AMERICAN PSYCHOLOGICAL ASSOCIATION

Personality Disorders: Theory, Research, and Treatment

© 2022 American Psychological Association
ISSN: 1949-2715

2022, Vol. 13, No. 4, 387–391
https://doi.org/10.1037/per0000584

COMMENTARY

Should the Demonstration of Improved Patient Outcome Be Necessary to Overhaul Diagnostic Approaches?: Comment on Bach and Tracy (2022)

Mark Zimmerman
Department of Psychiatry, Rhode Island Hospital, Providence, Rhode Island, United States
Department of Psychiatry and Human Behavior, Brown University

I will argue in this commentary that it is premature to change the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* approach toward diagnosing personality disorders to the alternative model for personality disorders (AMPD) because such a change in determining personality pathology will be disruptive, and such a disruption is warranted only when it can be demonstrated that patient outcomes are better. I briefly review studies comparing unstructured clinical evaluations with semistructured interviews for symptom disorders, and these studies have found that there is a high rate of missed diagnoses and misdiagnosis by usual clinical assessment. Although on common sense grounds it seems reasonable to assume that greater diagnostic precision will improve outcome, semistructured interviews are not recommended as the standard of care because no studies of adults have yet demonstrated that patient outcomes are better when semistructured interviews are used instead of unstructured clinical interviews. This logic equally applies to overhauling a diagnostic system. In fact, upon considering an approach toward subtyping patients based on outcome in treatment, it is unlikely that a study randomly assigning patients to competing diagnostic or assessment approaches would demonstrate superiority of one method over the other. Finally, I raise concerns about reliance on self-report scales to assess personality and note an absence of studies comparing self-report and interviewer assessments of the AMPD that computed or discussed the positive predictive value of self-report measures of the AMPD system.

Keywords: AMPD, diagnostic assessment, positive predictive value, treatment outcome subtyping, standard of care

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

At the outset I will stipulate that I agree with the points raised by Bach and Tracy (2022). That is, I agree that the alternative model for personality disorders (AMPD) is learnable, can be rated reliably, covers clinically relevant features of personality pathology, and captures the features of the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (*DSM–5*) Section II personality disorders such as borderline and narcissistic personality disorder and helps understand the heterogeneity in their clinical presentation.

However, I will argue in this commentary that it is premature to change the *DSM–5* Section II approach toward diagnosing personality disorders to the AMPD. The change in diagnosing personality disorders will be disruptive, and such a disruption should be adopted only when it can be demonstrated that patient outcomes are better. The promise or potential of improved patient care is insufficient—it must be established—and I do not believe that improved patient outcome will be shown. Moreover, I do not believe that a potential negative effect of the change to the AMPD—reliance on assessment by self-administered questionnaires—has been addressed in the scientific literature. In discussing these concerns, I will draw upon my experience as the principal investigator of the Rhode Island Methods to Improve Diagnostic Assessment and Services project (Zimmerman, 2016), which has examined assessment and nosologic issues for more than 25 years with the goal of improving patient outcomes as well as other literature that has attempted to improve recognition of specific diagnoses by utilizing self-report scales.

Should Semistructured Diagnostic Interviews Be the Standard of Care?

The impetus for initiating the Rhode Island Methods to Improve Diagnostic Assessment and Services project, in which research assessment methods would be incorporated into routine clinical practice, was my experience as a research assistant during which time I administered semistructured diagnostic interviews as part of several studies. As a resident in psychiatry, and then as an attending, my evaluations followed the format of a comprehensive semistructured interview, my evaluations were typically two or more hours in duration, longer than those of my colleagues, and, on average, I made more diagnoses.

Correspondence concerning this article should be addressed to Mark Zimmerman, Department of Psychiatry, Rhode Island Hospital, 593 Eddy Street, RI 02905, United States. Email: mzimmerman@lifespan.org

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Upon becoming director of our hospital's outpatient division, I hired research assistants to administer semistructured interviews to psychiatric outpatients presenting for treatment. Clinical reports were written and provided to the attending psychiatrist before the psychiatrist saw the patient. I considered this a win-win-win proposition—patients received a comprehensive, state-of-the-art, evaluation; clinicians were presented with a six to 10–page report, which they reviewed before seeing the patient; and we researchers developed a database. Our database now exceeds 10,000 patients who have been evaluated with semistructured interviews.

