A PLEA FOR PREREGISTRATION IN PERSONALITY DISORDERS RESEARCH: THE CASE OF PSYCHOPATHY

Bruno Verschuere, PhD, Franziska M. Yasrebi-de Kom, MSc, Iza van Zelm, MSc, and Scott O. Lilienfeld, PhD

In response to a crisis of confidence, several methodological initiatives have been launched to improve the robustness of psychological science. Given its real-world implications, personality disorders research is all too important to not follow suit. The authors offer a plea for preregistration in personality disorders research, using psychopathic personality (psychopathy) as a prominent case example. To suit action to word, the authors report on a preregistered study and use it to help refute common misconceptions about preregistration as well as to illustrate that the key strength of preregistration: transparency outweighs its (perceived) disadvantages. Although preregistration will not conclusively settle the many debates roiling the field of psychopathy and other personality disorders, it can help to verify the robustness of empirical observations that inform such debates.

Keywords: psychopathy, replication, preregistration, open science, Psychopathy Checklist Revised (PCL-R), prototypical analysis

In the past decade in particular, a growing cadre of scholars has questioned the replicability of psychological science writ large (Aarts et al., 2015; Bakker, van Dijk, & Wicherts, 2012; Button et al., 2013; Hagger et al., 2016; Ioannidis, 2005; Klein et al., 2014; Simmons, Nelson, & Simonsohn, 2011; Simons, 2014). These concerns have considerable merit. For at least five decades, most psychological studies have been known to be substantially underpowered (Cohen, 1962; Sedlmeier & Gigerenzer, 1989). Conjoined with the growing realizations that researcher flexibility in data analyses (often colloquially referred collectively to as “p-hacking” techniques) can allow them to “detect” statistically significant results in noise (Simmons et al., 2011), and the well-established finding that effects that do not reach statistical significance are less likely than others to be published (the “file drawer effect”; see Dickersin, 1990;
Rosenthal, 1979), these problems have led a growing chorus of researchers to conclude that many published findings in psychological science—and science more broadly—are inflated in magnitude or false (Ioannidis, 2005).

Still, the replicability crisis has thus far exerted relatively little impact on psychopathology research (Tackett et al., 2017), including work on personality disorders. In particular, it is unclear to what extent published research on personality disorders, including one of the most widely studied of such conditions, namely, psychopathic personality (psychopathy), is biased. Many of our arguments also apply to research on most or perhaps even all other personality disorders. We chose psychopathy as a case example given the considerable intrinsic interest in the disorder owing to its frequent antisocial manifestations; the small sample sizes common in forensic research; routine use of multiple psychopathy indices, which themselves frequently subsume multiple subdimensions; and numerous analytic choice-points (e.g., use of psychopathy diagnoses versus dimensional scores). In light of these issues, we speculate that the problem of researcher bias in the psychopathy field may be even more pronounced than for other personality disorders. Even if our conjecture is incorrect, however, all of our core arguments regarding the replicability crisis continue to apply to the personality disorders field more broadly.

Absence of evidence for bias in published research on personality disorders (including psychopathy) should not be taken as evidence of absence. Interestingly, the question of whether certain findings in the psychopathy domain are attributable to p-hacking has been raised by previous scholars (e.g., Lykken, 1995). Indeed, there are several reasons to assume that the literature on psychopathy is biased.

First, given that psychopathy research in forensic samples is very resource-intensive, studies—particularly those that rely on more than survey or file review—are often underpowered to detect even medium-to-large effects. To illustrate, a seminal brain imaging study showing poor fear conditioning in psychopathy (Birbaumer et al., 2005; 791 Google Scholar citations) contrasted 10 psychopathic individuals with 10 healthy comparison participants. Another seminal study that reported on aberrant affective language processing in psychopathy contrasted eight psychopathic inmates with eight nonpsychopathic inmates (Williamson, Harpur, & Hare, 1991; 548 Google Scholar citations). And a well-cited study on abnormal moral reasoning in psychopathy compared 10 psychopathic individuals with 10 nonpsychopathic controls (Blair, 1995; 1,630 Google Scholar citations). Even Lykken’s (1957) enormously influential study on fear conditioning in psychopathy (1,438 Google Scholar citations) contrasted only 19 primary psychopaths with 15 comparison participants. Although psychopathy research is understandably resource-intensive, such underpowered studies are highly undesirable from a statistical perspective: Even when expecting large effects, such sample sizes lack power (Button et al., 2013). Low statistical power increases the risk for false negatives, while also increasing the risk of false-positive findings given a well-known statistical phenomenon known as the “winner’s curse,” whereby statistically significant results obtained in small samples are likely to be either false or inflated in magnitude.
Second, a great number of “sexy” and arguably counterintuitive findings have been reported in the psychopathy field. To name only a few recent examples, compared with nonpsychopathic individuals, psychopathic individuals have been reported to be more likely to be night owls (Jonason, Jones, & Lyons, 2013), to use words related to sex and money and less likely to use words related to family and religion (Hancock, Woodworth, & Porter, 2013), to have a poorer sense of smell (Mahmut & Stevenson, 2012), to be less likely to yawn contagiously (Rundle, Vaughn, & Stanford, 2015), to take more selfies (McCain et al., 2016), to exhibit higher face width to height ratios (Anderl et al., 2016), and more likely to be ambidextrous (Shobe & Desimone, 2016). Many of these intriguing results may be genuine. Without a clear separation between exploration and confirmation, however, conventional inferential statistics do not have clear evidential value (de Groot, 1956). Furthermore, from a Bayesian standpoint, surprising findings, especially those that run counter to previous theory or research, may be less likely to replicate than other findings (Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011).

