

An investigation of diagnostic accuracy and confidence related to diagnostic checklists as well as gender biases in relation to mental disorders

Jan C. Cwik^{1*}, Fabienne Papen¹, Jan-Erik Lemke¹, Jürgen Margraf¹

¹Department of Clinical Psychology and Psychotherapy, Ruhr-Universität Bochum, Germany

Submitted to Journal:
Frontiers in Psychology

Specialty Section:
Psychology for Clinical Settings

ISSN:
1664-1078

Article type:
Original Research Article

Received on:
24 Aug 2016

Accepted on:
02 Nov 2016

Provisional PDF published on:
02 Nov 2016

Frontiers website link:
www.frontiersin.org

Citation:
Cwik JC, Papen F, Lemke J and Margraf J (2016) An investigation of diagnostic accuracy and confidence related to diagnostic checklists as well as gender biases in relation to mental disorders. *Front. Psychol.* 7:1813. doi:10.3389/fpsyg.2016.01813

Copyright statement:
© 2016 Cwik, Papen, Lemke and Margraf. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](http://creativecommons.org/licenses/by/4.0/). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Provisional

**An investigation of diagnostic accuracy and confidence related to diagnostic
checklists as well as gender biases in relation to mental disorders**

Jan Christopher Cwik, Fabienne Papen, Jan-Erik Lemke, and Jürgen Margraf

Mental Health Research and Treatment Center, Department of Psychology, Ruhr-
Universität Bochum, Germany

Corresponding author:

Jan C. Cwik

Dept. Clinical Psychology and Psychotherapy

Ruhr-Universität Bochum

Massenbergstraße 9-13

44787 Bochum

Germany

Tel.: +49 / 234 32 21503

Fax: +49 / 234 32 14369

jan.cwik@rub.de

Abstract

This study examines the utility of checklists in attaining more accurate diagnoses in the context of diagnostic decision-making for mental disorders. The study also aimed to replicate results from a meta-analysis indicating that there is no association between patients' gender and misdiagnoses.

To this end, 475 psychotherapists were asked to judge three case vignettes describing patients with Major Depressive Disorder, Generalized Anxiety Disorder, and Borderline Personality Disorder. Therapists were randomly assigned to experimental conditions in a 2 (diagnostic method: with using diagnostic checklists vs. without using diagnostic checklists) x 2 (gender: male vs. female case vignettes) between-subjects design. Multinomial logistic and linear regression analyses were used to examine the association between the usage of diagnostic checklists as well as patients' gender and diagnostic decisions.

The results showed that when checklists were used, fewer incorrect co-morbid diagnoses were made, but clinicians were less likely to diagnose MDD even when the criteria were met. Additionally, checklists improved therapists' confidence with diagnostic decisions, but were not associated with estimations of patients' characteristics. As expected, there were no significant associations between gender and diagnostic decisions.

Keywords: Diagnostic checklist; misdiagnosis; decision-making; gender bias; diagnostic accuracy; diagnostic confidence; therapeutic decisions.

1. Introduction

A comprehensive and systematic diagnostic process based on objective criteria is of particular importance for the treatment of mental disorders (Basco et al. 2000; Ehlert 2007). Nevertheless, several studies have shown high rates of misdiagnoses in daily practice (Bruchmüller and Meyer, 2009; Kales et al., 2005; Wolkenstein et al., 2011). Though patient factors, such as gender, are frequently assumed by clinicians to be a cause of misdiagnosis, in reality, among other factors, clinicians' diagnostic approach is a more prominent cause of misdiagnosis (Wolkenstein et al. 2011; Zimmerman and Mattia 1999; Cwik and Teismann 2016). The diagnostic process may involve subjective descriptions of the patient or may influence or be influenced by therapists' expectations (Langer and Abelson 1974; Rosenhan and Seligman 1989; Margraf and Schneider 2009). Clinicians may also tend toward looser interpretation and use of the diagnostic criteria of classification systems and might resort to other resources instead, such as their professional experiences or personal assumptions (Cwik and Teismann 2016; Bruchmüller, Margraf, and Schneider 2012; Wolkenstein et al. 2011; Bruchmüller and Meyer 2009; Meyer and Meyer 2009; Morey and Ochoa 1989; Garb 2013). As a result, in single diagnoses, clinicians tend to be more likely to make false-positive diagnoses over time, assigning a disorder label though not all required diagnostic criteria are fulfilled (Bruchmüller et al. 2011). Regarding comorbidity, the opposite tends to happen: the most salient diagnosis becomes diagnosed and comorbid disorders are missed (Garb 1998).

Structured diagnostic interviews are recommended in clinical practice to safeguard against misdiagnosis due to clinician bias (Silverman and Ollendick 2005; Schneider and Döpfner 2004; Joiner Jr. et al. 2005; Ehlert 2007). Although interviews

do help to reduce bias and misdiagnosis (Margraf and Schneider 2009), clinicians rarely use such instruments in daily practice (In-Albon et al. 2008; Suppiger et al. 2009; Suppiger et al. 2008; Bruchmüller et al. 2011). Therapists provided varying rationales for not using structured interviews: their clinical judgment is more useful (37%) and the length of the interviews (34%).

Diagnostic checklists provide an alternative to interviews and represent a compromise between clinicians' preference for less structured discussion and the inclusion of specific diagnostic criteria in the diagnostic process, enabling an evaluation process of higher reliability and validity than the status quo (Hiller et al. 1993). Diagnostic checklists are beneficial in that they can assess diagnostic criteria without interfering with the clinician's rapport or interview approach. Thus, open clinical judgment (OCJ) is still possible, but in combination with the assurance that all criteria of a potential disorder are considered systematically and that nonconformity with the criteria of a diagnosis is consciously noted. There is evidence that diagnostic checklists can help to attain reliable diagnoses and to increase diagnostic accuracy (Aebi, Winkler Metzke, and Steinhausen 2010; Biederman et al. 1993; Vaughn and Hoza 2013; Bronisch and Mombour 1994; Mokros et al. 2013).

However, accuracy of given diagnoses is only one essential component for a well-considered psychotherapy. Other important aspects are treatment planning, confidence with given diagnoses, and the estimation of therapy relevant patient characteristics (Arvilommi et al. 2007; Brownstein 2003; Schmidt et al. 2005; Meyer and Meyer 2009; Bruchmüller and Meyer 2009). However, Witteman and Kunst (1997) showed that professionals tend to insist on their first clinical opinion and then search for confirming information. Assuming that different therapists have different

underlying confirmation biases, therapeutic recommendations and estimations of patients' characteristics may vary when therapists do not use diagnostic tools to ensure a standardized consideration of all relevant aspects. In line with this, study results showed that neglecting diagnostic relevant information leads to inappropriate treatment planning (Drake, Alterman, and Rosenberg 1993; North et al. 1997). Furthermore, therapists suggest that the consideration of DSM criteria is not important for treatment planning (Zimmerman et al. 1993) and that their clinical judgment is more useful than diagnostic tools (Bruchmüller et al. 2011). More recently, Zimmerman (2016) pointed out in his systematic review that accurate diagnostics may have an impact on treatment recommendations.

Based on the assumption that the diagnostic process has an effect on treatment (Basco et al. 2000; Ehler 2007) and that an effective diagnostic process could improve treatment planning (Haynes and Williams 2003; Groenier et al. 2008), one could assume that diagnostic checklists – as effective diagnostic tools – will promote accurate diagnostics and adequate treatment recommendations and estimation of patient characteristics compared to open clinical judgment. Consequently, we hypothesized that the usage of diagnostic checklists influences diagnosticians' confidence with given diagnoses and other therapeutic aspects, such as taking patient characteristics or diagnosticians' appraisal of the subsequent psychotherapeutic process into consideration.