Our clinical-research group was the first to compare disorder prevalence rates in a cohort of patients whose diagnoses were based on unstandardized clinical interviews (i.e., usual clinical practice) and a cohort in which diagnoses were based on semistructured interviews. In one report, we examined whether clinicians underdetected comorbidity compared to research evaluations (Zimmerman & Mattia, 1999b) and found that many more patients were diagnosed with two or more *DSM–IV* symptom (formerly called Axis I) disorders when diagnoses were based on the Structured Clinical Interview for *DSM–IV* than when diagnoses were based on a standard clinical evaluation. Fifteen disorders were more frequently diagnosed when the Structured Clinical Interview for *DSM–IV* was used, and these differences cut across mood, anxiety, eating, somatoform, and impulse control disorder categories. When we examined patients' desire for treatment for the disorders that were not the primary reason for seeking treatment, we found that patients often wanted treatment to address their comorbid conditions (Zimmerman & Mattia, 2000). Thus, detecting diagnostic comorbidity was important from a consumer/patient perspective.

Underdiagnosis was not limited to the symptom disorders. In a separate report, we examined the diagnosis of borderline personality disorder (BPD) and found that diagnostic interviewers using the Structured Interview for *DSM–IV* personality were much more likely to diagnose BPD than clinicians using an unstructured interview (Zimmerman & Mattia, 1999a). Moreover, when the information from the semistructured interview was provided to the clinician then the clinicians diagnosed BPD more frequently; thus, it was not simply a matter of clinicians being reluctant to diagnose BPD during the initial diagnostic evaluation.

Our findings have been replicated and extended several times. Other studies, more narrowly focused on an individual or a limited number of psychiatric disorders such as posttraumatic stress disorder have likewise reported evidence of missed diagnosis in routine clinical practice.

In summary, studies of adults comparing unstructured clinical evaluations with semistructured interviews have found that there is a high rate of missed diagnoses and misdiagnosis by the usual clinical assessment. On common sense grounds, it seems reasonable that greater diagnostic precision will improve outcome. More complete and accurate diagnostic evaluations might impact upon patients' satisfaction with the diagnostic assessment, alliance with the treating clinician, selection of medication, or recommendation for psychotherapy. Patients want treatment for their comorbid conditions. However, no studies of adults have examined whether the use of standardized research interviews improves clinical outcomes.

The same is true for the AMPD. Bach and Tracy's overview highlights the *potential* for improved clinical outcomes. But lacking are studies that have *demonstrated* improved patient outcomes.

Why Improved Outcome Is Unlikely

The *DSM–5* includes several different ways to subtype patients. I will offer a different approach toward subtyping that has implications for whether to change assessment and diagnostic paradigms. This method of subtyping applies when considering a change to the standard of care for clinical evaluations, the use of screening scales (see the following text), or replacing one diagnostic system with another.

The framework toward subtyping is based on outcome in treatment. At its core, the key question when evaluating two competing assessment, diagnostic, or treatment methods is how many and which patients will demonstrate a *differential outcome*? That is, how many (and which) patients would get better in treatment when one type of assessment/diagnostic system/treatment is used as opposed to the other?

In this framework, there are four treatment outcome response subtypes. Group 1 consists of patients who respond to the nonspecific aspects of treatment. In studies of medication, this group can be thought of as placebo responders. The nonspecific aspects of treatment underlying the placebo response include the expectation of benefit, support from a clinician showing concern and attention, and symptom resolution due to the natural history of the disorder. For mood and anxiety disorders, recent meta-analyses have found placebo response rates of 35% to 40% for major depressive disorder (Papakostas et al., 2016), 30% to 40% for bipolar depression (Bartoli et al., 2018; Papakostas et al., 2016), 30% for bipolar mania (Bartoli et al., 2018; Welten et al. 2015), 45% to 50% for panic disorder (Zhang et al., 2020), 40–50% for generalized anxiety disorder (Li et al., 2017), and 35% to 40% for social anxiety disorder (Li et al., 2020). Appreciable placebo response rates have even been found for patients who are considered treatment resistant. Group 1 patients would have a positive response to treatment regardless of assessment method (e.g., unstandardized clinical interview vs. semistructured interview), diagnostic system (e.g., *DSM–5* section II vs. AMPD), or treatment (e.g., cognitive-behavior therapy vs. supportive therapy).

