

Informant-Reports of Personality Disorder: Relation to Self-Reports and Future Research Directions

E. David Klonsky, Thomas F. Oltmanns, and Eric Turkheimer
University of Virginia

Most instruments that assess personality disorder rely principally on self-report. However, there are major limitations to using self-report for the diagnosis of personality pathology. First, the self provides only one opinion. Second, like everyone else, people with personality disorders are frequently unable to view themselves realistically and are unaware of the effect of their behavior on other people. Using informant data, therefore, may improve the reliability and validity of diagnostic assessment. A review of the literature reveals that agreement between informant- and self-reports of personality disorder is modest at best, even though informants tend to agree with each other. Self-informant concordance appears to be higher for older subjects and for Cluster B traits (excluding narcissism). Further research should focus on methods of resolving discrepancies between self- and informant-reports of personality disorder and determining when self or informant data are more valid.

Key words: personality disorders, informants, self-report, assessment, Axis-II. [*Clin Psychol Sci Prac* 9:300-311, 2002]

WHY INFORMANT-REPORTS ARE IMPORTANT FOR ASSESSING PERSONALITY DISORDER

The valid assessment of personality disorder presents a challenge to clinicians and researchers alike. Numerous instruments have been created on the basis of DSM criteria to measure personality disorder (e.g., Clark, 1993; First, Spitzer, Gibbon, & Williams, 1995; Hyler et al., 1988;

Address correspondence to Thomas F. Oltmanns, Department of Psychology, University of Virginia, 102 Gilmer Hall, P.O. Box 400400, Charlottesville, VA 22904-4400. E-mail: tfo@virginia.edu.

Loranger, Susman, Oldham, & Russakoff, 1987; Pfohl, Blum, & Zimmerman, 1997). These instruments generally have demonstrated good reliability. Across 15 studies, interrater reliability (kappa) for semistructured interviews was calculated to be .60 (Zimmerman, 1994). However, the validity of these instruments remains open to question (Clark, Livesley, & Morey, 1997). Research has yielded mixed results regarding convergence between different instruments ostensibly assessing the same DSM criteria. Estimates of the median kappa for agreement among pairs of interviews have ranged from .25 (Farmer, 2000) to .53 (Skodol, Oldham, Rosnick, Kellman, & Hyler, 1991). For self-report inventories, the median correlation of agreement is .57 (Widiger & Coker, in press). Convergence between different instruments ostensibly assessing the same DSM criteria is less than ideal.

One major problem is that these instruments are based exclusively on self-report. This issue applies not only to paper-and-pencil self-report questionnaires but also to structured interviews, composed of questions answered verbally by the subject. Self-report has two major limitations. First, the self provides only one opinion, but basic psychometric theory favors the use of multiple sources of information. Many independent respondents are generally more reliable than a single source. Second, like everyone else, people with personality disorders are frequently unable to view themselves realistically and unaware of the effect of their behavior on other people. Their self-report is likely to include biased and misleading information or to omit important and relevant information. A more accurate description of personality problems might be obtained by collecting information from associates, friends, or family members.

A few researchers have chosen to rely exclusively on informant reports in assessing personality disorder. Cooney

(1996) sought to determine the prevalence of personality disorders in two Irish hospitals. He collected data from 78 consecutively admitted patients. ICD-10 draft criteria were used to diagnose both Axis I and Axis II disorders. Whereas Cooney diagnosed Axis I conditions on the basis of interview and patient records, he assessed personality disorders using reports from informants (i.e., next of kin, or the person who knew the subject best). Similarly, Pilgrim and Mann (1990) relied exclusively on informant data to determine the prevalence of personality disorders in psychiatric inpatients. They reasoned that using informant-reports and not self-reports avoids the inevitable distortions of patients with severe psychopathology. Of course, relying exclusively on informants has its own limitations, including the fact that informants provide only one kind of perspective and do not have access to all the thoughts and feelings of the person.

The majority of personality disorder studies continue to rely almost exclusively on subject-reports (i.e., either self-report questionnaires or structured interviews) to assess personality disorder. Several prominent personality disorder researchers have argued for the importance of incorporating informant-reports into studies on personality disorder (e.g., Clark et al., 1997; Grove & Tellegen, 1991; Westen & Shedler, 1999). Informants could be expected to improve diagnostic assessment in at least two ways. First, to the extent informant-reports provide information independent of data provided by self reports, informants will increase the reliability of assessment because random errors of measurement in self-report and informant-report will tend to cancel out. Second, if self-report data contain systematic biases, informant-report data may improve the validity of assessment by providing data without these biases, although informants will, of course, have biases of their own.

As a result of this need, researchers are beginning to collect information about personality disorder traits from informants. Such studies tend to take one of two methodological approaches. Most present their results (e.g., prevalence rates, correlations with other variables) separately for informant- and subject-based diagnoses (e.g., Peselow, Sanfilippo, & Fieve, 1994, 1995). Others combine subject and informant data by using clinicians who examine the informant- and self-report for each subject and use their own judgment to resolve discrepancies (e.g., Bernstein et al., 1997). Personality disorder prevalence rates or correlations with other variables are then derived from these diagnoses.

Such use of informant data is a step forward in personality disorder research, but many fundamental questions remain. How strong is self and informant agreement on personality disorder traits? What factors influence concordance between self- and informant-reports? If discordant, which source of information is more valid, and when? How should one resolve discrepancies between self- and informant-reports? Who reports more pathology, subjects or informants? What kind of informant should be selected? How many informants should be interviewed? To address these questions, this article presents a quantitative review of studies that used both informant- and self-reports of personality disorder.

STUDIES REVIEWED

A literature search was conducted to identify all empirical studies examining informant and subject reports of personality disorder. Theoretical and review papers and book chapters not including original data were excluded, as were unpublished dissertations.¹ A computerized literature search of PsycINFO was conducted, using the following search string: “(personality disorder OR personality disorders OR personality pathology OR axis-II) AND (informant OR informants OR peer OR peers OR friend OR friends).” The time period covered 1887 through February 2001. This search resulted in 182 citations. Twenty-nine studies met inclusion criteria. Examination of the reference lists of these articles revealed one additional study that met inclusion criteria, for a total of 30 meeting inclusion criteria. Of note, three studies were not included because they reported informant- but not self-reports of personality disorder (i.e., Cooney, Farren, & Clare, 1996; Pilgrim & Mann, 1990; Pilgrim, Mellers, Boothby, & Mann, 1993).

