 (0.85)</td>
<td>0.67 (0.86)</td>
<td>80 22 35 5 0.52</td>
</tr>
<tr>
<td>19. Revocation of conditional release</td>
<td>1.18 (0.90)</td>
<td>1.05 (0.95)</td>
<td>60 14 67 1 1.77</td>
</tr>
<tr>
<td>20. Criminal versatility</td>
<td>0.81 (0.78)</td>
<td>0.81 (0.77)</td>
<td>57 52 31 2 0.00</td>
</tr>
</tbody>
</table>

### Table B.1

| PCL-R total score | 21.82 (7.57) | 17.32 (7.12) | 31.60** |
| F1 Interpersonal | 3.60 (2.35) | 1.69 (1.84) | 60.40* |
| F2 Affective | 5.97 (1.74) | 5.51 (1.90) | 6.03* |
| F3 Lifestyle | 5.70 (2.68) | 4.92 (2.73) | 7.59** |
| F4 Antisocial | 5.01 (2.75) | 4.53 (2.83) | 2.72 |

Note. *p < .05, **p < .01 (two-tailed).
difference test, the model with constrained factor loadings and thresholds had significantly lower fit than the unconstrained model ($\Delta\chi^2 = 61.41; p < .01$), whereas the model with constrained factor loadings did not result in a significantly lower fit compared to the unconstrained model ($\Delta\chi^2 = 13.23; p = .51$).

### 3.3. Correlations with external measures

Table C.1 shows the mean scores and standard deviations of the external measures. A significant difference was found on the prosocial subscale of the OSAB ($t = 3.42, p < .01$), with patients with a personality disorder having a higher score. Correlations between the four PCL-R factors and the external correlate observed and self-reported measures are shown in Table D.1. For the patients with a personality disorder, the Interpersonal factor and the Affective factor were not significantly correlated with any of the external measures. The Lifestyle factor and the Antisocial factor (observed aggression and self-reported physical aggression) correlated significantly with the Antisocial factor ($r = .49, p < .05$). No other significant correlations with any of the external measures were found.

### 4. Discussion

We examined the four-factor structure of the PCL-R (Hare & Neumann, 2008) in a forensic psychiatric sample containing patients with a personality disorder and patients with a chronic psychotic disorder as their main diagnosis. A multi-group confirmatory factor analysis of the four-factor model of the PCL-R was performed. We tested not only the goodness of fit of this model, but also the degree of measurement invariance between the two patient groups, and the relations between the four factors and the external correlates.

The results indicate that the four-factor model had an acceptable fit for the total sample and both subsamples (albeit just below acceptable for the incremental index for the patients with a personality disorder). This conclusion is in accordance with recent studies that have focused on the four-factor model (Hare & Neumann, 2006; Hill et al., 2004; Mokros et al., 2011; Neumann, Hare et al., 2007, 2012; Neumann et al., 2006; Olver et al., 2012; Vitacco, Neumann, & Jackson, 2005). Furthermore, the PCL-R items demonstrated a reasonable degree of invariance across patients with a personality disorder and patients with a psychotic disorder. This suggests that the instrument can be administered in both patient groups and that the scores of these subgroups can be compared in a valid way. However, the more traditional chi-square difference approach suggests less strong evidence of scalar invariance and might therefore raise some questions about the applicability of the PCL-R in this rather heterogeneous group of patients under hospital order, along with the PCL-R rating methodology used in this study.
Further comparisons revealed that patients with a personality disorder had significantly higher scores on the Interpersonal factor, Affective factor and Lifestyle factor. There was no significant difference on the Antisocial factor, which was in contrast to the expectations as patients with a psychotic disorder detained under hospital order are more often first-offenders than patients detained under hospital order without any psychotic disorder (Nijman, Van Marle, & Kavelaars, 2006).

The threshold values of the items showed interesting results across the two subsamples. In general the threshold values of the patients with a psychotic disorder were higher than those of the patients with a personality disorder. This result may indicate that it is likely that the expression of psychopathic traits had to be evident above and beyond the psychotic symptoms of the psychotic patients before they received threshold ratings of 2 on the PCL-R. The thresholds of the items of the Interpersonal factor were relatively high for both groups of patients, but in particular for the patients with a psychotic disorder. This might indicate that high levels of these underlying traits must be present before these items can be scored as present (i.e., a charming and superficial communication style) on the PCL-R. These results are not surprising and might suggest that psychotic symptoms were not being confused with symptoms of psychopathy when the PCL-R was being administered to the patients with a psychotic disorder.