Group 2 includes patients who respond minimally or not at all two or more treatment efforts. For example, 30% of depressed patients have been estimated to be treatment resistant (Voineskos et al., 2020). As with the patients in Group 1, the method used to assess, classify, or treat will not make a difference in the outcome of treatment for the patients in Group 2.

Group 3 includes patients who respond specifically to the active ingredient(s) of an empirically supported procedure/diagnostic system. The response rate in Group 3 is additive to the Group 1 nonspecific treatment response rate (Scott et al., 2022). We currently have at our disposal pharmacological agents and psychotherapies with broad-based efficacy; consequently, assessment method and diagnostic system might not be important. Medications such as selective serotonin reuptake inhibitors have been found to be effective for depression, most anxiety disorders, eating disorders, impulse control disorders, obsessive–compulsive and related disorders, trauma-related disorder, and some somatoform disorders. Likewise, transdiagnostic therapies such as acceptance and commitment therapy and the unified protocol have a broad spectrum of effectiveness. In short, most of the problems for which individuals seek outpatient care have been found to be responsive to at least one of the new generation of medications or

psychotherapies; thus, the assessment method or classification approach are not critical.

Only the patients in Group 4 will exhibit a differential outcome. That is, it is only the patients in this group where the treatment, method of assessment, or classification scheme will make a difference. With respect to the AMPD, some patients in Group 4 will be more likely to respond to treatment when the AMPD approach is used because case formulation is more accurate. On the other hand, some patients in Group 4 will preferentially respond to the *DSM–5* Section II approach, perhaps because a diagnosis is helpful to patients who are reassured in knowing that their problems represent a known diagnostic entity. Thus, of all patients seeking mental health treatment, only those in Group 4 would show a differential outcome, and this group includes an admixture of patients some of whom preferentially respond to one approach and some who preferentially respond to the other.

The size of these four groups is open to debate, but I would hypothesize that no more than 20% of patients seeking behavioral health care belong to Group 4. Thus, for 80% of patients it would not matter how assessment was done or what diagnostic approach was used, and for the minority of 20% for whom it would make a difference, some would respond preferentially to one approach and others to the other approach. Given the likely small size of the differential outcome group, and the inclusion in this group of an admixture of patients who would preferentially respond to one or the other approach, it is unlikely that a study randomly assigning patients to the competing approaches will demonstrate superiority of one approach over the other.

Consequently, although on common sense grounds it seems reasonable to assume that greater diagnostic precision will improve outcome, and multiple studies have demonstrated that semistructured interviews provide a more complete diagnostic profile of patients presenting for treatment, I do not recommend semistructured interviews as the standard of care in assessing adult patients because no studies have yet demonstrated that patient outcomes are better when semistructured interviews are used instead of unstructured clinical interviews. This logic equally applies to overhauling the diagnostic system.

Self-Report Questionnaires, Screening, and Positive Predictive Value

Most of the literature on the AMPD has been based on self-report questionnaires. A PubMed search on January 10, 2022, on the terms AMPD and personality yielded 100 citations over the last 2 years. More than 80% of these studies relied on self-report questionnaires to assess the AMPD constructs.

My bias (perhaps self-serving) is that clinician evaluations are more valid than self-report scale- based assessments. To be sure, self-report scales have value, but I consider them the first stage in a 2-stage evaluation process. That is, indications of pathology or abnormality on self-report scales should not be accepted at face value but should be followed up with a clinician's assessment.

There are many instances in the research literature of the misuse of self-report scales as indicators of a diagnosis. The problem in these studies is the failure to discuss the modest positive predictive value of the scales that are used to identify individuals with the condition of interest.