Third, several effects in psychopathy research appear to be unstable. For instance, an increasing number of papers report on brain abnormalities in psychopathy, but there is a “remarkable heterogeneity” in imaging findings (Koenigs, Baskin-Sommers, Zeier, & Newman, 2011) that may be related to poor replicability. Likewise, results vary greatly whether and for which emotions (e.g., fear, sadness, disgust) psychopathic individuals exhibit deficits in recognition (Dawel, O’Kearney, McKone, & Palermo, 2012). Questions have also been raised concerning the replicability and robustness of key etiological models of psychopathy, including the response modulation model (Smith & Lilienfeld, 2015) and the low fear model (Hoppenbrouwers, Bulten, & Brazil, 2016).

Fourth, in psychopathy research, a great number of analytic decisions must typically be made. Such analytic choice-points may involve participant exclusion criteria (e.g., low IQ, psychosis, medication status, high scores on validity scales), covariates (e.g., for comorbid conditions, for social desirability), level of analyses (e.g., psychopathy total score, factor scores, or even subscale scores), the use of categorical (psychopathy versus nonpsychopathy) versus dimensional scores, potential moderator or stratification effects (e.g., psychopathy by trait anxiety), potential suppressor effects (e.g., should one control for other psychopathic traits when examining unique associations of certain psychopathic traits?), and transformation of psychopathy scores when they are skewed, to name only a few. To offer merely one example, some research teams testing the response modulation hypothesis of psychopathy have relied on Psychopathy Checklist Revised (PCL-R; Hare, 2003) total scores, others on PCL-R total scores stratified by trait anxiety, and still others on PCL-R factor scores (Lykken, 1995; Smith & Lilienfeld, 2015), rendering this body of literature challenging to evaluate without having access to the original data. Furthermore, because some psychopathy subdimensions, especially those relevant to boldness, exhibit low or at best modest relations with other subdimensions (Miller & Lynam, 2012), many psychopathy researchers have understandably elected to administer multiple psychopathy measures in
their studies to maximize content coverage. Combined with the further exploration of possible interaction effects between subdimensions, the potential for post hoc “cherry-picking” of psychopathy measures or effects to maximize the likelihood of statistically significant results (or what medical researchers term “outcome reporting bias”) is often high. Although reasonable justifications can often be offered for any number of such design and analytic decisions, it is difficult or impossible to properly evaluate them when it is unclear whether they were made “results-blind,” that is, prior to examining the initial findings (Simmons et al., 2011).

Fifth, many findings in psychopathy research are actively debated. In fact, there is considerable disagreement regarding virtually every aspect of psychopathy: its assessment (Watts et al., 2016), latent structure (Skeem & Cooke, 2010), childhood manifestations (Edens, Skeem, Cruise, & Cauffman, 2001), genetic underpinnings (Viding, 2004), neurobiological correlates (Koenigs et al., 2011), personality correlates (Miller & Lynam, 2015), the relative roles of adaptive versus maladaptive features (Verschuere et al., 2018), treatment responsivity (Salekin, 2002), gender, race, and cultural differences (Cooke, Michie, Hart, & Clark, 2005; Verona, Bresin, & Patrick, 2013), and implications for legal decision making (Boccaccini, Rufino, Jeon, & Murrie, 2017).