We found a total of 30 studies that included original informant and subject data on domains of personality pathology. Eight of these studies did not report self-informant agreement and were therefore not included in this review (i.e., Bernstein, Cohen, Skodol, Bezirgianian, & Brook, 1996; Brothwell, Casey, & Tyrer, 1992; Klein, Ouimette, Kelly, Ferro, & Riso, 1994; Pepper et al., 1995; Pfohl, Coryell, Zimmerman, & Stangl, 1986; Tyrer et al., 1984, 1990; Zimmerman, Pfohl, Stangl, & Corenthal, 1986). Five additional studies were excluded for methodological issues (Allard & Grann, 2000; Dowson, 1992b; Molinari, Kunik, Mulsant, & Rifai, 1998; Steketee, 1990; Vine & Steingart, 1994).²

Table 1. Methods and findings of studies reporting agreement between self and informant reports of the DSM personality disorders

Study	Self Instrument	Informant Instrument	n	Mean Age	Sample Type	# Informants	Median r	Median κ	Who Reports More Pathology?
Modestin & Puhan (2000)	SCID-IIQ	SCID-IIQ	32	31	Psychiatric patients	2		.13 Parent .14 Sibling	Informant
Dreessen et al. (1998)	SCID-II	SCID-II	42	33	Outpatients	1	.31		Self
McKeeman & Erickson (1997)	SCID-II	SCID-II	75	21	College	1	.27		Self
Bernstein et al. (1997)	SIDP-III	SIDP-III	62	40	In- and outpatients	1	.49	.33	Equivalent
Ferro & Klein (1997)	PDE	FHIPD	224	32	Outpatients and controls	1 or 2	.18	.10	Informant
Ouimette & Klein (1995)	SCID-IIQ	SCID-IIQ	97	19	College	1	.35		Not reported
Coolidge et al. (1995)	CATI	CATI	52	37	Married	2	.51 spouse .36 friends		Not reported
Peselow et al. (1995)	SIDP-III	SIDP-III	66	43	Outpatients	1		.31	Informant
Peselow et al. (1994)	SIDP-III	SIDP-III	58	42	Outpatients	1	.80		Informant
Riso et al. (1994)	PDE	PDE	105	32	Outpatients	1	.36	.06	Self
Dowson (1992a)	PDQ-R	PDQ-R	60	34	Psychiatric patients	1	.38		Self
Zimmerman et al. (1988)	SIDP-III	SIDP-III	66	37	Inpatients	1	.28		Informant
Median correlation coefficient across all studies							.36		
Median kappa coefficient across all studies								.14	

Note. CATI = Coolidge Axis-II Inventory; PDE = Personality Disorder Examination; PDQ-R = Personality Diagnostic Questionnaire-Revised; SCID-II = Structured Clinical Interview for DSM Axis-II; SCIDIQ = Structured Clinical Interview for DSM Axis-II-Questionnaire Version; SIDP-III = Structured Interview for DSM-III Personality.

The 17 remaining studies that report agreement between self- and informant-reports of personality disorder compose the focus of this review. Thirteen studies used DSM domains of personality disorder, and four did not. The methods and findings of the studies relying on DSM personality disorder diagnoses are presented in Table 1.³ The methods and findings of the non-DSM studies are presented in Table 2.

A PRELIMINARY METHODOLOGICAL CONSIDERATION

Not all studies listed in Tables 1 and 2 reported self-informant agreement in the same manner. Ten measured agreement between dimensional scores of personality disorder traits, three measured agreement between categorical diagnoses, and three used both methods. Pearson correlation coefficients were used to measure agreement in 5 of the 13 studies using dimensional scores; 5 others used intraclass correlations, and 3 did not specify kind of correlation. Intraclass correlations differ from Pearson correlations in that mean differences between raters are classified as error and contribute to lower correlations (Cronbach, Gleser, Nanda, & Rajaratnam, 1971). For this reason, intraclass correlations will generally be lower than Pearson cor-

relations computed on the same data and would probably be a more appropriate measure of interrater agreement in the studies being reviewed here. Kappa coefficients were calculated to measure diagnostic agreement when a categorical approach was utilized. An intraclass correlation computed between two sets of categorical diagnoses made by the informant and subject is mathematically equivalent to computing kappa (Fleiss, 1973).

ANALYSIS OF FINDINGS FROM THE LITERATURE

Summary of Self/Informant Agreement Across Studies

What kind of agreement would we expect to find between self- and informant-reports of the personality disorders? The literature on normal personality suggests that concordance between self- and other reports of personality traits ranges from modest to moderately high. Kenny (1994) summarized results from four studies examining self-other agreement on the Big Five factors. The median correlation found for extraversion was quite high ($r = .70$). However, median correlations for openness to experience, neuroticism, agreeableness, and conscientiousness were considerably lower (.51, .44, .42, and .39, respectively). Of all the Big Five traits, neuroticism may be most

Table 2. Methods and findings of studies reporting agreement between self and informant reports of non-DSM personality disorders

Study	Self Instrument	Informant Instrument	n	Mean Age	Sample Type	# Informants	Median r	Median κ	Who Reports More Pathology?
Ready et al. (2000)	SNAP	SNAP	189	20	College	1	.47		Not reported
Mann et al. (1999)	IPDE	SAP	90	34	Outpatients	1		.40	Informant
Hill et al. (1995)	APFA	APFA	56	41	Patients	1	.61		Equivalent
Tyrer et al. (1979)	PAS	PAS	15	Not reported	Inpatients	1	.12		Not reported
Median correlation							.47		
Median kappa								.40	

Note. APFA = Adult Personality Functioning Assessment; IPDE = International Personality Disorder Examination; PAS = Personality Assessment Schedule; SNAP = Schedule for Nonadaptive and Adaptive Personality.

similar to traits of personality disorder. We therefore might expect the modest self/other correlation for neuroticism to approximate the level of concordance between self- and informant-reports of personality disorder.