In addition to diagnostic methods, gender biases and their association with misdiagnoses of mental disorders became of scientific interest in the 1970s (Broverman et al. 1970). Currently and historically, several diagnoses tend to be

more prevalent in one gender, with a bias toward more prevalence in general in women (American Psychological Association 1975, 1978). Mental disorders with increased prevalence in one gender are of particular interest to psychiatrists and psychotherapists. Major Depressive Disorder (MDD), Generalized Anxiety Disorder (GAD), and Borderline Personality Disorder (BPD) have shown a greater prevalence in women (Seedat et al. 2009; Wittchen and Jacobi 2005; Morschitzky 2009; Angst et al. 2002), with some potential overdiagnosis.

Widiger and Spitzer (1991) postulated that gender-biased mis- or overdiagnosis can occur on two levels: in relation to the application of the diagnostic criteria and/or related to the diagnostic criteria themselves. In line with this, some evidence indicates that a diagnostic category that is, in clinician's opinion, linked to a particular gender, is more likely to be given when the symptomatic behavior of the patient is concordant with traditional gender stereotypes (Sprock 1996; Crosby and Sprock 2004; Flanagan and Blashfield 2005). Clinicians tend to pathologize symptoms when there is a large gap between symptoms and traditional gender characteristics of a patient (Sprock 1996; Möller-Leimkühler 2005). For instance, Möller-Leimkühler (2005) reports that emotional expressions in a depressive man or aggressive behavior in a depressive woman are not in line with traditional gender characteristics and are associated with misdiagnosis of major depressive disorder. According to these findings, Ford and Widiger (1989) illustrated the effects of gender biases in a study on personality disorders. They showed that even if women fulfilled the diagnostic criteria for Antisocial Personality Disorder, clinicians more frequently diagnosed Histrionic Personality Disorder, despite nonfulfillment of the diagnostic criteria.

Assuming that therapists are aware of the connection between gender and prevalence, one could expect consistent gender biases in diagnostic decisions. However, evidence of the influence of gender stereotypes on diagnostic decisions is mixed (Gomes and Abramowitz 1976; Abramowitz et al. 1976; Teri 1982; Hansen and Reekie 1990; Heatherington, Stets, and Mazzarella 1986). For example, therapists made significantly more MDD diagnoses when the described person in a case vignette was a woman in comparison to a man (Bertakis et al. 2001; Lewis et al. 2006; Wrobel 1993). However, other studies reported no effect of patient's gender in MDD (López et al. 1993; Kales et al. 2005; Hansen and Reekie 1990; Case et al. 1999). More generally, Garb (1997) found that women are no more likely to be given a mental health diagnosis than men. Additionally, a recent meta-analysis (Cwik et al., under revision) of ours showed that patient's gender is not a causal factor for misdiagnoses of mental disorders when examining results across 22 studies.

The aim of the present study was to investigate two main questions. First, we wanted to investigate diagnostic accuracy when using diagnostic checklists for mental disorders as compared with OCJ. Additionally, we were interested in finding out to what extent the usage of checklists increases diagnosticians' confidence with given diagnoses. We hypothesized that clinicians' diagnostic accuracy and confidence would be higher when using checklists.

Secondly, this study aimed to investigate the relationship between misdiagnoses and patients' gender. Based on the results of our meta-analysis, one could assume that a diagnostic gender bias is absent when it is investigated in mental disorders in general, but present in specific mental disorders that are more

frequent in females (e.g., MDD, GAD, BPD). Even in these disorders, we expected that patients' gender is not a cause of misdiagnosis or of different estimations of diagnosticians' confidence with given diagnoses.

Finally, we conducted an exploratory analysis of how the usage of checklists and patients' gender is associated with the following diagnostic aspects: clinicians' estimation of the severity of diagnoses, patients' motivation for treatment, expected number of treatment sessions until a significant improvement of symptoms, and clinicians' recommendation for therapeutic interventions.

2. Methods

The Ethics Committee of the Ruhr-Universität Bochum approved the study. All participants gave their informed consent before they were able to start with the questionnaire.

2.1 Participants and Procedure

An invitation to the study was sent out to mailing addresses of psychotherapists that were available by using the search functions of the Chamber of Psychotherapists in all federal states of Germany and of the "Deutsche Psychotherapeutenvereinigung" (German Association of Psychotherapists). In this email, participants received background information about the study and received a link to the online survey. The online survey was completed anonymously and participants could only proceed to the next questionnaire once all questions had been answered. A response of 834 invitees was achieved. Participants were randomly allocated to conditions. The exclusion rate was 359, due to incompleteness of 357 (42.81%) diagnostic surveys and

implausible information of two surveys, thus an effective response rate of $n = 475$ (56.95%) was achieved (checklist (CL): $n = 245$ vs. OCJ: $n = 230$; female vignettes: $n = 241$ vs. male vignettes: $n = 234$). The exclusion of the incomplete surveys ensured that there was no missing data in the remaining 475 surveys. The data collection took place between April and August 2011.

Participants' mean age was 46.67 years ($SD = 11.02$), and 326 participants (68.8%) were women. The sample comprised 79.2% clinicians who had completed cognitive behavioral psychotherapy training, 22.5% who had completed training in analytic psychotherapy, and 24.6% who had completed another type of psychotherapeutic training (including gestalt, client-centered, systemic therapists). Of all psychotherapists, 75.4% had completed a single therapeutic training (e.g., only a training in CBT), 22.9% had completed two trainings (e.g., trainings in CBT and gestalt therapy), and 1.7% had completed three (e.g., trainings in CBT, gestalt therapy, and psychoanalysis) or more trainings. Overall, 82.9% of the participants worked in their own practice, with a mean of 16 years ($SD = 10.13$) of professional experience. When asked to rate their clinical experience in using the ICD-10 (WHO, 1992) and the DSM-IV-TR (APA, 2000) (1 = "no experience" to 5 = "very experienced"), the participants rated their level of experience with ICD-10 higher ($M = 4.37$, $SD = 0.71$) than their experience with DSM-IV-TR ($M = 2.53$, $SD = 1.09$); there was a significant difference between experiences with both classification systems ($t(474) = -32.35$, $p < 0.001$).

2.2 Materials

2.2.1 Case vignettes

Each participant received an online survey with a cover letter and a link to the survey including three case vignettes (available as supplementary material). The case vignettes were constructed on the basis of DSM-IV-TR and were extended to include essential criteria for ICD-10, so that the underlying disorders could be unambiguously diagnosed according to both classification systems.

The first case vignette described a middle-aged patient with a severe episode of MDD. The vignette contained all information necessary to clearly diagnose a MDD episode without psychotic symptoms, according to DSM-IV-TR and ICD-10, except the criterion for suicidal behavior. The second vignette described a patient fulfilling criteria for GAD, based on a DSM-III case description (Spitzer et al. 1991). For this case vignette, we shortened the original description attuned to German culture and to DSM-IV-TR criteria. To ensure that participating therapists who usually use ICD-10 diagnostic criteria in their daily routine would be able to diagnose GAD, we added all required symptoms of ICD-10. The last vignette described a patient fulfilling all general criteria for a personality disorder according to DSM-IV-TR as well as seven required criteria of a BDP. It was based on a case description of Zaudig, Wittchen, and Sass (2000).