Of course, some of these concerns may be overstated. Researchers in clinical psychological science and psychiatry, who are accustomed to dealing with messy and imperfect data, may generate less biased results than those in other psychological disciplines (Tackett et al., 2017), although there are few good data bearing on this possibility. Also, the psychopathy field has seen its share of replications. For instance, associations between psychopathy and five-factor model traits (for a review, see Lynam & Miller, 2015), criminal behavior (Hemphill, Hare, & Wong, 1998; Salekin, Rogers, & Sewell, 1996), and startle eye blink modulation (e.g., Baskin-Sommers, Curtin, & Newman, 2013; Benning, Patrick, & Iacono, 2005; Patrick, Bradley, & Lang, 1993), to name a few examples, have been conceptually replicated in many studies (but for importance of direct replications, see Simons, 2014; Zwaan, Etz, Lucas, & Donnellan, 2017). Critically, however, none of these studies was preregistered. Preregistration offers tremendous benefits not only for initial studies but also for replication studies.

Preregistration refers to a time-stamped report of the research plan (for an introduction to preregistration, see de Groot, 2014; Nosek, Ebersole, DeHaven, & Mellor, 2018; Wagenmakers, Weszels, Borsboom, van der Maas, & Kievit, 2012). It is not so much about the why (why do you want to test a specific hypothesis?) as it is about the how (how will you test the hypothesis?). Throughout this article, we will argue that the key advantage of preregistration is transparency: It clarifies which decisions were made during which stage of the research (e.g., prior to or after inspecting the data), helping the reader to evaluate the strength of the presented evidence. By setting constraints on researcher degrees of freedom, and by more clearly distinguishing exploratory from confirmatory research, preregistration can enhance confidence on findings in psychopathy research. Despite the preregistration revolution (Nosek et al., 2018), the development of tools that make preregistration straightforward
PREREGISTER PSYCHOPATHY RESEARCH!

(e.g., aspredicted.org and several sources on the Open Science Framework, including several preregistration forms, https://osf.io/zab38/wiki/home/), and the clear benefits for the scientific community as well as individual researchers (https://www.psychologicalscience.org/observer/seven-selfish-reasons-for-preregistration#.WEASRaIrJZ0), psychopathy researchers have yet to embrace preregistration. A survey among researchers showed that the most common reasons for not preregistering studies are that (a) preregistration is not necessary when conducting exploratory or descriptive research, (b) preregistration is not required and does not necessarily assure higher quality, (c) preregistration only serves to avoid (the exceptional case of) fraud, (d) researchers do not know how to preregister, and (e) preregistration brings about extra (and unnecessary) burden (Washburn et al., 2018). In Table 1, we specified these and additional arguments against preregistration to the case of psychopathy research, and we discuss their validity. Although some of the arguments possess a degree of validity, many are largely or entirely invalid, stemming from common misconceptions about preregistration or an unwarranted lack of appreciation of the benefits of preregistration. In sum, we argue that psychopathy research, mirroring personality disorders research at large, would benefit enormously from preregistration.

We strongly advocate for preregistration in psychopathy research. To suit action to the word and illustrate our points, we present a report on a preregistered study. Because of the perceived obstacles for preregistered replication in clinical psychological science (Tackett et al., 2017), discussing some of the hurdles we encountered in our imperfect preregistration may reveal how they can be overcome. Our preregistered study concerns a direct replication of Verschuere and te Kaat (2018). In their search for core psychopathy features, Verschuere and te Kaat (2018) asked Dutch forensic mental health professionals to rank the relevance of 20 features to the construct of psychopathy. The features were the items of the PCL-R (Hare, 2003), the most widely used instrument in clinical and forensic practice for the assessment of psychopathy. The exploratory analyses showed that (a) the affective-interpersonal features of psychopathy were deemed more important than the lifestyle-antisocial features, (b) Callous/lack of empathy, Conning/manipulative, and Lack of remorse or guilt specifically were deemed to be most relevant, and (c) this subjective ranking aligned surprisingly well with three psychometric indices of feature importance. Given the exploratory nature of the exclusion criteria and the statistical analyses, it is difficult to ascertain how much confidence should be given to the robustness of those findings. Our preregistered study aimed at replicating these three key observations.1

We strongly advocate for preregistration in psychopathy research. To suit action to the word and illustrate our points, we present a report on a preregistered study. Because of the perceived obstacles for preregistered replication in clinical psychological science (Tackett et al., 2017), discussing some of the hurdles we encountered in our imperfect preregistration may reveal how they can be overcome. Our preregistered study concerns a direct replication of Verschuere and te Kaat (2018). In their search for core psychopathy features, Verschuere and te Kaat (2018) asked Dutch forensic mental health professionals to rank the relevance of 20 features to the construct of psychopathy. The features were the items of the PCL-R (Hare, 2003), the most widely used instrument in clinical and forensic practice for the assessment of psychopathy. The exploratory analyses showed that (a) the affective-interpersonal features of psychopathy were deemed more important than the lifestyle-antisocial features, (b) Callous/lack of empathy, Conning/manipulative, and Lack of remorse or guilt specifically were deemed to be most relevant, and (c) this subjective ranking aligned surprisingly well with three psychometric indices of feature importance. Given the exploratory nature of the exclusion criteria and the statistical analyses, it is difficult to ascertain how much confidence should be given to the robustness of those findings. Our preregistered study aimed at replicating these three key observations.1