![Factor loadings of patients with a personality disorder and patients with a psychotic disorder.](image)

![Threshold values of the items of the PCL-R in the total sample (n = 411), patients with a personality disorder (n = 269) and patients with a psychotic disorder (n = 142). Note: Items are ordered in such a way as they represent the factors.](image)

Please cite this article as: Zwets, A.J., et al., The four-factor model of the Psychopathy Checklist—Revised: Validation in a Dutch forensic inpatient sample, *International Journal of Law and Psychiatry* (2015), [http://dx.doi.org/10.1016/j.ijlp.2015.01.016](http://dx.doi.org/10.1016/j.ijlp.2015.01.016)
Several low threshold values were also evident. For both groups, the threshold values of item 6 (lack of remorse or guilt), item 10 (poor behavior controls), item 16 (failure to accept responsibility for own actions), and item 19 (revocation of conditional release) were rather low. However, these items also had high percentages of scores in which they were judged to apply to the patient (an item-score of 2). Therefore, these low threshold values may be scored as 2, even if an item does not fully apply. On the other hand, these low threshold values might be due, in part, to the specific samples used in the current study—i.e., high risk and severely disordered forensic psychiatric inpatients.

Another purpose of this study was to examine the relations between various factors of the PCL-R with external correlates. For this reason, the four PCL-R factors were correlated with indexes of observed and self-reported aggressive behavior and observed prosocial behavior. The correlations for the patients with a personality disorder showed a number of interesting findings in which the Lifestyle factor and the Antisocial factor were related to observed and self-reported (physical) aggression. Use of a latent variable approach to represent inpatient aggression provided the best evidence of an association between PCL-R factors and aggression across both samples. This association is in line with the results of other studies (e.g., Hornsveld et al., 2007; Walters, 2003a) and may provide support for the validity of the PCL-R in regard to aggressive behavior in patients with a personality disorder as their main diagnosis.

The pattern of correlations between the PCL-R factors and the external correlates for the patients with a psychotic disorder was less clear-cut. The relation with observed aggression could not be confirmed, whereas earlier studies demonstrated an association between psychopathy and aggression for patients with schizophrenia (Rice & Harris, 1992; Tengström et al., 2000). Nevertheless, when a latent aggression variable was used, we did find a moderately strong association with the PCL-R Antisocial factor (see also Fullam & Dolan, 2008; Hill et al., 2004). Notably, most previous studies have focused on aggression in terms of violent recidivism, whereas the current study used hospital staff observations, carried out in an environment in which aggressive behavior was inhibited due to multiple factors. These inhibiting factors include, among others, an intensive day-program, a relatively high staff–patient ratio, and the use of antipsychotic medication, which is related to a decrease of aggressive behavior in patients with schizophrenia (e.g., Buckley, 1999; Lammers, 2006; Ruedrich et al., 2008). Nevertheless, the current study also revealed a significant association between self-reported physical aggression and the PCL-R, especially the Lifestyle factor and the Antisocial factor, in patients with a psychotic disorder.

Our study should be interpreted with respect to several limitations. First of all, the sample of this study was a very selective group of patients, which may have repercussions for the generalization of the findings. These patients had all committed a serious violent offense and were diagnosed with a severe mental disorder. Secondly, it should be noticed that PCL-R scores were largely based on file study and not on a combination of file study and a structured interview. According to Hare and Neumann (2006) this is a disadvantage because “there may not be sufficient information to adequately score the items that tap interpersonal and affective features” (p. 66). A third limitation of the study is that in our opinion the structured and controlled environment in an institution with a relatively high patient–staff ratio has an attenuating effect on the patients’ behavior, and probably results in relatively low scores on the observation scale and certain self-report questionnaires (see also Hornsveld et al., 2009). Fourth, although the sample size of the total group and the subgroup of patients with a personality disorder was sufficient for the analyses, the sample size of the patients with a chronic psychotic disorder was rather small. Fifth, due to practical considerations, the external measures were limited to aggression related and personality measures and did not include other relevant measures like impulsivity, hostility or empathy (Bogaerts, Polak, Spreen, & Zwets, 2012).