To validate the diagnostic criteria of the underlying disorders, seven licensed psychotherapists reviewed all vignettes. All psychotherapists diagnosed the correct disorder without an additional (comorbid) diagnosis in each vignette.

2.2.2 Diagnostic questionnaire

All participants received the same diagnostic questionnaire. In this questionnaire, therapists were asked to diagnose the case vignettes. They were able to choose up

to three options. One option of which was “*no disorder present*”. The other options were 12 listed diagnoses. In addition, therapists of both groups were asked how reliable they rated the selected diagnosis (0=“*insufficient information*” to 100=“*very reliable*”).

Furthermore, therapists were asked to rate the assumed extent of mental, social and job-related impairment caused by the described symptoms (0=“*mentally healthy*” to 100=“*mentally ill*”), patients’ motivation for a therapeutic treatment (0=“*not at all motivated*” to 100=“*highly motivated*”), and severity of the described disorder (1=“*no mental disorder*” to 8=“*severe mental disorder*”), and were asked to estimate how many treatment sessions would be needed until significant improvement could be expected. Finally, therapists were asked which therapeutic orientation for treatment they would advise.

2.2.3 Diagnostic checklists

Checklists were only presented to participants in the checklist condition. Initially, the checklists were not visible to participants. First participants had to choose and click on a diagnosis that determined the checklist they would be presented. For instance, if a participant decided for MDD and clicked on MDD diagnosis, the corresponding MDD checklist opened. Then the participant was able to verify the given diagnosis or to falsify it, switch to another diagnosis and receive the corresponding checklist (e.g., choosing Dysthymia diagnosis and receiving the checklist for Dysthymia).

The checklists listed particular keywords extracted from DSM-IV-TR criteria for each diagnosis. Subsequent to each symptom, participants had to decide whether

the symptom was described or not. Finally, at the end of the checklist, participants had to decide whether the criteria were fulfilled.

2.3 Statistical analysis

Although participants were initially randomized, due to exclusion of incomplete 359 surveys, several between-group significant differences in demographic variables were observed. For the diagnostic procedure condition, significant between-group differences were observed several demographic variables. Significant group differences were observed in: age ($t(473) = -6.80$; $p < 0.001$), professional experience ($t(473) = -5.60$; $p < 0.001$), occupational setting ($t(473) = 4.17$; $p < 0.001$), and therapeutic direction (behavior therapy) ($t(473) = -1.03$; $p = 0.034$). For the gender condition, significant between-group differences were observed in age ($t(473) = -4.86$; $p < 0.001$), professional experience ($t(473) = -5.60$; $p < 0.001$), occupational setting ($t(473) = 4.31$; $p < 0.001$), and clinical practice in usage of ICD-10 ($t(473) = -2.65$; $p = 0.008$). For all other demographic variables, no significant differences between groups were observed. These significant variables were used for the calculation of propensity scores (PS) for each condition: a PS for each participant for gender condition and another PS for diagnostic procedure condition. The PS were calculated using logistic regression with the respective conditions as dependent variables and the demographic variables showing significant differences between groups as independent variables (Austin 2011; Bartak et al. 2009; Weinberger et al. 2009; D'Agostino Jr 1998; Rosenbaum and Rubin 1983).

Next, multinomial logistic regression analyses were performed for each case vignette, with each of the two conditions (diagnostic procedure and patients' gender)

as independent variables and diagnostic decisions and treatment recommendations as dependent variables. Within each condition, the corresponding PS for that condition were used for weighting each participant's contribution, to control for significant group differences.

Additionally, multiple linear regression analyses were used to analyze the nature of correctly diagnosed cases within each case vignette after weighting for differences in demographic variables. The respective conditions (diagnostic procedure and patient's gender) were used as independent variables. Therapists' confidence of diagnosis, patients' motivation for treatment, severity of diagnosis and the expected number of treatment sessions were included as dependent variables. Furthermore, model fit and relative strength of associations between conditions and diagnostic decisions were assessed using Nagelkerke's R^2 and Cohen's f^2 .

All results were Bonferroni-corrected for multiple comparisons. Data analysis was conducted using SPSS version 22.0 for Mac (IBM Corporation 2013).

3. Results

3.1 Diagnostic decisions

3.1.1 Checklists and diagnostic decisions

As can be seen in Table 1, results from the multinomial logistic regression analysis revealed that the usage of OCJ was significantly associated with making more false comorbid diagnoses in patients with MDD ($p < 0.005$), GAD ($p < 0.001$), and BPD ($p < 0.001$), compared to making the correct diagnosis and to therapists using checklists. Furthermore, the usage of OCJ was also significantly associated with making false diagnoses in patients with GAD ($p < 0.001$) and patients with BPD

($p = 0.002$), compared to making the correct diagnosis and to therapists using checklists.

Contrary to our hypothesis that clinicians' diagnostic accuracy would be higher when using checklists, the usage of checklists was significantly associated with therapist's decision to refrain from making a diagnosis in the MDD case, compared to making correct diagnostic decisions and therapists using OCJ ($p < 0.001$). This means that there were more false-negative diagnoses when using checklists compared to using OCJ.

3.1.2 Patients' gender and diagnostic decisions

To examine the association between patients' gender and diagnostic decisions, all three case vignettes were given to participants either in a male or in a female version. As indicated in Table 1, there was no significant association between patients' gender and giving false comorbid diagnoses, false diagnoses or no diagnoses in patients with MDD, GAD or BPD.

- Insert Table 1 about here please -

3.2 Treatment recommendations

3.2.1 Checklists and therapists' treatment recommendations

With respect to treatment recommendations, there was no significant association between the diagnostic procedure and recommendations in the MDD and the GAD cases. Pertaining to the treatment recommendations in the BPD case, therapists who used checklists recommended significantly more frequently dialectic-behavioral

therapy (DBT) as preferable therapy ($p = 0.007$) than psychotherapists in the OCJ condition (see Table 2).

3.2.2 Patients' gender and therapists' treatment recommendations

With respect to treatment recommendations, there was no significant association between the patients' gender and treatment recommendations (see Table 2).

- Insert Table 2 about here please -

3.3 Confidence with given diagnoses and estimations of patients' characteristics

Multiple linear regression analyses were conducted, using both conditions as independent variables and the appraisal of patients' motivation for treatment, severity of given diagnoses, confidence in diagnoses, and psychotherapists' estimation regarding the expected number of treatment sessions until the patient experiences significant improvement as dependent variables. Additionally, corresponding propensity scores were included in the regression models to control for significant differences in demographic variables.

3.3.1 Checklists and therapists' confidence and estimations of patients' characteristics

As can be seen in Table 3, there was a significant association between the usage of checklists and the therapist's confidence in the diagnosis in all case vignettes (MDD: $\beta = -0.161$; $p < 0.001$; GAD: $\beta = -0.206$; $p < 0.001$; BPD: $\beta = -0.205$; $p < 0.001$).

With regard to the estimation of disorders' severity ratings and patients' motivation for treatment as well as the estimation of expected treatment sessions, no significant associations were found.

3.3.2 Patients' gender and therapists' confidence and estimations of patients' characteristics

As Table 3 illustrates, there was a significant association between men with MDD and men with BPD and an appraisal of lower motivation for treatment (MDD: $\beta = -0.142$, $p = 0.003$; BPD: $\beta = -0.159$, $p = 0.001$), but not in case of men with GAD ($\beta = -0.051$, $p = 0.286$). With respect to the confidence with diagnostic decisions, the estimation of the severity of the diagnoses and the number of expected treatment sessions, there were no significant associations observed with patients' gender in any of the three cases.