1. Replications fall on a continuum from “direct” or “close” replications (recreating the original study as closely and faithfully as possible) to “conceptual” replications (testing the same hypothesis in a different way) (Brandt et al., 2014; Lykken, 1968). We consider our replication to be a direct one, because the changes made to the original study design were deemed small and unlikely to significantly change the findings. These changes were the following: (a) The instructions no longer restricted psychopathy to criminal psychopathy to avoid what would de-emphasize antisocial features; (b) we used the exact PCL-R item descriptions as in the Dutch PCL-R manual (Vertommen, Verheul, De Ruiter, & Hildebrand, 2002) rather than rephrasing them to make them maximally understandable; and (c) we recruited and included forensic behavioral experts only, specifying the required profession and experience to qualify as a forensic behavioral expert.
<table>
<thead>
<tr>
<th>Argument</th>
<th>Evaluation of argument and possible solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>I am afraid that preregistration hinders creativity and scientific discovery.</td>
<td>Preregistration does not prevent exploration. Exploration is explicitly allowed. Preregistration clarifies which findings were explicitly anticipated, and which discoveries were made after results were known (e.g., by identifying subtypes or using novel analyses).</td>
</tr>
<tr>
<td>Preregistration is fine for easy research in undergraduate samples, but it does not apply to hard-to-recruit (forensic) samples.</td>
<td>Why not? The fact that special samples are harder to assemble is irrelevant from a statistical perspective. The difficulty in collecting sufficiently large samples may be overcome by (a) teaming up and conducting multilab studies, for example, through the Society for Scientific Study of Psychopathy (<a href="https://www.psychopathy-society.org/en/">https://www.psychopathy-society.org/en/</a>) or Psychological Science Accelerator (<a href="https://psyciac">https://psyciac</a> crippleme.readme_file.html or <a href="https://epf.nl/projecten/lnd-bhs">https://epf.nl/projecten/lnd-bhs</a>).</td>
</tr>
<tr>
<td>The data have already been collected.</td>
<td>Preferably preregistration is done before data collection. Nevertheless, the use of archival data does not preclude the possibility of preregistration as long as there are guarantees that (part of) the data have not been inspected yet.</td>
</tr>
<tr>
<td>I am breaking new ground and don’t know yet exactly what to expect.</td>
<td>Great, go ahead and explore! Just be clear to readers what it is: exploration. Preregistration in no way precludes exploratory research as long as it is explicitly declared to be exploratory.</td>
</tr>
<tr>
<td>My hypothesis is counterintuitive.</td>
<td>Having a counterintuitive hypothesis is a great reason for preregistration. Preregistration shows that you did not HARK—Hypothesize After the Results are Known (Kerr, 1998).</td>
</tr>
<tr>
<td>The use of different psychopathy measures/factors/scales and/or ways to combine them (e.g., interactions, suppressor effects) provides a richer picture.</td>
<td>It does. But it also increases the number of statistical tests, potentially inflating the false-positive rate (de Groot, 2014).</td>
</tr>
<tr>
<td>Preregistration is just extra work—I know what I am doing.</td>
<td>You may know exactly what you are doing and why, but there is no way for readers to know. Plus, p hacking and HARKing may often be largely unintentional, especially when they are conducted over a period of many months, allowing researchers to forget what analyses they had conducted and predicted earlier.</td>
</tr>
<tr>
<td>Fraud is exceptional/I should not be considered a possible fraud/Preregistration is not necessary when you are honest.</td>
<td>Fraud is indeed exceptional, and fraud detection is not a prime purpose of preregistration. False positives are, however, far from exceptional (Ioannidis, 2005), and perhaps only about 50% of findings in psychology replicate (Camerer et al., 2018). Ironically, inadvertent biases in human decision making are among the best replicable findings in psychology research (Klein et al., 2014), and researchers are not immune to them. Preregistration exemplifies the late physicist Richard Feynman’s astute point that science, at its best, is a recipe for minimizing (of course, not eliminating) the odds that we are fooled.</td>
</tr>
<tr>
<td>Submitting my research proposal for review before conducting the study slows down research and/or is not practically feasible (e.g., given availability of research assistants or interns).</td>
<td>Preregistration is often confused with preregistered reports (Chambers, 2013). Preregistered reports are a special kind of preregistrations because they involve peer review. Preregistered reports have—when accepted after peer review—the advantage of in principle acceptance of publication, irrespective of study findings, thereby reducing publication bias. A disadvantage of preregistered reports is that there is unknown and sometimes considerable time between proposing the idea and the start of data collection. A possible solution is to submit well in advance. Another solution is to stick to preregistration, which can be quite streamlined and simple (see, e.g., aspredicted.org).</td>
</tr>
<tr>
<td>One can preregister after the data were analyzed.</td>
<td>One can. This is called fraud.</td>
</tr>
</tbody>
</table>
METHOD