For example, a recent study reported no difference between patients with and without BPD in their response to electroconvulsive therapy (ECT) (Yip et al., 2021). The authors concluded that their data "provides the clinician with a rational for proceeding with ECT among depressed patients, notwithstanding comorbid BPD." In this study, the researchers "diagnosed" BPD with the McLean Screening Inventory for BPD (MSI-BPD). A second recent study of the relationship between BPD and ECT also used the MSI-BPD to identify BPD and also found no difference in outcome between the patients who did and did not screen positive for BPD (Lee et al., 2019).

In both studies using the MSI-BPD the prevalence of BPD was 21%. A review of nine studies of the MSI-BPD in adults found that at the cutoff of 7 used by the authors of both ECT studies the sensitivity of the MSI-BPD was 81.7% and the specificity of the scale was 63.2% (Zimmerman & Balling, 2021). Based on these values of sensitivity and specificity, and a prevalence of BPD of 21%, the positive predictive value of the scale is only 37.0%. That is, about two thirds of patients in the BPD group would not have been diagnosed with BPD had they been interviewed. Neither study of ECT discussed the limited positive predictive value of the screening test. Thus, conclusions regarding equal efficacy of ECT in depressed patients with and without BPD may be erroneous.

Self-report screening questionnaires have been used in general populations surveys, and researchers have drawn conclusions about disorder prevalence, the psychosocial impairment and public health burden associated with the disorder, and the frequency with which disorders are underrecognized, underdiagnosed, and undertreated. For example, several studies have used the Mood Disorders Questionnaire (MDQ), a self-report screening questionnaire for bipolar disorder, as a case-finding instrument (Zimmerman, 2012) and expressed public health and treatment alarms regarding the underdiagnosis of bipolar disorder. Even as recently as this past year, another large epidemiological study of bipolar disorder, reportedly the first in the United Kingdom, used the Mood Disorders Questionnaire as a case-finding measure and drew inappropriate conclusions about the disorder's recognition and treatment (Zimmerman, 2021).

I have similar concerns about the use of self-report scales in assessing the AMPD. Self-report scales assessing the *DSM* personality disorders have been recognized as "overdiagnosing" personality disorders, and they are best conceptualized as screening measures that are to be followed by a clinical evaluation. The complexity of the AMPD model may lend itself to an overreliance on self-report scales such as the Personality Inventory for *DSM–5* (PID-5) without adequate clinician follow-up. I am unaware of any studies comparing self-report and interviewer assessments of the AMPD that computed or discussed positive predictive value. One might argue that measures of the AMPD reflect the dimensional nature of personality and discussions of positive predictive are inappropriate because they are based on a categorical diagnostic approach. Although PID-5 scores are distributed continuously, they will likely be interpreted categorically—particularly by clinicians. In recognition of how clinicians need context to interpret scores and identify high scorers on the PID-5 (a categorical exercise), Miller and colleagues (Miller et al., 2022) recently presented a table of mean scores plus 1, 1.5 and 2 *SD*s based on a normative sample. This data

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

provides the foundation for categorical utilization of the PID-5 in clinical practice. It is my hope that future research using self-report scales to assess the AMPD discuss the issue of positive predictive value and explicitly recommend that clinicians verify the results of a self-administered questionnaire.

The Research to Be Done

When interviewing psychologists for resident and postdoctoral fellowship positions in the Brown University Psychology Consortium, I routinely ask applicants if they believe that semi-structured interviews should be the standard of care. Because the Brown program tends to recruit research-oriented applicants, almost all applicants have extensive prior experience administering semistructured diagnostic interviews. When asked the standard of care question, most applicants acknowledge the potential benefit of semistructured interviews but raise questions about compensation. When asked to put aside the compensation issue, most support the use of semistructured interviews as the standard of care. (In fact, such evaluations are reimbursed when billed as psychological testing.) Yet, when the residents and fellows join our practice as staff psychologists none employ such interviews as part of their practice. When queried, they defend the adequacy of their clinical assessments. I would not expect otherwise.