- Insert Table 3 about here please -

4. Discussion

The present study examined the diagnostic accuracy of psychotherapists with the use of checklists in the diagnostic process and investigated diagnostic accuracy as related to patient gender. We hypothesized that the usage of checklists would result in significantly higher diagnostic accuracy. After controlling for group differences in several therapist demographic variables and therapist training/experience, this hypothesis was mostly supported by the data with one exception: for the MDD case

vignette, when using diagnostic checklists therapists made significantly more false-negative diagnoses compared to when using OCJ.

With regard to the association between the use of checklists and misdiagnoses, psychotherapists who used OCJ were more likely to diagnose an additional comorbid disorder in all of the three disorder cases. In the literature on misdiagnoses, there is evidence that therapists base diagnoses more on comparison to prototypes than on the meeting of diagnostic criteria (Blashfield and Herkov 1996; Westen and Shedler 2000; Crosby and Sprock 2004; Garb 1996). This could explain the difference between the diagnostic decisions of therapists who used checklists and those that did not.

Furthermore, it seems that therapists rely on various heuristics while making a diagnosis, based on their professional experience and, as prior studies also showed, their appraisal of symptoms and diagnoses is not often evident (Morey and Ochoa 1989; Crosby and Sprock 2004; Blashfield and Herkov 1996). In their theory of heuristics and biases, Tversky and Kahneman (1974) point out that even experts are prone to errors of judgment. Concerning the prevalence and the overdiagnosing of Mood Disorders in this study, it seems that psychotherapists seem to be prone to a representativeness bias (Blumenthal-Barby and Krieger 2015; Tversky and Kahneman 1974) with respect to their knowledge of the prevalence and regarding the symptoms of mood and hyperarousal reported in the vignettes. It also seems that the last mentioned aspect triggers a focus on particular symptom complexes that provoke a neglect effect with respect to other relevant symptoms. Furthermore, it seems that the representativeness bias draws diagnosticians' attention to particular symptom complexes, while other relevant symptoms are neglected. Regarding

representativeness bias, it could be surmised that the less frequent the diagnosis, and thus, the less familiar therapists are with the diagnosis, the more they will benefit from using checklists to attain correct and to avoid false diagnoses.

Contrary to our expectations, psychotherapists were more likely to refrain from making a diagnosis in the MDD case vignette when they used checklists. The results of a study by Garb (2007) showed that the usage of computer-administered interviews and checklists was associated with more false-positive diagnoses. Thus, Garb recommends a combination of such a diagnostic instrument with clinical judgment.

More information was collected using the computer-administered interviews and checklists than traditional clinical interviews. Admittedly, in the study of Garb (2007), more symptoms were revealed and thus, it is not surprising that more diagnoses were made. Contrary, in the present study checklists were used to help clinicians making diagnoses from vignettes and thus, this study is involved more with data integration than data collection. Furthermore, although psychotherapists were given checklist information, they were not given the diagnostic criteria for MDD. As Garb (1996) illustrated, if clinicians make diagnoses by comparing clients to prototypes, they may not know the diagnostic criteria for MDD and may have continued to compare the MDD case vignette to their prototype for MDD. Then, if the checklist contained information psychotherapists do not usually collect, that information may not be part of their prototype, and thus the vignette may seem more dissimilar to their prototype resulting in fewer diagnoses of MDD. Thus, one could assume that giving clinicians checklist information does not mean the representativeness heuristic is no longer descriptive. However, it is possible that they

were unable to clarify uncertainties with respect to single symptoms. In the case of doubt, they might have become more conservative in their judgment and decided to refrain from making a diagnosis. To deal with this problem, it could be helpful to provide additional lists with exemplary descriptions. However, Margraf and Schneider (2009) pointed out, while referring to the study of Wittchen and Unland (1991), that checklists do not protect against confirmation bias in the diagnostic process. Therefore, the reliability and validity of diagnoses based on the usage of checklists depends on the clinician's training as well as on the homogeneity of the patients.

Concerning the recommendation of DBT in the BPD case, therapists who used checklists recommended DBT two and a half times as often as therapists who made a diagnosis based on OCJ. It is possible that checklists help therapists to maintain an overview of the full extent of the disorder, whereas therapists relying on OCJ may neglect problem areas or misjudge symptoms. Thus, therapists using checklists tend to consider a broad range of problems, especially cognitive and behavioral deficits, while simultaneously considering BPD specific dysfunctional behaviors like self-injury. Therefore, therapists using checklists were prepared to recommend an adequate therapeutic intervention (APA, 2005).

Regarding the hypothesis that the use of checklists is significantly associated with higher confidence with diagnostic decisions, the results of the study confirm our hypothesis: Therapists who used checklists reported significantly higher confidence with their diagnostic decisions in all three case vignettes compared to psychotherapists who did not use checklists. The results indicate that even a low-threshold diagnostic instrument such as a checklist is associated with increased

diagnostician certainty when compared to OCJ, which is in line with prior results (Vicente et al. 2007). Considering the fact that psychotherapists made significantly more false-positive diagnoses, it is questionable whether higher confidence with diagnostic decisions related to the use of checklists is desirable.

We investigated the association of patients' gender with the accuracy of diagnostic decisions. Based on the results of our meta-analysis, we expected that there would be no significant association between patients' gender and diagnostic decisions in all three case vignettes. In line with our expectations, we generally found no significant association between patients' gender and diagnostic decisions of psychotherapists or treatment recommendations. However, an association was found between therapists' estimations of motivation for treatment in male patients with MDD and BPD. Admittedly, the choice of disorders that were described in the case vignettes could explain this result. Alternatively, in the '70s and '80s, diagnosticians may also have become more aware of gender-related biases and thus tried to counteract them.

However, Widiger and Spitzer (1991) postulated that biases due to patients' gender can occur on two levels: on application of the diagnostic criteria and on the diagnostic criteria themselves. Accordingly, clinicians are more likely to make gender-linked diagnoses if patients' symptomatic behaviors correspond to gender stereotypes (Sprock 1996; Crosby and Sprock 2004; Flanagan and Blashfield 2005). Likewise, experts tend to pathologize when there is a great gap between patients' symptoms and traditional gender characteristics (Sprock 1996; Möller-Leimkühler 2005). A number of stereotypes are reported, e.g., in relation to differences in

symptom expression between men and women (Vredenburg, Krames, and Flett 1986). The case vignettes used in our study did not revert to different these aspects and thus, generalization of these results may be more limited. Future studies should therefore consider these stereotype-related misdiagnoses in their case descriptions.

Finally, it should be mentioned that diagnostic criteria vary between ICD and DSM and that there is no gold standard for making "correct" diagnoses. Thus, simply showing that the usage of checklists can improve diagnosis by using the very criteria that have been used to make the classification of the disorder, is in itself not sufficient to improve the diagnostic process or improve decisions on adequate therapy. As recommended by Garb (2007) a combination of such diagnostic instruments with clinical judgment of a therapist is therefore highly recommended. Accordingly, diagnostic instruments – like diagnostic checklists or interviews – should be seen as helpful diagnostic tools, but they should be clinician administered.

Limitations

A limitation of this study may be the use of case vignettes. Thus, therapists were not able to request additional information during the diagnostic process and might have been limited in their decisions. Future studies should analyze the diagnostic process in detail and inquire as to which information therapists would additionally need for their diagnostic decisions. Studies may also examine why therapists decided not to make a diagnosis (e.g., which criterion was not fulfilled). This could clarify why therapists using checklists more often refrained from making a diagnosis, although a mental disorder was evident. Furthermore, the checklists used were created

specifically for this study. Thus, the results regarding the effects of checklists may be limited to these specific checklists. Finally, future studies should include a greater selection of possible diagnoses and investigate comorbid diagnoses to illustrate more ecologically valid results.