The study was approved by the ethical committee of the Psychology Department of the University of Amsterdam and classified under number 2018-CP-8748. We preregistered the hypotheses, analyses, and inferential criteria on the Open Science Framework: https://osf.io/bgwxq/register/565fb3678c5e4a66b5582f67. After the start of data collection, we made an amendment to the preregistration with regard to one of the exclusion criteria: https://osf.io/zh635/register/564d31db8c5e4a7c9694b2be (discussed below in the Participants section).

PARTICIPANTS

We ran an a priori power analysis with the program G*Power 3.1.9.2 (Faul, Erdfelder, Lang, & Buchner, 2007) for each of the three targeted effects. To test the hypothesis that affective-interpersonal features would be deemed more important by raters than lifestyle-antisocial features, we planned a repeated-measures analysis of variance (ANOVA) with PCL-R factor as a within-subjects factor (four levels: affective, interpersonal, lifestyle, antisocial), expecting a main effect of PCL-R factor. The minimal required sample size to test whether
there is a main effect of PCL-R factor on the ranking of the PCL-R items ($f = .25$, $\alpha = 0.05$, $1-\beta = 0.8$, one-tailed) was found to be $n = 32$ when a modest ($r = .3$) correlation between measures was assumed. To test the hypothesis that the features Callous/lack of empathy, Conning/manipulative, and Lack of remorse or guilt are among the most important items, we used the logic of the Helmert contrast and planned to conduct three paired sample $t$ tests, comparing (a) Callous/lack of empathy with the mean of subsequently ranked features, (b) Conning/manipulative with the mean of the subsequently ranked features, and (c) Lack of remorse or guilt with the mean of the subsequently ranked features. For these $t$ tests ($d = .5$, $\alpha = 0.05$, $1-\beta = 0.8$, one-tailed), a minimum of $n = 27$ was required. To test the hypothesis that the ranking of the forensic behavioral experts would align with three psychometric indices of feature importance (i.e., network centrality, item-total correlation, and item response theory (IRT) discrimination parameter), we planned three Spearman’s rho correlations that required a minimum of $n = 67$ for $r = .3$, $\alpha = 0.05$, $1-\beta = 0.8$, one-tailed. Thus, the minimum required sample size for the three effects was $n = 67$.

Data collection began on April 13, 2018. Our stopping rule was that data collection would be terminated on July 1, 2018, if the minimum required sample size ($n = 67$) would be reached or to continue until the minimum required sample size was reached. Data collection was completed on July 1, 2018.

The sample consisted of Dutch-speaking forensic behavioral experts, defined as participants who reported having at least 1 year of work experience (including the time of internships) as behavioral experts in the forensic setting. Behavioral experts included the following professions: sociotherapist, psychologist, psychological assistant, psychiatrist, pedagogue, and sexologist.

Exclusion criteria were (a) having less than 1 year of work experience as a behavioral expert in the forensic setting, and (b) not completing the survey. We initially had imposed a third exclusion criterion—spending less than 4 minutes on the ranking page—but we realized soon after the start of the data collection that the necessary information to evaluate this criterion was not recorded. Furthermore, it became apparent that this criterion was probably unrealistic because indirect estimates indicated it would lead to exclusion of up to half of the sample. We therefore made an amendment to our preregistration (https://osf.io/zh635/register/564d31db8c5e4a7c9694b2be), explaining why we chose to drop this third exclusion criterion (for subsidiary analyses applying the exclusion criterion, see https://osf.io/xk5ze/).