To summarize, the four-factor model (Hare, 2003; Hare & Neumann, 2008) had an acceptable fit for the PCL-R data we collected in our institution. A reasonable degree of measurement invariance could be established between patients with a personality disorder and patients with a psychotic disorder, which supports comparisons between these groups of patients. However, a more traditional approach suggests questionable evidence for invariance at the level of item thresholds. Thus, investigators, at this point should proceed with caution when assuming that PCL-R scores level among patients with personality disorders versus those with psychotic disorders and that these groups have equivalent levels of psychopathy when they have the same PCL-R score. Therefore, more research about the measurement invariance of the PCL-R is needed in a larger sample of Dutch patients under hospital order. At this moment, testing for measurement invariance is rarely done (see Cyders, 2013; Vandenberg & Lance, 2000), though measurement invariance has to be established to make valid comparison between groups. Furthermore, the validity of this model was partially supported by meaningful correlations with external measures of anger and aggression in the group of patients with a personality disorder. However, such results were absent in the group of patients with a psychotic disorder, possibly due to multiple confounding variables such as the effects of medication. Our results provide a contribution to the validation of the PCL-R in various subgroups of patients. In today’s forensic psychiatry, very important decisions are based on PCL-R total and factor scores. Therefore, future research should focus on validating the PCL-R in different groups as these decisions should be supported by empirical evidence.

References


Please cite this article as: Zwets, A.J., et al., The four-factor model of the Psychopathy Checklist—Revised: Validation in a Dutch forensic inpatient sample, *International Journal of Law and Psychiatry* (2015), http://dx.doi.org/10.1016/j.ijlp.2015.01.016

---

**Table C.1**

Mean scores and standard deviations of external correlates.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Subscale</th>
<th>Total group (n = 411)</th>
<th>Personality disorder (n = 269)</th>
<th>Psychotic disorder (n = 142)</th>
<th>Statistical analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>r</td>
<td></td>
</tr>
<tr>
<td>OSAB Aggressive</td>
<td>15.67 (4.99)</td>
<td>15.64 (4.98)</td>
<td>15.74 (5.02)</td>
<td>−0.18</td>
<td></td>
</tr>
<tr>
<td>Prosocial</td>
<td>30.26 (7.67)</td>
<td>31.31 (7.28)</td>
<td>28.25 (8.02)</td>
<td>3.42**</td>
<td></td>
</tr>
<tr>
<td>AQ-SF Physical</td>
<td>8.13 (2.78)</td>
<td>8.13 (2.89)</td>
<td>8.14 (2.51)</td>
<td>−0.03</td>
<td></td>
</tr>
<tr>
<td>Verbal</td>
<td>7.48 (2.21)</td>
<td>7.34 (2.32)</td>
<td>7.80 (2.13)</td>
<td>−1.41</td>
<td></td>
</tr>
<tr>
<td>STAS Trait anger</td>
<td>16.73 (5.47)</td>
<td>16.87 (5.57)</td>
<td>16.36 (5.23)</td>
<td>0.64</td>
<td></td>
</tr>
</tbody>
</table>

Note: "p < .05, **p < .01 (two-tailed); PCL-R = Psychopathy Checklist—Revised; OSAB = Observation Scale Aggressive Behavior; AQ-SF = Aggression Questionnaire—Short Form; STAS = State-Trait Anger Scale.

---

**Table D.1**

Pearson’s r correlations between the PCL-R four-factor scores and external variables.

<table>
<thead>
<tr>
<th>Group</th>
<th>OSAB</th>
<th>AQ-SF</th>
<th>STAS</th>
<th>PCL-R</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pers.D.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>.09</td>
<td>.14</td>
<td>.16</td>
<td>.09</td>
<td>.33</td>
</tr>
<tr>
<td>F2</td>
<td>.11</td>
<td>.07</td>
<td>.03</td>
<td>.07</td>
<td>.19</td>
</tr>
<tr>
<td>F3</td>
<td>.23**</td>
<td>.03</td>
<td>.39**</td>
<td>.12</td>
<td>.27**</td>
</tr>
<tr>
<td>F4</td>
<td>.25**</td>
<td>.14</td>
<td>.41**</td>
<td>.18</td>
<td>.18</td>
</tr>
<tr>
<td>PsyD.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>.15</td>
<td>.22</td>
<td>−.02</td>
<td>−.18</td>
<td>−.30</td>
</tr>
<tr>
<td>F2</td>
<td>.06</td>
<td>.01</td>
<td>.21</td>
<td>.05</td>
<td>.03</td>
</tr>
<tr>
<td>F3</td>
<td>.18</td>
<td>−.09</td>
<td>.30*</td>
<td>.05</td>
<td>.13</td>
</tr>
<tr>
<td>F4</td>
<td>.06</td>
<td>−.12</td>
<td>.45*</td>
<td>.02</td>
<td>.01</td>
</tr>
</tbody>
</table>

Note: "p < .05, **p < .01 (two-tailed). Aggression LV = aggression latent variable which includes OSAB aggression and AQ-SF physical aggression.