Conclusions

In sum, the present study supports research indicating that diagnostic checklists can improve the accuracy of mental disorder diagnoses, and could be a useful tool to avoid false-positive diagnostic decisions. Nevertheless, the results of our study revealed that the use of checklists could also lead to more false-negative diagnoses when compared with OCJ. Furthermore, the use of checklists provides a higher level of confidence in diagnostic decisions compared with OCJ and is also associated with more correct treatment recommendations.

With regard to the investigation of an association between patients' gender and misdiagnoses, the results of this study revealed no gender bias and no association in relation to most diagnostic decisions.

Acknowledgments

Fabienne Papen and Jan-Erik Lemke carried out their master's thesis in connection with this study. We thank Kristen Lavalley, Simon E. Blackwell, and Helen Copeland-Vollrath for their critical commentaries and for editing the manuscript. We acknowledge support by the RUB international funding program, the German Research Foundation (Deutsche Forschungsgemeinschaft; DFG) and the Open Access Publication Funds of the Ruhr-Universität Bochum, Germany.

Author's disclosure

All authors read and approved the final manuscript and agreed to authorship.

All authors state their compliance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). They also agree to the ethical standards of the Faculty of Psychology's Ethical Commission of the Ruhr-Universität Bochum.

Sources of funding

Jan C. Cwik's position was funded by an Alexander von Humboldt professorship, awarded to Jürgen Margraf. The study was not funded by any granting agency. There has been no prior publication.

Ethic standards

The study was approved by the Ethics Committee of the Faculty of Psychology at the Ruhr-Universität Bochum.

Conflict of interest

All authors declare that they have no conflict of interest and that they are not affiliated or supported by industrial companies.

References

Abramowitz, S. I., H. B. Roback, J. M. Schwartz, A. Yasuna, C. V. Abramowitz, and B. Gomes. 1976. "Sex Bias in Psychotherapy: A Failure to Confirm." *American*

Journal of Psychiatry 133: 706–9.

Aebi, M., C. Winkler Metzke, and H.-C. Steinhausen. 2010. “Accuracy of the DSM-Oriented Attention Problem Scale of the Child Behavior Checklist in Diagnosing Attention-Deficit Hyperactivity Disorder.” *Journal of Attention Disorders* 13: 454–63.

American Psychological Association. 1975. “Report of the Task Force on Sex Bias and Sex-Role Stereotyping in Psychotherapeutic Practice.” *American Psychologist* 30: 1169–75.

— — —. 1978. “Guidelines for Therapy with Women.” *Psychologist* 33: 1122–23.

Angst, J., A. Gamma, M. Gastpar, J.-P. Lépine, J. Mendlewicz, and A. Tylee. 2002. “Gender Differences in Depression - Epidemiological Findings from the European DEPRES I and II Studies.” *European Archives of Psychiatry and Clinical Neuroscience* 252: 201–9.

Arvilommi, P., K. S. Suominen, O. K. Mantere, S. Leppämäki, H. Valtonen, and E. T. Isometsä. 2007. “Adequacy of Treatment Received by Diagnosed and Undiagnosed Patients with Bipolar I and II Disorders.” *Journal of Clinical Psychiatry* 68: 102–110.

Association, American Psychiatric. 2005. *Leitlinien Zur Behandlung Der Borderline Persönlichkeitsstörung*. Bern: Huber.

Austin, Peter C. 2011. “An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies.” *Multivariate Behavioral Research* 46 (3): 399–424. doi:10.1080/00273171.2011.568786.

Bartak, Anna, Marieke D Spreeuwenberg, Helene Andrea, Jan J V Busschbach,

Marcel A Croon, Roel Verheul, Paul M. G. Emmelkamp, and Theo Stijnen. 2009. "The Use of Propensity Score Methods in Psychotherapy Research: A Practical Application." *Psychotherapy and Psychosomatics* 78 (1): 26–34.
doi:10.1159/000162298.

Basco, Monica Ramirez, Jeff Q Bostic, Dona Davies, A. John Rush, Brad Witte, William Hendrickse, and Vicki Barnett. 2000. "Methods to Improve Diagnostics Accuracy in a Community Mental Health Setting." *American Journal of Psychiatry* 157: 1599–1605.

Bertakis, K. D., L. J. Helms, E. J. Callahan, R. Azari, P. Leigh, and Robbinsm J. A. 2001. "Patient Gender Differences in the Diagnosis of Depression in Primary Care." *Journal of Womens Health and Gender-Based Medicine* 10: 689–98.

Biederman, J., S. V. Faraone, A. Doyle, B. K. Lehman, I. Kraus, J. Perrin, and M. T. Tsuang. 1993. "Convergence of the Child Behavior Checklist with Structured Interview-Based Psychiatric Diagnoses of ADHD Children with and without Comorbidity." *Journal of Child Psychology and Psychiatry* 34 (7): 1241–51.
doi:10.1111/j.1469-7610.1993.tb01785.x.

Blashfield, Roger, and M. Herkov. 1996. "Investigating Clinician Adherence to Diagnosis by Criteria: A Replication of Morey and Ochoa (1989)." *Journal of Personality Disorders* 10: 219–28.

Blumenthal-Barby, Jennifer S., and Heather Krieger. 2015. "Cognitive Biases and Heuristics in Medical Decision Making: A Critical Review Using a Systematic Search Strategy." *Medical Decision Making* 35 (4): 539–57.
doi:10.1177/0272989X14547740.

- Bronisch, T., and W. Mombour. 1994. "Comparison of a Diagnostic Checklist with a Structured Interview for the Assessment of DSM-III-R and ICD-10 Personality Disorders." *Psychopathology* 27: 312–20.
- Broverman, I. K., D. M. Broverman, F. E. Clarkson, P. S. Rosenkrantz, and S. R. Vogel. 1970. "Sex-Role Stereotypes and Clinical Judgments of Mental Health." *Journal of Consulting and Clinical Psychology* 34: 1–7.
- Brownstein, Aaron L. 2003. "Biased Predecision Processing." *Psychological Bulletin* 129 (4): 545–68. doi:10.1037/0033-2909.129.4.545.
- Bruchmüller, Katrin, Jürgen Margraf, and Silvia Schneider. 2012. "Is ADHD Diagnosed in Accord with Diagnostic Criteria? Overdiagnosis and Influence of Client Gender on Diagnosis." *Journal of Consulting and Clinical Psychology* 80 (1): 128–38. doi:10.1037/a0026582.
- Bruchmüller, Katrin, Jürgen Margraf, Andrea Suppiger, and Silvia Schneider. 2011. "Popular or Unpopular? Therapists' Use of Structured Interviews and Their Estimation of Patient Acceptance." *Behavior Therapy* 42 (4): 634–43. doi:10.1016/j.beth.2011.02.003.
- Bruchmüller, Katrin, and Thomas D. Meyer. 2009. "Diagnostically Irrelevant Information Can Affect the Likelihood of a Diagnosis of Bipolar Disorder." *Journal of Affective Disorders* 116: 148–51. doi:10.1016/j.jad.2008.11.018.
- Case, S. M., R. Hatala, J. Blake, and G. S. Golden. 1999. "Does Sex Make a Difference? Sometimes It Does and Sometimes It Doesn't." *Academic Medicine* 74: 37–40.
- Crosby, J. P., and J. Sprock. 2004. "Effect of Patient Sex, Clinician Sex, and Sex

Role on the Diagnosis of Antisocial Personality Disorder: Models of Underpathologizing and Overpathologizing Biases.” *Journal of Clinical Psychology* 60 (6): 583–604.