Of the 167 participants who accessed the website, 119 completed the survey, 5 had less than 1 year of work experience as a behavioral expert in the forensic setting, and 15 had a profession other than behavioral expert in the forensic setting. Data of one participant were lost due to a technical error. The data for all participants, included and excluded, are available on https://osf.io/uzg8t/. The final sample consisted of 98 participants (31% male). The vast majority were university educated (86%). Most were psychologists (59%). Only a minority (38%) had taken the 3-day formal PCL-R training. Participants were, on average, 38.01 years old ($SD = 11.31$) and had 10.32 years of work experience ($SD = 8.78$).
MATERIALS
An online Dutch Qualtrics survey was programmed (the original: https://osf.io/wp7h2/; English translation: https://osf.io/rds27/). After providing informed consent, each participant was provided with the 20 items of the PCL-R one below the other, with the order determined at random for each participant. The item description followed the exact description provided in the PCL-R manual (Vertommen, Verheul, De Ruiter, & Hildebrand, 2002) with the exception of the item *Revocation of conditional release*, which was abbreviated due to an overly lengthy description of the Dutch translation (124 characters). Participants were instructed to order the items by dragging them to the desired position on the page according to the item’s importance, whereby a score of 1 (top of the page) indicated *very important* and a score of 20 (bottom of the page) indicated *least important*. After completion of the ranking, participants provided information on gender, age, education, years of work experience in the forensic setting, profession, and whether they had attended a PCL-R training. Lastly, they were provided with a list of possible sources of information that they may have used to determine their ranking (news, education, personal experience with individuals with the psychopathy diagnosis, fictional movies, intuition, diagnostic assessment, scientific articles, and “other,” with the option to enter a description of the last option in a textbox). Participants were asked to rank-order these sources in descending order of importance following the same method used when ordering the PCL-R items.

PROCEDURE
The online Dutch Qualtrics survey was distributed (a) by direct, broad e-mailing to behavioral experts working in a forensic setting, (b) on social media (LinkedIn, Twitter, Facebook, and a specific platform used by Dutch forensic experts: https://www.knapp-efp.nl), (c) at a Dutch forensic conference, and (d) by site visits to forensic institutions.

RESULTS
Given the directional hypotheses, all *p* values are reported using one-tailed testing and an alpha of .05.

CONFIRMATORY ANALYSES
*Target effect #1: Affective-interpersonal features are deemed more important than lifestyle-antisocial features.* We calculated the harmonic means of judged item importance for each PCL-R factor and plotted them as shown in Figure 1. The repeated-measures ANOVA on judged importance with PCL-R factor score as the within-subjects factor (four levels: affective, interpersonal, lifestyle, antisocial) showed a significant main effect of PCL-R factor, *F*(2.66, 257.82) = 178.69,
greenhouse–geisser correction. Planned follow-up paired sample t tests revealed that the affective features were rated as more important than the lifestyle features, \( t(97) = 17.59, p < .001, d = 1.78 \), and the antisocial features, \( t(97) = 18.48, p < .001, d = 1.87 \), and that the interpersonal features were rated as more important than the lifestyle features, \( t(97) = -12.52, p < .001, d = 1.27 \), and the antisocial features, \( t(97) = -14.14, p < .001, d = 1.43 \).

**Target effect #2:** Callous/lack of empathy, Conning/manipulative, and Lack of remorse or guilt are among the core psychopathy features. Using the logic of the Helmert contrast, the ranking of Callous/lack of empathy was higher than the average ranking of the subsequent 19 items (\( M = 10.91, SD = .14 \)), \( t(97) = 28.86, p < .001, d = 2.91 \), the ranking of Lack of remorse or guilt was higher than the average ranking of the subsequent 18 items (\( M = 11.39, SD = .27 \)), \( t(97) = -19.84, p < .001, d = 2.00 \), and the ranking of Conning/manipulative was higher than the average ranking of the subsequent 17 items (\( M = 11.67, SD = .33 \)), \( t(97) = 17.71, p < .001, d = 1.79 \) (see Table 2).

**Target effect #3:** The subjective ranking of the experts aligns with psychometric indices of item importance. Spearman’s rho correlation coefficients between the expert ranking and ranked network centrality (Verschuere et al., 2018), \( r = 0.66, p < .001 \), ranked item-total correlation (Hare et al., 1990), \( r = 0.73, p < .001 \), and ranked item response parameters (Bolt, Hare, Vitale, & Newman, 2004), \( r = 0.83, p < .001 \), were significant, positive, and large (i.e., \( r_s > .5 \)).

**EXPLORATORY ANALYSES**

There was a very strong relation between the current expert rankings and those reported in the original study (Verschuere & te Kaat, 2018), \( r = 0.98, p < .001 \).

Participants reported basing their rankings in descending order of importance on the following sources: education (\( M = 2.34, SD = 1.48 \)), personal experience (\( M = 2.56, SD = 1.57 \)), scientific research (\( M = 3.65, SD = 1.74 \)), diagnostic assessments (\( M = 3.83, SD = 1.84 \)), intuition (\( M = 4.83, SD = 1.88 \)), news (\( M = 5.27, SD = 1.56 \)), fiction (\( M = 6.60, SD = 1.24 \)), and other sources (\( M = 6.93, SD = 1.73 \)).