What data might encourage clinicians to change their practice? In the interviews of potential residents and fellows my question of whether semistructured interviews should be the standard of care is followed by a question asking the applicant to design a study that would support the use of semistructured interviews as the standard of care. The same question applies to the adoption of the AMPD.

Here is one such, ultrabriefly-described, proposal. Convene a consortium of graduate school programs who advocate the AMPD. Let's assume 10 programs join the consortium. Each program makes outreach to 6 therapists in their community and offers to have graduate students assist the therapist in conducting initial evaluations for the therapist's next 10 patients. At each consortium site, the therapists are randomized to receive a report from the graduate student evaluator which includes information about personality based on either the *DSM–5* section II or AMPD approach. This would result in 600 patients being recruited, with the therapists of 300 receiving a *DSM–5* Section II based description and the therapists of 300 receiving an AMPD based description. Each patient would complete some symptom, functioning, and quality of life assessments at baseline and again at 6 and 12 months. Would the null hypothesis be rejected? Based on the above-described conceptual model of outcome subtyping I would bet no.

Conclusions

The current *DSM–5* Section II approach toward diagnosing personality disorders is not perfect. Others have elaborated the limitations of the Section II categorical approach toward personality disorder diagnosis so I will not repeat them here. In advocating for a more complete overhaul of the current diagnostic nomenclature to a dimensional approach, Lahey et al. (2021) noted that it will be essential to demonstrate that a hierarchical dimensional diagnostic approach improves patient outcomes. If patient outcomes are not demonstrably better, it will be difficult to convince the clinical community that it is worth the effort to learn a new diagnostic language.

An empirically supported system of classifying personality pathology is, of course, highly desirable. But let's not throw out the proverbial baby with the bathwater. While there may be problems with the current *DSM–5* Section II approach, there is also evidence of its validity. Before the clinical community is asked to put forth the time and effort to learn the AMPD, it should be shown data that the change in diagnostic approach will improve the care provided to patients. Specifically, there need to be studies demonstrating that more patients get better. I hope that this will be proved. But doubt it.

References

Bach, B., & Tracy, M. (2022). Clinical utility of the alternative model of personality disorders: A 10th year anniversary review. *Personality Disorder: Theory, Research and Treatment*, *13*(4), 369–379. https://doi.org/10.1037/per0000527

Bartoli, F., Clerici, M., Di Brita, C., Riboldi, I., Crocamo, C., & Carrà, G. (2018). Effect of clinical response to active drugs and placebo on antipsychotics and mood stabilizers relative efficacy for bipolar depression and mania: A meta-regression analysis. *Journal of Psychopharmacology*, *32*(4), 416–422. https://doi.org/10.1177/0269881117749851

Lahey, B. B., Moore, T. M., Kaczkurkin, A. N., & Zald, D. H. (2021). Hierarchical models of psychopathology: Empirical support, implications, and remaining issues. *World Psychiatry*, *20*(1), 57–63. https://doi.org/10.1002/wps.20824

Lee, J. H., Kung, S., Rasmussen, K. G., & Palmer, B. A. (2019). Effectiveness of electroconvulsive therapy in patients with major depressive disorder and comorbid borderline personality disorder. *The Journal of ECT*, *35*(1), 44–47. https://doi.org/10.1097/YCT.0000000000000533

Li, X., Hou, Y., Su, Y., Liu, H., Zhang, B., & Fang, S. (2020). Efficacy and tolerability of paroxetine in adults with social anxiety disorder: A meta-analysis of randomized controlled trials. *Medicine*, *99*(14), Article e19573. https://doi.org/10.1097/MD.0000000000019573

Li, X., Zhu, L., Su, Y., & Fang, S. (2017). Short-term efficacy and tolerability of venlafaxine extended release in adults with generalized anxiety disorder without depression: A meta-analysis. *PLoS ONE*, *12*(10), Article e0185865. https://doi.org/10.1371/journal.pone.0185865

Miller, J. D., Bagby, R. M., Hopwood, C. J., Simms, L. J., & Lynam, D. R. (2022). Normative data for PID-5 domains, facets, and personality disorder composites from a representative sample and comparison to community and clinical samples. *Personality Disorders: Theory, Research, and Treatment*. Advance online publication. https://doi.org/10.1037/per0000548