Cwik, Jan Christopher, Eva Charlotte Merten, Marcella Lydia Woud, Lorika Shkreli, Silvia Schneider, and Jürgen Margraf. n.d. “Biased Diagnoses? Meta-Analyses of the Effects of Gender, Ethnicity and Unstructured Assessment.” *Under Revision*.

Cwik, Jan Christopher, and Tobias Teismann. 2016. “Misclassification of Self-Directed Violence.” *Clinical Psychology and Psychotherapy*. doi:10.1002/cpp.2036.

D’Agostino Jr, Ralph B. 1998. “Tutorial in Biostatistics: Propensity Score Methods for Bias Reduction in the Comparison of a Treatment to a Non-Randomized Control Group.” *Statistics in Medicine* 17: 2265–81.

Drake, R. E., A. I. Alterman, and S. R. Rosenberg. 1993. “Detection of Substance Use Disorders in Severely Mentally Ill Patients.” *Community Mental Health Journal* 29: 175–92.

Ehlert, Ulrike. 2007. “Eine Psychotherapie Ist Immer Nur so Gut Wie Ihre Diagnostik.” *Verhaltenstherapie* 17 (2): 81–82. doi:10.1159/000103156.

Flanagan, Elizabeth H., and Roger K. Blashfield. 2005. “Gender Acts as a Context for Interpreting Diagnostic Criteria.” *Journal of Clinical Psychology* 61 (12): 1485–98. doi:10.1002/jclp.20202.

Ford, Maureen R., and Thomas A. Widiger. 1989. “Sex Bias in the Diagnosis of Histrionic and Antisocial Personality Disorders.” *Journal of Consulting and*

Clinical Psychology 57: 301–5.

- Garb, Howard N. 1996. "The Representativeness and Past-Behavior Heuristics in Clinical Judgment." *Professional Psychology: Research and Practice* 27 (3): 272–77. doi:10.1037/0735-7028.27.3.272.
- — —. 1997. "Race Bias, Social Class Bias, and Gender Bias in Clinical Judgment." *Clinical Psychology: Science and Practice* 4 (2): 99–120. doi:10.1111/j.1468-2850.1997.tb00104.x.
- — —. 1998. *Studying the Clinician: Judgment Research and Psychological Assessment*. Washington, DC: American Psychological Association.
- — —. 2007. "Computer-Administered Interviews and Rating Scales." *Psychological Assessment* 19 (1): 4–13.
- — —. 2013. "Cognitive and Social Factors Influencing Clinical Judgment in Psychiatric Practice." *World Psychiatry* 12 (2): 108–10. doi:10.1002/wps.20045.
- Gomes, B., and S. I. Abramowitz. 1976. "Sex-Related Patient and Therapist Effects on Clinical Judgment." *Sex Roles* 2: 1–13.
- Groenier, Marleen, Jules M. Pieters, Casper D. Hulshof, Pascal Wilhelm, and C. L. M. Witteman. 2008. "Psychologists' Judgements of Diagnostic Activities: Deviations from a Theoretical Model." *Clinical Psychology and Psychotherapy* 15 (4): 256–65. doi:10.1002/cpp.587.
- Hansen, F. J., and L.-J. Reekie. 1990. "Sex Differences in Clinical Judgment of Male and Female Therapists." *Sex Roles* 23: 51–64.
- Haynes, S. N., and A. E. Williams. 2003. "Case Formulation and Design of Behavioral Treatment Programs: Matching Treatment Mechanisms to Causal

Variables for Behavior Problems.” *European Journal of Psychological Assessment* 19: 164–174.

Heatherington, L., J. Stets, and S. Mazzarella. 1986. “Whither the Bias: The Female Client’s ‘edge’ in Psychotherapy?” *Psychotherapy: Theory, Research, Practice, Training* 23: 252–56.

Hiller, Wolfgang, Michael Zaudig, Werner Mombour, and Thomas Bronisch. 1993. “Routine Psychiatric Examinations Guided by ICD-10 Diagnostic Checklists (International Diagnostic Checklists).” *European Archives of Psychiatry and Clinical Neuroscience* 242: 218–23.

IBM Corporation. 2013. “SPSS.” Armonk, New York: IBM Corporation.

In-Albon, Tina, Andrea Suppiger, Barbara Schlup, Sascha Wendler, Jürgen Margraf, and Silvia Schneider. 2008. “Validität Des Diagnostischen Interviews Bei Psychischen Störungen (DIPS Für DSM-IV-TR).” *Zeitschrift Für Klinische Psychologie Und Psychotherapie* 37 (1): 33–42. doi:10.1026/1616-3443.37.1.33.

Joiner Jr., Thomas E., Rheeda L. Walker, Jeremy W. Pettit, Marisol Perez, and Kelly C. Cukrowicz. 2005. “Evidence-Based Assessment of Depression in Adults.” *Psychological Assessment* 17 (3): 267–77. doi:10.1037/1040-3590.17.3.267.

Kales, Helen C., Harold W. Neighbors, Frederic C. Blow, Kiran K. K. Taylor, Leah Gillon, Deborah E. Welsh, Susan M. Maixner, and Alan M. Mellow. 2005. “Race, Gender, and Psychiatrists’ Diagnosis and Treatment of Major Depression among Elderly Patients.” *Psychiatric Services* 56 (6): 721–28.

Langer, E. J., and R. P. Abelson. 1974. “A Patient by Any Other Name...: Clinician

Group Difference in Labeling Bias.” *Journal of Consulting and Clinical Psychology* 42: 4–9.

Lewis, R., R. M. Lamdan, D. Wald, and M. Curtis. 2006. “Gender Bias in the Diagnosis of a Geriatric Standardized Patient: A Potential Confounding Variable.” *Academic Psychiatry* 30: 392–96.

López, S. R., A. Smith, B. H. Wolkenstein, and V. Charlin. 1993. “Gender Bias in Clinical Judgment: An Assessment of the Analogue Method’s Transparency and Social Desirability.” *Sex Roles* 28: 35–45.

Margraf, Jürgen, and Silvia Schneider. 2009. “Diagnostik Psychischer Störungen Mit Strukturierten Interviews.” In *Lehrbuch Der Verhaltenstherapie, Band 1: Grundlagen, Diagnostik, Verfahren, Rahmenbedingungen*, edited by Jürgen Margraf and Silvia Schneider, 3rd ed., 344–45. Heidelberg: Springer.

Meyer, Friederike, and Thomas D. Meyer. 2009. “The Misdiagnosis of Bipolar Disorder as a Psychotic Disorder: Some of Its Causes and Their Influence on Therapy.” *Journal of Affective Disorders* 112: 174–83.
doi:10.1016/j.jad.2008.04.022.

Mokros, A., P. Hollerbach, J. Nitschke, R. Eher, and E. Habermeyer. 2013. “Normative Data for the Psychopathy Checklist-Revised in German-Speaking Countries - A Meta-Analysis.” *Criminal Justice and Behavior* 40: 1397–1412.

Möller-Leimkühler, Anne Maria. 2005. “Geschlechterrolle Und Psychische Erkrankung.” *Journal Für Neurologie, Neurochirurgie Und Psychiatrie* 6 (3): 29–35.