The most notable finding from the literature on self- and informant-reports of DSM personality disorder is that self/informant concordance is modest at best. Median correlation and kappa coefficients for studies reporting self-informant agreement on the DSM personality disorders are presented in Table 1. Across the 10 studies using continuous measures, the median correlation is .36. For the six studies using categorical diagnoses, median kappa is .14. It is not surprising that the median kappa is lower than the median correlation, because the categorical approach to personality disorders is more susceptible to measurement error. For dichotomous diagnoses, small errors of measurement that would make little difference in a continuous score will tend to change the categorical diagnoses of patients falling close to the diagnostic criterion.

The range of self/informant correlations reported is wide. The median correlations across studies ranged from .18 to .80. What accounts for the substantial variation in self/informant correlations across studies? Several variables may cause the correlation to vary. We will consider the possible roles of the assessment instrument administered, type of sample, kind of informant, sample age, and sample size.

Does Self/Informant Agreement Vary by Type of Assessment Instrument?

Several different measures of personality disorder have been used by the studies examined in this article (i.e., SIDP, PDE, IPDE, SCID-II, SCID-IIQ, PDQ-R, SNAP,

FHIPD, and CATI) (these acronyms are spelled out in the Table 1 and Table 2 notes). These instruments are similar in that they all assess each personality disorder criterion one at a time. In addition, those appearing most often in this literature (i.e., SIDP, PDE, SCID-II) have similar reliabilities (Zimmerman, 1994). However, distinctions among the various measures may affect the kind of information obtained from subjects and informants. First, the SCID-IIQ, PDQ-R, and CATI are self-report questionnaires, whereas the SIDP, PDE, IPDE, SCID-II, and FHIPD are structured interviews. Though both questionnaires and interviews assess the same personality disorder criteria, interpretive standards for questionnaires tend to yield more abnormal findings than interviews (Zimmerman, 1994). In addition, interviews allow for responses to be followed up, elaborated on, and clarified. A person's idiosyncratic understanding of a criterion can be detected and rectified through the administrator's clinical judgment. For these reasons, self-report questionnaires and structured interviews may differ in how they assess personality pathology.

A second manner in which instruments assessing the DSM personality disorders differ is in the organization of the items measured. The SCID-II contains separate modules for each class of diagnoses. That is, items are grouped by diagnosis. In contrast, items in the SIDP and PDE are organized into topical sections rather than by disorder. The domains of personality disorder being targeted by the former system of organization may become apparent to the subject since questions relating to each domain are asked consecutively. This kind of diagnostic organization may therefore be more susceptible to impression management. Topically oriented interviews may better conceal

the diagnostic categories being measured and thereby lead to a more accurate assessment.

Third, the content of items across measures differs, even though the items attempt to measure the same personality disorder criteria. For example, the SIDP, PDE, and PDQ often employ different questions to measure identical criteria. It should not be surprising then that the convergent validity of the various structured interviews is less than ideal, and convergence between self-report questionnaires and structured interviews still worse (Farmer, 2000; Zimmerman, 1994).

Last, personality disorder instruments may differ in how they are adapted for use with an informant. Both the SCID-II and SIDP-IV note the option of interviewing an informant by asking the questions in the third person. The SIDP-IV also stars particular questions to be asked of an informant that are a subset of the ones asked of the respondent and mentions that suitable follow-up questions may need to be composed or adapted (Pfohl et al., 1997). However, for neither instrument is there a separate version intended for use with informants or comprehensive instructions about how to adapt the instrument to collect information from informants. To our knowledge, the Standardized Assessment of Personality (SAP; Mann, Jenkins, Cutting, & Cowen, 1981) represents the first personality disorder measure created specifically to gather informant data. It was intended as a brief interview of an informant that could be conducted in person or over the telephone. The SAP begins with a sequence of questions designed to elicit key words relevant to the categories of personality disorder. If the key words are elicited, further investigation of the relevant personality disorder categories is conducted. More recently, the SNAP (Clark, 1993) was adapted for use with informants. The SNAP is a factor analytically derived self-report inventory composed of 375 true/false items designed to assess trait dimensions in the domain of personality disorders. It also includes diagnostic scales corresponding to each of the DSM-III-R personality disorders. A collateral-report version of the SNAP (SNAP-CRV) was created by rewording all questions in the third person (Ready, Clark, Watson, & Westerhouse, 2000). Additionally, some questions were rephrased because their meaning became awkward or altered as a result of being reworded in the third person. Unfortunately, the SNAP-CRV does not contain diagnostic scales corresponding to the DSM-IV personality disorders.

Given the many differences among inventories measuring personality disorder, it is reasonable to hypothesize that the self/informant correlation may differ depending on which instrument is used to measure personality disorder. We examined whether self/informant concordance varied for self-report questionnaires versus structured interviews, by comparing agreement coefficients between studies using questionnaires and interviews. The median correlation for self-report questionnaires was .38 ($N = 4$), whereas the median correlation for structured interviews was .31 ($N = 7$). (Kappa coefficients were not compared since only one study employing a questionnaire computed values for kappa.) Self/informant agreement appears to be roughly equivalent for questionnaires and interviews, though the small sample of studies available does not allow this issue to be resolved conclusively.

Next, we compared self/informant agreement among the personality disorder instruments. Unfortunately, the number of studies using each personality disorder instrument is too small to make any kind of conclusive determination. Three studies using the SIDP-III reported correlations of .28, .49, and .80. A fourth SIDP-III study reported a kappa of .31. Two studies using the SCID-II reported correlations of .31 and .27. There may be a slight trend for agreement to be higher for studies measuring personality disorder using the SIDP-III versus the SCID-II, though this trend is small and clearly not definitive. The remaining six studies measuring DSM categories using a variety of other instruments yielded median correlations ranging from .18 to .51 and kappas ranging from .06 to .40. Perhaps due primarily to the small number of studies available for this analysis, no clear pattern emerged regarding the relationship between self/informant agreement and different measures of DSM personality disorders.