In sum, the three key observations of Verschuere and te Kaat (2018) were replicable. Although encouraging, the results would arguably have been more convincing had the replication attempt been conducted by a replication team that was fully independent of the original team of researchers and had we made an effort to assure nonoverlap between the original and the replication samples (an exploratory analysis excluding possible overlap also replicated the three targeted effects: https://osf.io/xk5ze/). These limitations notwithstanding, the findings point to a growing consensus concerning what may be considered the core features of psychopathy, which have been variously described

---

\(^2\) Upon request of a reviewer, we report \( \omega^2 \) (Olejnik & Algina, 2003, calculated with JASP (JASPTeam, n.d.), as a measure of effect size. We did not preregister the use of \( \omega^2 \) and consequently do not use it to evaluate our hypotheses.
FIGURE 1. Box plot of judged importance of PCL-R factors. The box represents 50% of the observations, with the band inside the box displaying the mean. The whiskers (the lines extending vertically from the boxes) extend to 1.5 times the interquartile range from the box. The dots are outliers, defined as values beyond the whiskers. Item importance is reverse scored for ease of interpretation (higher scores display higher importance).

TABLE 2. PCL-R Items and Factors in Descending Order of Ranked Item Importance (M, SD)

<table>
<thead>
<tr>
<th>PCL-R Item</th>
<th>PCL-R Factor (Hare, 2003)</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Callous/Lack of Empathy</td>
<td>Affective</td>
<td>2.74</td>
<td>2.66</td>
</tr>
<tr>
<td>Lack of Remorse or Guilt</td>
<td>Affective</td>
<td>4.04</td>
<td>3.39</td>
</tr>
<tr>
<td>Conning/Manipulative</td>
<td>Interpersonal</td>
<td>4.86</td>
<td>3.63</td>
</tr>
<tr>
<td>Glibness/Superficial Charm</td>
<td>Interpersonal</td>
<td>6.51</td>
<td>4.11</td>
</tr>
<tr>
<td>Shallow Affect</td>
<td>Affective</td>
<td>7.64</td>
<td>4.61</td>
</tr>
<tr>
<td>Pathological Lying</td>
<td>Interpersonal</td>
<td>8.11</td>
<td>4.53</td>
</tr>
<tr>
<td>Grandiose Sense of Self-Worth</td>
<td>Interpersonal</td>
<td>8.55</td>
<td>4.60</td>
</tr>
<tr>
<td>Failure to Accept Responsibility for Own Actions</td>
<td>Affective</td>
<td>8.64</td>
<td>4.42</td>
</tr>
<tr>
<td>Parastic Lifestyle</td>
<td>Lifestyle</td>
<td>9.07</td>
<td>5.04</td>
</tr>
<tr>
<td>Need for Stimulation/Proneness to Boredom</td>
<td>Lifestyle</td>
<td>11.15</td>
<td>4.28</td>
</tr>
<tr>
<td>Early Behavioral Problems</td>
<td>Antisocial</td>
<td>11.77</td>
<td>4.66</td>
</tr>
<tr>
<td>Irresponsibility</td>
<td>Lifestyle</td>
<td>12.09</td>
<td>3.68</td>
</tr>
<tr>
<td>Impulsivity</td>
<td>Lifestyle</td>
<td>12.27</td>
<td>4.63</td>
</tr>
<tr>
<td>Poor Behavioral Controls</td>
<td>Antisocial</td>
<td>12.30</td>
<td>4.31</td>
</tr>
<tr>
<td>Criminal Versatility</td>
<td>Antisocial</td>
<td>12.89</td>
<td>4.83</td>
</tr>
<tr>
<td>Juvenile Delinquency</td>
<td>Antisocial</td>
<td>13.71</td>
<td>4.60</td>
</tr>
<tr>
<td>Promiscuous Sexual Behavior</td>
<td>Other</td>
<td>15.49</td>
<td>3.92</td>
</tr>
<tr>
<td>Many Short-term Marital Relationships</td>
<td>Other</td>
<td>15.62</td>
<td>3.86</td>
</tr>
<tr>
<td>Lack of Realistic Long-term Goals</td>
<td>Lifestyle</td>
<td>15.96</td>
<td>3.78</td>
</tr>
<tr>
<td>Revocation of Conditional Release</td>
<td>Antisocial</td>
<td>16.58</td>
<td>3.45</td>
</tr>
</tbody>
</table>

Note. PCL-R = Psychopathy Checklist-Revised. Judged importance ranges from 1 = most important to 20 = least important.
(with relatively minor differences in content) as guiltlessness and lovelessness (McCord & McCord, 1964), callous-unemotional or CU traits (Frick & White, 2008), antagonism (Miller & Lynam, 2015), meanness (Patrick, Fowles, & Krueger, 2009), and coldheartedness (Lilienfeld & Andrews, 1996), and that “Antagonism, should be considered a core feature, perhaps the core feature, of psychopathy” (Sherman, Lynam, & Heyde, 2014, p. 275; see also Lynam & Miller, 2015; Miller, Lynam, Widiger, & Leukefeld, 2001).