Morey, L. C., and E. S. Ochoa. 1989. “An Investigation of Adherence to Diagnostic

Criteria: Clinical Diagnosis of the DSM-III Personality Disorders.” *Journal of Personality Disorders* 3: 180–92.

Morschitzky, H. 2009. *Angststörungen: Diagnostik, Konzepte, Therapie, Selbsthilfe*. 4th ed. Wien: Springer.

North, Carol S., David E. Pollio, Sanna J. Thompson, Daniel A. Ricci, Elizabeth M. Smith, and Edward L. Spitznagel. 1997. “A Comparison of Clinical and Structured Interview Diagnoses in a Homeless Mental Health Clinic.” *Community Mental Health Journal* 33 (6): 531–43.
<http://www.ncbi.nlm.nih.gov/pubmed/9435999>.

Rosenbaum, B Y Paul R, and Donald B Rubin. 1983. “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika* 70 (1): 41–55.

Rosenhan, D. L., and M. E. P. Seligman. 1989. *Abnormal Psychology*. 2nd ed. New York: Norton.

Schmidt, Norman B., Danette Salas, Rebecca Bernert, and Chris Schatschneider. 2005. “Diagnosing Agoraphobia in the Context of Panic Disorder: Examining the Effect of DSM-IV Criteria on Diagnostic Decision-Making.” *Behaviour Research and Therapy* 43 (9): 1219–29. doi:10.1016/j.brat.2004.09.004.

Schneider, Silvia, and Manfred Döpfner. 2004. “Leitlinien Zur Diagnostik Und Psychotherapie von Angst- Und Phobischen Störungen Im Kindes- Und Jugendalter: Ein Evidenzbasierter Diskussionsvorschlag.” *Kindheit Und Entwicklung* 13 (2): 80–96.

Seedat, S., K. M. Scott, M. C. Angermeyer, P. Berglund, E. J. Bromet, T. S. Brugha,

- K. Demyttenaere, et al. 2009. "Cross-National Associations between Gender and Mental Disorders in the WHO World Mental Health Surveys." *Archives of General Psychiatry* 66: 785–95.
- Silverman, W. K., and T. H. Ollendick. 2005. "Evidence-Based Assessment of Anxiety and Its Disorders in Children and Adolescents." *Journal of Clinical Child and Adolescent Psychology* 34: 380–411.
- Spitzer, R. L., M. Gibbon, A. E. Skodol, J. B. Williams, and M. B. First. 1991. "Falldarstellungen Diagnostisches Und Statistisches Manual Psychischer Störungen (DSM-III-R)." In *Falldarstellungen Diagnostisches Und Statistisches Manual Psychischer Störungen (DSM-III-R)*, edited by R. L. Spitzer, M. Gibbon, A. E. Skodol, J. B. Williams, and M. B. First, 257–58. Weinheim: Beltz Verlag.
- Sprock, J. 1996. "Abnormality Ratings of DSM-III-R Personality Disorder Criteria for Males vs. Females." *The Journal of Nervous and Mental Disease* 184: 314–16.
- Suppiger, Andrea, Tina In-Albon, Stephanie Hendriksen, Ernst Hermann, Jürgen Margraf, and Silvia Schneider. 2009. "Acceptance of Structured Diagnostic Interviews for Mental Disorders in Clinical Practice and Research Settings." *Behavior Therapy* 40: 272–79.
- Suppiger, Andrea, Tina In-Albon, Chantal Herren, Klaus Bader, Silvia Schneider, and Jürgen Margraf. 2008. "Reliabilität Des Diagnostischen Interviews Bei Psychischen Störungen (DIPS Für DSM-IV-TR) Unter Klinischen Routinebedingungen." *Verhaltenstherapie* 18 (4): 237–44.
doi:10.1159/000169699.
- Teri, L. 1982. "Effects of Sex and Sex-Role Style on Clinical Judgment." *Sex Roles* 8:

639–49.

- Tversky, Amos, and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185 (4157): 1124–31.
doi:10.1126/science.185.4157.1124.
- Vaughn, A. J., and B. Hoza. 2013. "The Incremental Utility of Behavioral Rating Scales and a Structured Diagnostic Interview in the Assessment of Attention-Deficit/hyperactivity Disorder." *Journal of Emotional and Behavioral Disorders* 21: 227–39.
- Vicente, B., R. Kohn, I. Levav, F. Espejo, S. Saldivia, and N. Sartorius. 2007. "Training Primary Care Physicians in Chile in the Diagnosis and Treatment of Depression." *Journal of Affective Disorders* 98: 121–27.
- Vredenburg, Karel, Lester Krames, and Gordon L. Flett. 1986. "Sex Differences in the Clinical Expression of Depression." *Sex Roles* 14: 37–49.
- Weinberger, Andrea H, Paul K Maciejewski, Sherry A McKee, Erin L Reutenauer, and Carolyn M Mazure. 2009. "Gender Differences in Associations between Lifetime Alcohol, Depression, Panic Disorder, and Posttraumatic Stress Disorder and Tobacco Withdrawal." *The American Journal on Addictions* 18 (2): 140–47.
doi:10.1080/10550490802544888.
- Westen, D., and J. Shedler. 2000. "A Prototype Matching Approach to Diagnosing Personality Disorders: Toward DSM-V." *Journal of Personality Disorders* 14: 109–26.
- Widiger, T. A., and R. L. Spitzer. 1991. "Sex Bias in the Diagnosis of Personality Disorders: Conceptual and Methodological Issues." *Clinical Psychology Review*

11: 1–22.

Wittchen, Hans-Ulrich, and F. Jacobi. 2005. "Size and Burden of Mental Disorders in Europe: A Critical Review and Appraisal of 27 Studies." *European Neuropsychopharmacology* 15: 357–76.

Wittchen, Hans-Ulrich, and H. Unland. 1991. "Neue Ansätze Zur Symptomerfassung Und Diagnosestellung Nach ICD-10 Und DSM-III-R: Strukturierte Und Standardisierte Interviews." *Zeitschrift Für Klinische Psychologie* 20: 321–42.

Witteman, Cilia, and Harald Kunst. 1997. "Planning the Treatment of a Depressed Patient." *Clinical Psychology and Psychotherapy* 4 (3): 157–71.

Wolkenstein, Larissa, Katrin Bruchmüller, Petra Schmid, and Thomas D. Meyer. 2011. "Misdiagnosing Bipolar Disorder - Do Clinicians Show Heuristic Biases?" *Journal of Affective Disorders* 130 (3): 405–12. doi:10.1016/j.jad.2010.10.036.

Wrobel, N. H. 1993. "Effect of Patient Age and Gender on Clinical Decisions." *Professional Psychology: Research and Practice* 24: 206–12.

Zaudig, Michael, Hans-Ulrich Wittchen, and Henning Saß. 2000. *DSM-IV Und ICD-10 Fallbuch: Fallübungen Zur Differentialdiagnose Nach DSM-IV Und ICD-10*. Göttingen: Hogrefe.

Zimmerman, Mark. 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." *The Canadian Journal of Psychiatry* 61 (2): 71–79. doi:10.1177/0706743715625935.

Zimmerman, Mark, V. C. Jampala, F. S. Sierles, and M. A. Taylor. 1993. "DSM-III and DSM-III-R: What Are American Psychiatrists Using and Why?"

Comprehensive Psychiatry 34 (6): 360–74.