Papakostas, G. I., Martinson, M. A., Fava, M., & Iovieno, N. (2016). Demographic variables, design characteristics, and effect sizes of randomized, placebo-controlled, monotherapy trials of major depressive disorder and bipolar depression. *The Journal of Clinical Psychiatry*, *77*(05), e619–e624. https://doi.org/10.4088/JCP.14r09767

Scott, A. J., Sharpe, L., Quinn, V., & Colagiuri, B. (2022). Association of single-blind placebo run-in periods with the placebo response in randomized clinical trials of antidepressants: A systematic review and meta-analysis. *JAMA Psychiatry*, *79*(1), 42–49. https://doi.org/10.1001/jamapsychiatry.2021.3204

Voineskos, D., Daskalakis, Z. J., & Blumberger, D. M. (2020). Management of reatment-resistant depression: Challenges and strategies. *Neuropsychiatric Disease and Treatment*, *16*, 221–234. https://doi.org/10.2147/NDT.S198774

Welten, C. C., Koeter, M. W., Wohlfarth, T., Storosum, J. G., van den Brink, W., Gispen-de Wied, C. C., Leufkens, H. G. M., & Denys, D. A.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

(2015). Placebo response in antipsychotic trials of patients with acute mania: Results of an individual patient data meta-analysis. *European Neuropsychopharmacology*, *25*(7), 1018–1026. https://doi.org/10.1016/j.euroneuro.2015.03.010

Yip, A. G., Ressler, K. J., Rodriguez-Villa, F., Siddiqi, S. H., & Seiner, S. J. (2021). Treatment outcomes of electroconvulsive therapy for depressed patients with and without borderline personality disorder: A retrospective cohort study. *The Journal of Clinical Psychiatry*, *82*(2), Article, 19m13202. https://doi.org/10.4088/JCP.19m13202

Zimmerman, M. (2012). Misuse of the Mood Disorders Questionnaire as a case-finding measure and a critique of the concept of using a screening scale for bipolar disorder in psychiatric practice. *Bipolar Disorders*, *14*(2), 127–134. https://doi.org/10.1111/j.1399-5618.2012.00994.x

Zimmerman, M. (2016). A review of 20 years of research on overdiagnosis and underdiagnosis in the Rhode Island Methods to Improve Diagnostic Assessment and Services (MIDAS) project. *Canadian Journal of Psychiatry*, *61*(2), 71–79. https://doi.org/10.1177/0706743715625935

Zimmerman, M. (2021). Using screening scales for bipolar disorder in epidemiologic studies: Lessons not yet learned. *Journal of Affective Disorders*, *292*, 708–713. https://doi.org/10.1016/j.jad.2021.06.009

Zimmerman, M., & Balling, C. (2021). Screening for borderline personality disorder with the McLean Screening Instrument: A review and critique of the literature. *Journal of Personality Disorders*, *35*(2), 288–298. https://doi.org/10.1521/pedi_2019_33_451

Zimmerman, M., & Mattia, J. I. (1999a). Differences between clinical and research practices in diagnosing borderline personality disorder. *The American Journal of Psychiatry*, *156*(10), 1570–1574. https://doi.org/10.1176/ajp.156.10.1570

Zimmerman, M., & Mattia, J. I. (1999b). Psychiatric diagnosis in clinical practice: Is comorbidity being missed? *Comprehensive Psychiatry*, *40*(3), 182–191. https://doi.org/10.1016/S0010-440X(99)90001-9

Zimmerman, M., & Mattia, J. I. (2000). Principal and additional *DSM–IV* disorders for which outpatients seek treatment. *Psychiatric Services*, *51*(10), 1299–1304. https://doi.org/10.1176/appi.ps.51.10.1299

Zhang, B., Wang, C., Cui, L., Gao, J., Wang, C., Tan, X., & Fang, S. (2020). Short-term efficacy and tolerability of paroxetine versus placebo for panic disorder: A meta-analysis of randomized controlled trials. *Frontiers in Pharmacology*, *11*, Article 275. https://doi.org/10.3389/fphar.2020.00275

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.