Self/informant agreement in four studies measuring non-DSM domains of personality pathology, however, yielded somewhat higher levels of agreement (see Table 2). These studies focused on dimensions of personality disorder different from those proposed by DSM. Even though non-DSM domains of personality disorder are less often studied, it is worth considering them since the DSM classification of Axis II pathology may not be the most valid system for classifying personality pathology (Clark et al., 1997; Klonsky, 2000; Westen & Shedler, 1999, 2000; Widiger, 1993).

The instruments used in these studies were the SNAP, IPDE, APFA, and PAS. The SNAP includes 15 factor analytically derived dimensions of personality disorder termed negative temperament, positive temperament, disinhibition, mistrust, manipulativeness, aggression, self-harm, eccentric perceptions, dependency, exhibitionism, entitlement, detachment, impulsivity, propriety, and workaholism (see Ready et al., 1999). The IPDE contains criteria intended to correspond to the ICD-10 personality disorders (see Mann et al., 1999). The APFA was designed to measure personality dysfunction associated with six domains of interpersonal and social role performance (i.e., work, love relationships, friends, nonintimate social contacts, negotiations, and coping) (see Hill, Fudge, Harrington, Pickles, & Rutter, 1995). The PAS was designed to be a schedule for rating personality disorders that assesses 24 personality attributes such as rigidity, aggression, lability, and shyness (see Tyrer, Alexander, Cicchetti, Cohen, & Remington, 1979). The median correlations reported by the studies using the APFA, SNAP, and PAS were .61, .47, and .12, respectively, which can be compared to the median correlation of .36 found in DSM studies. The study using the IPDE yielded a median kappa of .40, in contrast to the median kappa of .14 for DSM studies. The few studies using non-DSM domains of personality disorder produced moderately higher self/informant concordance. However, more studies using non-DSM domains of personality disorder would be needed before we could conclude with any certainty that they produce higher self/informant agreement.

Does Self/Informant Agreement Vary by Type of Sample?

The type of sample may affect the correlation. For example, the self/informant correlation may be different for psychiatric patients than for nonpatient participants. One may speculate that nonclinical subjects have better insight than patients, since patients are more likely to suffer from the lack of self-insight associated with personality pathology. Therefore, nonpatients may be more likely to agree with their informants. It is also possible, however, that by virtue of receiving mental health treatment, patients become more aware of their personality pathology than nonclinical participants and therefore more likely to agree with their informants. To address this question, we compared the median self/informant correlations for studies using mental health patients to those using nonclinical subjects. Results indicated that median self/informant

correlations were comparable for patient samples ($r = .38$, $n = 8$) and nonclinical samples ($r = .40$, $n = 4$). However, given the small number of studies, and the considerable variance in the magnitude of the self/peer correlations reported, it is not possible to determine whether sample type is a factor.

Does Self/Informant Agreement Vary by Type of Informant?

We also considered whether the type of informant could affect the self/informant correlation. For example, perhaps spouses know the subject best and therefore are more likely to agree with the subject. Relatives and friends do not interact with the subject in as wide a variety of contexts and may therefore be less likely to agree with the subject. Unfortunately, most studies allowed the subject to choose the informant, which could be a spouse, sibling, parent, child, nonimmediate relative, or friend, and analyses were not carried out separately by informant type. A thorough comparison of levels of self/informant agreement for different types of informants was therefore not possible. Only two studies allowed for such a comparison. Coolidge, Burns, and Mooney (1995) used two informants per subject, a spouse and a friend. Coolidge et al. computed subject/informant correlations separately for the spouses and friends and found that agreement with the subject was higher for spouses ($r = .51$) than friends ($r = .36$). Modestin and Puhon (2000) compared subject/parent agreement to subject/sibling agreement. They found essentially similar levels of agreement between subjects and parent ($k = .13$) and sibling ($k = .14$) informants. Unfortunately, further comparisons with parents, siblings, second-degree relatives, or nonfriend acquaintances have not been conducted.

Does Self/Informant Agreement Vary by the Age of the Sample?

We also explored whether self/informant concordance differed with the age of the subjects. Mean age ranged from 19 to 43 in the 13 studies examining self/informant agreement among the 10 DSM personality disorders. A Spearman correlation between each sample's mean age and median self/informant correlation yielded an r of .69 ($p = .03$, $N = 10$). This result indicates that self/informant concordance may increase with subject age. Several factors may help to explain this finding. It is possible that older people are less defensive or more honest about their negative qualities, or that traits are more stable in older people and therefore easier for informants to read. It is also possible that older targets and their informants are better

acquainted since they could have had more years knowing each other.

Does Self/Informant Agreement Vary by Sample Size?

Last, we considered whether sample size has any relation to the magnitude of the self/informant correlation. There may be a bias to publish studies with statistically significant correlations, and smaller sample sizes would require higher correlations to be significant. It is therefore possible that median correlations would be higher for studies with smaller sample sizes. Sample size ranged from 42 to 224 in the 13 studies examining self/informant agreement among the 10 DSM personality disorders. We computed a Spearman correlation between each study's sample size and median correlation and found a sizable inverse relationship ($r = -.53, p = .12, N = 10$). This result indicates that in the literature smaller sample sizes are associated with higher self/informant correlations.

Who Reports More Pathology, the Self or the Informant?

For both research and clinical purposes, it would be important to know if either subjects or informants tend to systematically over- or underreport personality pathology. There are reasons to suspect tendencies in each direction for both subjects and informants. Since the subject presumably knows the most information about himself or herself, we may expect the self to be able to report about more personality disorder traits than an informant. Moreover, subjects likely attempt to conceal personality disorder traits from informants, decreasing the number of traits that an informant can observe. On the other hand, the informant may be less likely to deny negative traits that the subject would choose not to acknowledge in an interview or on a self-report measure. An examination of the literature did not settle the issue. Of the twelve studies reporting the magnitude of self versus informant ratings, informants reported more personality pathology in six, subjects' reports of personality disorder were higher in four, and informant- and self-reports were comparable in two studies. It remains inconclusive whether the subjects themselves or informants report more personality pathology.