Zimmerman, Mark, and Jill I. Mattia. 1999. "Psychiatric Diagnosis in Clinical Practice: Is Comorbidity Being Missed?" *Comprehensive Psychiatry* 40 (3): 182–91.

Provisional

Table 1. Results of three multinomial logistic regression analyses associating usage of checklists and patients' gender with diagnostic decisions (reference category = correct diagnostic decision).

Vignette	Condition	Comparisons			Model fit index			
		comorbid vs. correct	false vs. correct	no vs. correct	$X^2(df)$	p	N	Nagelkerke F^2
MDD	procedure	OR^a [95% CI] 2.62 [1.34 – 5.10]*	OR ^a [95% CI] 1.81 [0.88 – 3.75]	OR ^a [95% CI] 0.05 [0.01 – 0.21]**	$X^2(df)$ 82.73 (12)	p < 0.001	N 475	Nagelkerke F^2 0.19
	gender	1.89 [1.01 – 3.56]	0.90 [0.46 – 1.78]	0.88 [0.42 – 1.83]				
GAD	procedure	OR^a [95% CI] 5.34 [2.68 – 10.64]**	OR^a [95% CI] 5.06 [2.61 – 9.83]**	OR ^a [95% CI] 0.61 [0.25 – 1.42]	$X^2(df)$ 96.39 (12)	p < 0.001	N 475	Nagelkerke F^2 0.21
	gender	1.12 [0.62 – 2.04]	0.70 [0.39 – 1.23]	1.17 [0.51 – 2.67]				
BPD	procedure	OR^a [95% CI] 4.81 [2.83 – 8.17]**	OR^a [95% CI] 5.01 [1.82 – 13.80]**	OR ^a [95% CI] 1.33 [0.73 – 2.45]	$X^2(df)$ 58.86 (12)	p < 0.001	N 475	Nagelkerke F^2 0.13
	gender	1.42 [0.87 – 2.31]	0.56 [0.23 – 1.35]	0.98 [0.53 – 1.78]				

Note: *OR* = odds ratio; 95% *CI* = 95% confidence interval; MDD = Major Depressive Disorder; GAD = Generalized Anxiety Disorder; BPD = Borderline Personality Disorder; correct = correct diagnosis given; comorbid = correct diagnosis given with a false comorbid diagnosis; false = only false diagnosis/diagnoses given; no = no diagnosis/diagnoses given. Significant results are presented in bold letters and were all Bonferroni corrected: * $p < 0.016$; ** $p < 0.003$.

^a Adjusted for significant differences demographic data by using propensity scores.

Table 2. Results of three multinomial logistic regression analyses associating usage of checklists and patients' gender with treatment recommendations (reference category = CBT/DBT).

Vignette	Condition	Comparisons			Model fit index			
		CT vs. CBT/DBT	AT vs. CBT/DBT	OTH vs. CBT/DBT	$X^2(df)$	p	N	Nagelkerke F^2
MDD	procedure	<i>OR</i> ^a [95% <i>CI</i>] 0.87 [0.33 – 2.29]	<i>OR</i> ^a [95% <i>CI</i>] 1.26 [0.72 – 2.19]	<i>OR</i> ^a [95% <i>CI</i>] 0.96 [0.58 – 1.61]	66.66 (12)	< 0.001	475	0.15
	gender	0.83 [0.32 – 2.19]	1.09 [0.63 – 1.89]	1.16 [0.69 – 1.93]				
GAD	procedure	<i>OR</i> ^a [95% <i>CI</i>] 0.56 [0.29 – 1.11]	<i>OR</i> ^a [95% <i>CI</i>] 1.06 [0.58 – 1.92]	<i>OR</i> ^a [95% <i>CI</i>] 1.21 [0.72 – 2.02]	56.84 (12)	< 0.001	475	0.13
	gender	0.80 [0.41 – 1.56]	1.49 [0.81 – 2.72]	1.18 [0.71 – 1.97]				
BPD	procedure	<i>OR</i> ^a [95% <i>CI</i>] 1.11 [0.58 – 2.12]	<i>OR</i> ^a [95% <i>CI</i>] 1.81 [0.70 – 3.55]	<i>OR</i> ^a [95% <i>CI</i>] 2.54 [1.38 – 4.70]**	55.58 (12)	< 0.001	475	0.13
	gender	0.85 [0.44 – 1.63]	1.37 [0.70 – 2.67]	0.83 [0.47 – 1.48]				

Note: *OR* = odds ratio; 95% *CI* = 95% confidence interval; MDD = Major Depressive Disorder; GAD = Generalized Anxiety Disorder; BPD = Borderline Personality Disorder; CBT = cognitive-behavioral therapies; CT = cognitive therapies; AT = analytic therapies; OTH = other therapies; DBT = dialectic-behavioral therapy. Significant results are presented in bold letters and were all Bonferroni corrected: * $p < 0.016$; ** $p < 0.003$.

^a Adjusted for significant differences demographic data by using propensity scores

Table 3. Results of the multiple linear regression models associating usage of checklists and patients' gender with confidence of diagnoses, motivation for treatment, severity of diagnoses and number of expected treatment sessions.

				Model fit index				Effect size
MDD vignette	Condition	β^a	p	R^2	Adjusted R^2	$F(df)$	p	f^2
Confidence of diagnosis (0-100)	procedure	-0.161	0.001*	0.036	0.028	4.44 (4, 470)	0.002	0.037
	gender	-0.006	0.905					
Motivation for treatment (0-100)	procedure	0.101	0.035	0.040	0.032	4.96 (4, 470)	0.001	0.042
	gender	-0.142	0.003*					
Severity of diagnosis (1-8)	procedure	-0.113	0.017	0.056	0.048	7.00 (4, 470)	<0.001	0.059
	gender	0.046	0.333					
Number of sessions	procedure	-0.068	0.154	0.034	0.026	4.10 (4, 470)	0.003	0.035
	gender	0.072	0.131					
<i>GAD vignette</i>								
Confidence of diagnosis (0-100)	procedure	-0.206	<0.001*	0.041	0.033	5.06 (4, 470)	0.001	0.043
	gender	0.021	0.651					
Motivation for treatment (0-100)	procedure	0.029	0.553	0.008	-0.001	0.91 (4, 470)	0.456	0.008
	gender	-0.051	0.286					
Severity of diagnosis (1-8)	procedure	-0.055	0.240	0.074	0.067	9.46 (4, 470)	<0.001	0.080
	gender	0.030	0.525					
Number of sessions	procedure	0.012	0.802	0.019	0.011	2.32 (4, 470)	0.056	0.019
	gender	0.027	0.569					
<i>BPD vignette</i>								
Confidence of diagnosis (0-100)	procedure	-0.205	<0.001*	0.049	0.041	6.08 (4, 470)	<0.001	0.052
	gender	-0.073	0.124					
Motivation for treatment (0-100)	procedure	-0.038	0.422	0.029	0.021	3.52 (4, 470)	0.008	0.030
	gender	-0.159	0.001*					
Severity of diagnosis (1-8)	procedure	-0.030	0.537	0.020	0.012	2.41 (4, 470)	0.048	0.020
	gender	0.085	0.076					
Number of sessions	procedure	-0.056	0.245	0.031	0.023	3.80 (4, 470)	0.005	0.032
	gender	0.107	0.025					

Note: MDD = Major Depressive Disorder; GAD = Generalized Anxiety Disorder; BPD = Borderline Personality Disorder. Significant results are presented in bold letters and were all Bonferroni corrected: * $p < 0.004$.

^a Adjusted for significant differences in demographic data by using propensity scores.