DISCUSSION

Given the abundant evidence of widespread biases in research and publication practices in a great variety of disciplines, both within and outside of psychology (Bakker et al., 2012), we adopt the conservative position that there is likely to be a systemic problem with personality disorders research at large, until proven otherwise. To illustrate this point, we have selected one prominent and widely researched personality condition, namely psychopathy, as a case example. As noted earlier, there are several grounds, such as the small sample sizes in most forensic research, routine use of multiple psychopathy indices, and numerous analytic choice-points (e.g., use of total versus subscale scores, use of categorical psychopathy diagnoses versus dimensional scores), for speculating that the problem of researcher bias in the psychopathy field may be even more pronounced than in most other domains of personality disorder research. We have therefore made a plea for preregistration in psychopathy research. We also reported on a preregistered study. The current literature review and our imperfect preregistration lead us to make three points for further consideration.

First, preregistrations are helpful but imperfect safeguards against investigator bias. To illustrate, we preregistered an exclusion criterion that we later considered invalid, and that would have unnecessarily harmed the power of the study. Although extensive thought and piloting can improve them, researchers will often be confronted with imperfect preregistrations: What to do with unexpected skewness of the data? What to do if you did not preregister the decision to exclude severely intoxicated participants? Or more extremely, what if one simply preregistered the wrong analyses? One could easily use these possibilities to argue against preregistration. But “preregistrations are a plan, not a prison” (https://cos.io/blog/preregistration-plan-not-prison/). Indeed, even imperfect preregistrations serve their purpose: They make key decisions in data collection, analyses, and reporting transparent and explicit, and they force researchers to more explicitly distinguish the exploratory versus confirmatory modes of the scientific endeavor.

Second, despite the preregistration of our research plan, a reviewer noted several shortcomings (e.g., assure nonoverlap between the original study and the replication study) that we could have addressed with an improved methodology. This observation makes the case for a specific form of preregistration: the Preregistered Report format (Chambers, 2013). With this format, one submits the preregistered research plan to a journal for peer review, and acceptance of the proposal leads to in-principle acceptance of the research,
irrespective of the findings. This approach not only reduces publication bias, but also allows peer reviewers to weigh in during early stages of the research. In the past 5 years, more than 148 journals (https://cos.io/rr/) have adopted the Preregistered Report format, including several journals that publish psychopathy research (e.g., British Journal of Clinical Psychology, Cognition & Emotion, Clinical Psychology in Europe, Journal of Research in Personality, Legal & Criminological Psychology, and Psychological Science).

Third, as psychologists, we know that change is hard, especially when it involves changing what we have been doing for years. Psychopathy researchers should be helped with making the change to preregistration. Some of the common reasons not to preregister are that it is not necessary, not required, and not rewarded (Washburn et al., 2018; see also Table 1). Here we see a key role for The Society for the Scientific Study of Psychopathy (SSSP), which promotes education in, conduct of, and communication of psychopathy research (https://www.psychopathysociety.org/en/). There are several ways in which the society could help overcome common hurdles to adopt preregistration, including (a) encouraging its members to preregister their studies, (b) encouraging those members who serve on journal board committees for those journals to adopt the Preregistered Report format, and (c) awarding open science badges (for preregistration, but also for sharing data and materials) to contributions to the biannual SSSP meeting.

In sum, given its real-world impact, psychopathy research, along with other domains of personality disorders research, is all too important to neglect the replication crisis in psychology. Preregistrations can help correct biases in psychopathy research and, we hope, enhance the robustness and accuracy of findings in this field. With these points in mind, we strongly encourage researchers in the personality disorders field more generally to heed these recommendations.

REFERENCES


PREREGRINTER PSYCHOPATHY RESEARCH!

Assessment, 2, 338–341. https://doi.org/10.1037/1040-3590.2.3.338
JASP Team. (n.d.). JASP version 0.9.2 [Computer software]. Amsterdam, the Netherlands: Author. https://jasp-stats.org/
Salekin, R. T. (2002). Psychopathy and therapeutic pessimism: Clinical lore or clinical