Is Self/Informant Agreement Higher for Some Personality Disorders Than Others?

In the literature on normal personality traits, there appear to be differences in self/informant agreement depending on the personality trait being investigated. For example,

Kenny (1994) found better agreement between subjects and informants for extraversion than for the other four Big-Five personality traits. We examined whether self/informant concordance differed for different types of personality disorder. Correlation and kappa coefficients for each of the DSM personality disorders are presented in Table 3.

Across studies, median self/informant correlations for individual personality disorders ranged from .29 for narcissistic personality disorder to .56 for antisocial personality disorder. The median correlation was .35 for the Cluster A personality disorders, .45 for the Cluster B personality disorders, and .35 for the Cluster C personality disorders. A similar pattern of results emerged when examining agreement in terms of kappa coefficients. Median kappa coefficients for the Cluster A, Cluster B, and Cluster C personality disorders were .18, .20, and .11, respectively. There did not appear to be substantial differences in self/informant agreement among the personality disorders. In general, self/informant agreement tended to be highest for the Cluster B personality disorders excluding narcissism. Agreement was between low and modest for the Cluster A and C personality disorders. This pattern is consistent with Kenny (1994), in that agreement was found to be highest for the more extraverted forms of personality pathology (i.e., Cluster B traits).

A study examining self/informant agreement on the 15 trait and temperament scales of the SNAP yielded a similar pattern of results (Ready et al., 2000). Agreement was highest for aggression ($r = .62$) and lowest for entitlement ($r = .27$). Though these correlations are higher than those found in studies on DSM categories of personality disorder, the pattern of results found in Ready et al. is consistent with the pattern generated by DSM-oriented studies. Agreement was highest for the conceptually related categories of DSM antisocial personality disorder and SNAP aggression. Agreement was lowest for the conceptually related domains of DSM narcissism and SNAP entitlement.

Do Informants Agree With Each Other?

Though we have discussed several factors that may influence agreement between self- and informant-reports of personality disorder, the fact remains that self/informant concordance is fairly low. One possible explanation is that informants simply do not provide valid information. We tested this hypothesis in part by examining studies that used more than one informant. If informants do not

Table 3. Agreement between self and informant reports for each of the 10 DSM personality disorders

Study	PAR r (κ)	STP r (κ)	SCH r (κ)	ASP r (κ)	BOR r (κ)	NAR r (κ)	HIS r (κ)	AVD r (κ)	DEP r (κ)	OBS r (κ)
Modestin & Puhan (2000)										
Parents	(.05)	(.15)	(-.05)	(.11)	(.24)	(.41)	(.08)	(.41)	(.27)	(.03)
Siblings	(.02)	(.54)	(.39)	(.33)	(.31)	(-.09)	(.06)	(.61)	(.06)	(.15)
Dreessen et al. (1998)	.61	.28			.47			.27	.25	.39
McKeeman & Erickson (1997)	.27	.30	.08	.31	.32	.20	.42	.29	.10	.19
Bernstein et al. (1997)	.36 (.17)	.50 (.43)			.50 (.37)		.45 (.45)	.32 (.22)		.49 (.29)
Ferro & Klein (1997)	.18 (.10)	.17 (-.01)	.34 (.19)	.40 (.28)	.27 (.15)	.14 (.07)	.18 (.04)	.26 (.21)	.15 (.05)	.17 (.11)
Ouimette & Klein (1995)	.59	.47	.29	.13	.42	.35	.21	.39	.28	.19
Coolidge et al. (1995)										
Spouses	.27	.28	.57	.58	.52	.29	.63	.59	.60	.52
Friends	.34	.24	.43	.37	.22	.22	.40	.50	.46	.61
Peselow et al. (1995)	(.26)	(.27)	(.36)	(.36)	(.16)	(.26)	(.35)	(.43)	(.45)	(.31)
Peselow et al. (1994)										
At baseline	.92	.90	.80	.90	.89	.94	.93	.91	.84	.92
At recovery	.77	.73	.57	.81	.80	.80	.81	.66	.54	.72
Riso et al. (1994)	.42 (-.05)	.34 (-)	-.50 (-.01)	.55 (-.04)	.43 (.29)	(-.21 (-.04)	.42 (.15)	.36 (.34)	.23 (-.05)	.19 (-.03)
Dowson (1992)	.23	.38	.61	.75	.59	.44	.38	.35	.46	.32
Zimmerman et al. (1988)	.28	.15	.40	.61	.58	.29	.20	.27	.30	.17
Median correlation	.35	.32	.46	.56	.48	.29	.42	.36	.30	.35
Median kappa	(.10)	(.31)	(.18)	(.25)	(.28)	(.12)	(.15)	(.34)	(.11)	(.11)

Note. PAR = Paranoid Personality Disorder; STP = Schizotypal Personality Disorder; SCH = Schizoid Personality Disorder; ASP = Antisocial Personality Disorder; BOR = Borderline Personality Disorder; NAR = Narcissistic Personality Disorder; HIS = Histrionic Personality Disorder; AVD = Avoidant Personality Disorder; DEP = Dependent Personality Disorder; OBS = Obsessive-Compulsive Personality Disorder.

agree with each other, then it becomes less likely that any given informant offers a valid view of the subject's personality disorder traits. If informants do agree with each other, even if only minimally with the subjects, the possibility remains open that informants have a meaningful perspective to contribute.

Three studies collected data from more than one informant and reported agreement among informants. Coolidge et al. (1995) administered an Axis II inventory to 52 married subjects and then administered a significant-other form about the targets to their spouses and a close friend. Agreement between spouses and friends ($r = .41$) was slightly greater than that between subjects and friends ($r = .36$), but not higher than between subjects and spouses ($r = .51$). Modestin and Puhan (2000) also compared personality disorder data collected from subjects with that obtained from two informants. Subjects were administered the questionnaire version of the SCID-II, and for each subject the same questionnaire was completed by one parent and one sibling. Concordance between parents and siblings ($\kappa = .29$) was more than twice the magnitude of both subject/parent ($\kappa = .13$) and subject/sibling ($\kappa = .14$) agreement. Oltmanns, Turkheimer, and Strauss (1998) examined self/peer agreement in three samples of college student women. They used a peer nomination procedure that allowed participants in each sample to be each other's informants. So, for example, in

their first sample of 41 sorority members, each participant was a subject, as well as an informant for the other 40 members. As a result, there were 40 informants for each participant. Across the three samples, the median self/peer correlation was quite low ($r = .13$). Nevertheless, there was considerable agreement among judges regarding the expression of personality traits in their peers. Across the three groups of participants, median reliability was .81. In general, it appears that subject/informant agreement is lower than that among informants. Thus, even though subjects and informants often disagree on their reports of personality disorder, it is likely that informants have a reliable perspective to offer.

METHODOLOGICAL LIMITATIONS OF THE LITERATURE

The body of literature on self- and informant-reports of personality disorder suffers from several major limitations. First, it is surprising that many studies collecting both self and informant data do not report levels of agreement. Eight studies on personality disorder collect data from both subjects and informants and do not report either correlation or kappa coefficients. Second, of those studies reporting levels of agreement, some report correlations and others report kappa coefficients. Computing correlations is most appropriate when taking a dimensional view of personality disorder, and computing kappa coefficients is appropriate when measuring personality disorder cate-

gorically. A dimensional approach to assessing personality disorder is favored over a categorical approach (Clark et al., 1997; Widiger, 1993). For the purposes of a review paper or meta-analysis, it would be preferable if studies measured personality disorder dimensionally. Third, of those studies that measure personality disorder dimensionally, many use Pearson rather than intraclass correlations. As noted earlier, intraclass correlations are probably more appropriate for measuring self/informant agreement because Pearson correlations do not classify mean differences between raters as error.

Fourth, 26 of the 30 studies gathering information from informants used only one informant. Of the remaining four studies, one used either one or two informants depending on availability, and two used two informants. This is problematic since a single informant can offer only a single perspective. Such a report is not likely to be reliable in either the lay or psychometric sense. Though a one-informant design is more cost-effective and practical, it will be less reliable than a multi-informant design. The fourth study using more than one informant, Oltmanns et al. (1998), illustrates the increase in reliability of several informants over only one.

A fifth limitation is the method of selecting the informant. In 29 of the 30 personality disorder studies using informant data, the informant was selected by the subject and was a friend, significant other, or relative of the subject. This might be described as the “letter of recommendation” problem. The informants who are friends or relatives of the subject, and who are selected by the subject, will most likely view the subject in a positive light. In this sense, the informant’s report becomes more a letter of recommendation than an accurate, objective appraisal of the subject’s personality disorder traits. Particularly because personality disorder (by definition) disrupts interpersonal relationships, an informant selected at random from a group of people who know the subject well is more likely to be frank about the subject’s personality pathology. Granted, researchers may not always be able to select informants at random due to consent issues. When possible, however, this approach may be preferable to having the subject select an informant of his or her choice.

CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

A review of the literature on informant- and self-reports of personality yielded several findings. Self/informant agreement on the DSM personality disorders is modest at

best ($r = .36$, $\kappa = .14$). Agreement is somewhat higher for the few studies examining non-DSM domains of personality pathology ($r = .47$). Self/informant concordance does not appear to vary as a function of the type of Axis-II measure used (e.g., SCID versus SIDP, questionnaire versus structured interview) or of the type of sample examined (e.g., patients versus nonpatients), though more research is needed before any definitive conclusions can be made. Agreement between subjects and informants improves as the subjects’ age increases. Agreement tends to be higher for studies with smaller sample sizes. There is a slight trend in the literature for informants to report more personality pathology than subjects. Self/informant concordance on personality disorder traits is highest for Cluster B pathology (excluding narcissism), lower for Cluster A and Cluster C pathology, and lowest for traits related to narcissism/entitlement. Informant/informant agreement appears to be higher than that between subjects and informants.

Though a lot is now known about the relation between informant and subject reports of personality disorder, several important questions remain. When informant and subject information is discordant, which should be trusted? Future research should compare the validity of self- and informant-reports of personality disorder for various traits and various situations. One such study used a sample of Air Force recruits to compare the predictive validity of peer and self-reports. Oltmanns, Turkheimer, and Fiedler (1999) demonstrated that peer reports of personality disorder might be as good or better than self-reports at predicting early separation from the Air Force. A related unanswered question is, how should one resolve discrepancies between self- and informant-reports? A few researchers have grappled with this question, but no definitive solution has been reached. For example, in one study, “when the patient and informants’ responses conflicted, the rater considered which source gave the clearest examples and most consistent answers and then used his or her own clinical judgment” (Zimmerman et al., 1986, p. 262). Several other studies have adopted similar approaches to the issue, ultimately relying on some form of “clinical judgment” to resolve self/informant discrepancies (e.g., Bernstein et al., 1996, 1997; Klein et al., 1994). Future research may focus on what kinds of traits are more accurately reported by subjects and which by informants. A question not yet addressed in the literature is, does it matter who selects the informant? As mentioned

previously, the studies described in this review asked the subject to choose his or her own informant. Future research should determine whether information given by an informant selected by the subjects differs from that provided by an informant who knows the subject but was not chosen by that person. Last, to what extent is it beneficial to use more than one informant? Oltmanns et al. (1998) demonstrated substantial increases in reliability of peer reports for 40 versus a single rater. Future research should examine how having additional informants increases the quantity, quality, and validity of information about a subject's personality disorder features.

NOTES

1. Our literature yielded only three dissertations. One was found not to have included informant data, one was not available from its host institution, and one reported levels of self/informant agreement for a subset of the personality disorders, and the magnitude of agreement reported was consistent with that reported by published studies. We therefore decided to limit the papers included in this review to those readily available.

2. Allard and Grann (2000) was excluded because more than half of their 42 subjects were psychotic; Dowson (1992b) was excluded because it uses the same sample as Dowson (1992a), which was included in this review; Steketee (1990) was excluded because of a very small sample size ($N = 11$) and because it reported percent agreement rather than kappa or correlation coefficients; Molinari et al. (1998) and Vine and Steingart (1994) were excluded because they computed self/informant agreement at the level of the DSM Clusters rather than for the individual personality disorders.

3. One study (i.e., Oltmanns, Turkheimer, & Strauss, 1998) was not included in Table 1 because it included only three of the DSM personality disorders.

ACKNOWLEDGMENT

Preparation of this article was supported in part by grant MH51187 from the National Institute of Mental Health. We thank Will Grove for his thoughtful comments on an earlier version of this manuscript.

REFERENCES

Allard, K., & Grann, M. (2000). Personality disorders and patient-informant concordance on DIP-Q self-report in a forensic

psychiatric inpatient setting. *Nordic Journal of Psychiatry*, *54*, 195–200.

- Bernstein, D. P., Cohen, P., Skodol, A., Bezirgianian, S., & Brook, J. S. (1996). Childhood antecedents of adolescent personality disorders. *American Journal of Psychiatry*, *153*, 907–913.
- Bernstein, D. P., Kasapis, C., Bergman, A., Weld, E., Mitropoulou, V., Horvath, T., Klar, H. M., Silverman, J., & Siever, L. J. (1997). Assessing Axis II disorders by informant interview. *Journal of Personality Disorders*, *11*, 158–167.
- Brothwell, J., Casey, P. R., & Tyrer, P. (1992). Who gives the most reliable account of a psychiatric patient's personality? *Irish Journal of Psychological Medicine*, *9*, 90–93.
- Clark, L. A. (1993). Schedule for Nonadaptive and Adaptive Personality (SNAP). Minneapolis: Regents of the University of Minnesota.
- Clark, L. A., Livesley, W. J., & Morey, L. (1997). Special feature: Personality disorder assessment: The challenge of construct validity. *Journal of Personality Disorders*, *11*, 205–231.
- Coolidge, F. L., Burns, E. M., & Mooney, J. A. (1995). Reliability of observer ratings in the assessment of personality disorders: A preliminary study. *Journal of Clinical Psychology*, *51*, 22–28.
- Cooney, J. M. (1996). Personality disorder among first ever admissions to an Irish public and private hospital. *Irish Journal of Psychological Medicine*, *13*, 6–8.
- Cooney, J. M., Farren, C. K., & Clare, A. W. (1996). Personality disorder among first ever admissions to an Irish public & private hospital. *Irish Journal of Psychological Medicine*, *13*, 6–8.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1971). *The dependability of behavioral measurements*. New York: Wiley.
- Dowson, J. H. (1992a). Assessment of DSM-III-R personality disorders by self-report questionnaire: The role of informants and a screening test for co-morbid personality disorders (STCPD). *British Journal of Psychiatry*, *161*, 344–352.
- Dowson, J. H. (1992b). DSM-III-R narcissistic personality disorder evaluated by patients' and informants' self-report questionnaires: Relationships with other personality disorders and a sense of entitlement as an indicator of narcissism. *Comprehensive Psychiatry*, *33*, 397–406.
- Dreessen, L., Hildebrand, M., & Arntz, A. (1998). Patient-informant concordance on the Structured Clinical Interview for DSM-III-R personality disorders (SCID-II). *Journal of Personality Disorders*, *12*, 149–161.
- Farmer, R. F. (2000). Issues in the assessment and conceptualization of personality disorders. *Clinical Psychology Review*, *20*, 823–851.
- Ferro, T., & Klein, D. N. (1997). Family history assessment of personality disorders: I. Concordance with direct interview and between pairs of informants. *Journal of Personality Disorders*, *11*, 123–136.

- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1995). The structured clinical interview for DSM-III-R personality disorders (SCID-II): I. Description. *Journal of Personality Disorders, 9*, 83–91.
- Fleiss, J. L. (1973). *Statistical methods for rates and proportions*. New York: Wiley.
- Grove, W. M., & Tellegen, A. (1991). Problems in the classification of personality disorders. *Journal of Personality Disorders, 5*, 31–41.
- Hill, J., Fudge, H., Harrington, R., Pickles, A., & Rutter, M. (1995). The Adult Personality Functioning Assessment (APFA): Factors influencing agreement between subject and informant. *Psychological Medicine, 25*, 263–275.
- Hyder, S. E., Rieder, R. O., Williams, J. B. W., Spitzer, R. L., Hendler, J., & Lyons, M. (1988). The Personality Diagnostic Questionnaire: Development and preliminary results. *Journal of Personality Disorders, 2*, 229–237.
- Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford Press.
- Klein, D. N., Ouimette, P. C., Kelly, H. S., Ferro, T., & Riso, L. P. (1994). Test-retest reliability of team consensus best-estimate diagnoses of Axis I and II disorders in a family study. *American Journal of Psychiatry, 151*, 1043–1047.
- Klonsky, E. D. (2000). The DSM classification of personality disorder: Clinical wisdom or empirical truth? *Journal of Clinical Psychology, 56*, 1615–1621.
- Loranger, A. W., Susman, V. L., Oldham, J. M., & Russakoff, L. M. (1987). The personality disorder examination: A preliminary report. *Journal of Personality Disorders, 1*, 1–13.
- Mann, A. H., Jenkins, R., Cutting, J. C., & Cowen, P. J. (1981). The development and use of a standardized assessment of abnormal personality. *Psychological Medicine, 11*, 839–847.
- Mann, A. H., Raven, P., Pilgrim, J., Khanna, S., Velayudham, A., Suresh, P., Channabasavanna, S. M., Janca, A., & Sartorius, N. (1999). An assessment of the Standardized Assessment of Personality as a screening instrument for the International Personality Disorder Examination: A comparison of informant and patient assessment for personality disorder. *Psychological Medicine, 29*, 985–989.
- McKeeman, J. L., & Erickson, M. T. (1997). Self and informant ratings of SCID-II personality disorder items for nonreferred college women: Effects of item and participant characteristics. *Journal of Clinical Psychology, 53*, 523–533.
- Modestin, J., & Puhan, A. (2000). Comparison of assessment of personality disorder by patients and informants. *Psychopathology, 33*, 265–270.
- Molinari, V., Kunik, M. E., Mulsant, B., & Rifai, A. H. (1998). The relationship between patient, informant, social worker, and consensus diagnoses of personality disorder in elderly depressed inpatients. *American Journal of Geriatric Psychiatry, 6*, 136–144.
- Oltmanns, T. F., Turkheimer, E., & Fiedler, E. R. (1999). *Peer assessment of personality disorders: Predictive validity*. Paper presented at the 6th International Congress on the Disorders of Personality, Geneva, Switzerland.
- Oltmanns, T. F., Turkheimer, E., & Strauss, M. E. (1998). Peer assessment of personality traits and pathology in female college students. *Assessment, 5*, 53–65.
- Ouimette, P. C., & Klein, D. N. (1995). Test-retest stability, mood-state dependence, and informant-subject concordance of the SCID-Axis II questionnaire in a nonclinical sample. *Journal of Personality Disorders, 9*, 105–111.
- Pepper, C. M., Klein, D. N., Anderson, R. L., Riso, L. P., Ouimette, P. D., & Lizardi, H. (1995). DSM-III-R Axis II comorbidity in dysthymia and major depression. *American Journal of Psychiatry, 152*, 239–247.
- Peselow, E. D., Sanfilippo, M. P., & Fieve, R. R. (1994). Patients' and informants' report of personality traits during and after major depression. *Journal of Abnormal Psychology, 103*, 819–824.
- Peselow, E. D., Sanfilippo, M. P., & Fieve, R. R. (1995). Relationship between hypomania and personality disorders before and after successful treatment. *American Journal of Psychiatry, 152*, 232–238.
- Pfohl, B., Blum, N., & Zimmerman, M. (1997). *Structured Interview for DSM-IV Personality (SIDP-IV)*. Washington, DC: American Psychiatric Press.
- Pfohl, B., Corywell, W., Zimmerman, M., & Stangl, D. (1986). DSM-III personality disorders: Diagnostic overlap and internal consistency of individual DSM-III criteria. *Comprehensive Psychiatry, 27*, 21–34.
- Pilgrim, J. A., & Mann, A. H. (1990). Use of the ICD-10 version of the Standardized Assessment of Personality to determine the prevalence of personality disorder in psychiatric in-patients. *Psychological Medicine, 20*, 985–992.
- Pilgrim, J. A., Mellers, J. D., Boothby, H. A., & Mann, A. H. (1993). Inter-rater and temporal reliability of the Standardized Assessment of Personality and the influence of informant characteristics. *Psychological Medicine, 23*, 779–786.
- Ready, R. E., Clark, L. A., Watson, D., & Westerhouse, K. (2000). Self- and peer-reported personality: Agreement, trait ratability, and the “self-based heuristic.” *Journal of Research in Personality, 34*, 208–224.
- Riso, L. P., Klein, D. N., Anderson, R. L., Ouimette, P. C., & Lizardi, H. (1994). Concordance between patients and informants on the Personality Disorder Examination. *American Journal of Psychiatry, 151*, 568–573.
- Skodol, A. E., Oldham, J. M., Rosnick, L., Kellman, H. D., & Hyler, S. E. (1991). Diagnosis of DSM-III-R personality disorders: A comparison of two structured interviews. *International Journal of Methods and Psychiatric Research, 1*, 13–26.
- Steketee, G. (1990). Personality traits and disorders in obsessive-compulsives. *Journal of Anxiety Disorders, 4*, 351–364.

- Tyrer, P., Alexander, M. S., Cicchetti, D., Cohen, M. S., & Remington, M. (1979). Reliability of a schedule for rating personality disorders. *British Journal of Psychiatry*, *135*, 168–174.
- Tyrer, P., Cicchetti, D. V., Casey, P. R., Fitzpatrick, K., Oliver, R., Balter, A., Giller, E., & Harkness, L. (1984). Cross-national reliability study of a schedule for assessing personality disorders. *Journal of Nervous and Mental Disease*, *172*, 718–721.
- Tyrer, P., Seivewright, N., Ferguson, B., Murphy, S., Darling, C., Brothwell, J., Kingdon, D., & Johnson, A. L. (1990). The Nottingham study of neurotic disorder: Relationship between personality status and symptoms. *Psychological Medicine*, *20*, 423–431.
- Vine, R. G., & Steingart, A. B. (1994). Personality disorder in the elderly depressed. *Canadian Journal of Psychiatry*, *39*, 392–398.
- Westen, D., & Shedler, J. (1999). Revising and assessing Axis II, Part II: Toward an empirically based and clinically useful classification of personality disorders. *American Journal of Psychiatry*, *156*, 273–285.
- Westen, D., & Shedler, J. (2000). A prototype matching approach to diagnosing personality disorders: Toward DSM-V. *Journal of Personality Disorders*, *14*, 109–126.
- Widiger, T. A. (1993). The DSM-III-R categorical personality disorder diagnoses: A critique and an alternative. *Psychological Inquiry*, *4*, 75–90.
- Widiger, T. A., & Coker, L. A. (in press). Assessing personality disorders. In J. N. Butcher (Ed.), *Clinical personality assessment* (2nd ed.). New York: Oxford University Press.
- Zimmerman, M. (1994). Diagnosing personality disorders: A review of issues and research methods. *Archives of General Psychiatry*, *51*, 225–245.
- Zimmerman, M., Pfohl, B., Coryell, W., Stangl, D., & Corenthal, C. (1988). Diagnosing personality disorder in depressed patients: A comparison of patient and informant interviews. *Archives of General Psychiatry*, *45*, 733–737.
- Zimmerman, M., Pfohl, B., Stangl, D., & Corenthal, C. (1986). Assessment of DSM-III personality disorders: The importance of interviewing an informant. *Journal of Clinical Psychiatry*, *47*, 261–263.

Received March 20, 2001; revised August 17, 2001; accepted August 30, 2001