Information order effects in clinical psychological diagnoses

Jan Christopher Cwik | Jürgen Margraf

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

Correspondence
Jan C. Cwik, Department of Clinical Psychology and Psychotherapy, Ruhr-Universität Bochum, Massenbergrstraße 9-13, 44787 Bochum, Germany.
Email: jan.cwik@rub.de

Funding information
Alexander von Humboldt-Stiftung

Despite the wide application and long history of diagnostic systems, several sources of diagnostic errors remain in the criterion-based diagnosing of mental disorders. The aim of this study was to investigate whether the presentational order of diagnosis-relevant information and pretreatment reports predict diagnostic errors. One hundred twenty psychotherapists participated in the present online study. The study employed a 2 (symptom presentation: core symptoms at vignette’s beginning vs. core symptoms at the end of the case vignette) × 2 (pretreatment report: receiving a pretreatment report with an incongruent diagnosis to the case vignette vs. receiving no pretreatment report) between-subjects experimental design, with random assignment. Participants were asked to make diagnoses after reading three case vignettes describing patients with different disorder constellations. Additionally, participants rated their confidence in the diagnoses and their estimation of the severity of each diagnosed condition. Results indicated that order of symptom descriptions predicted the correctness of diagnostic decisions, with a recency effect causing more fully correct diagnostic decisions in cases where diagnostic information was presented last. Receiving incongruent pretreatment reports was predictive for diagnostic errors. In conclusion, the results of this study indicate that diagnoses of mental disorders can depend on the way symptoms are presented or reported.

Key Practitioner Message:
• Therapists’ diagnostic decisions are not influenced by pretreatment reports.
• Diagnostic decisions are affected by information order effects.
• Diagnostic accuracy of psychotherapists is debatable.
• High rate of misdiagnoses in case vignette with comorbid disorders.

KEYWORDS
Decoy Effect, Diagnostic Accuracy, Diagnostic Errors, Pretreatment Report, Recency Effect

1 | INTRODUCTION

An accurate diagnosis is critical in choosing an appropriate course of therapy for mental health patients. Diagnostic classification systems for mental disorders, including the 10th edition of the International Classification of Diseases and Related Health Problems (ICD-10; World Health Organization [WHO], 1992) and the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association [APA], 2013), provide a clear structure for making diagnoses. Diagnosticians assess a core symptom (e.g., the exposure to a traumatic event in posttraumatic stress disorder [PTSD]) and a number of other symptoms as well as estimate the degree of suffering and exclude criteria associated with other disorders. Diagnosticians can and should refrain from assessing further symptoms of a disorder when core symptoms are not present. Comorbid diagnoses can complicate the process. While structured interviews can provide the diagnostician with a reliable and valid tool for systematically making and ruling out diagnoses (In-Albon et al., 2008; Suppiger et al., 2008), they are still rarely used in everyday clinical practice; indeed only in about 15% of clinical interviews (Bruchmüller, Margraf, Suppiger, & Schneider, 2011). Psychotherapists tend to believe their open clinical judgment—that is a diagnostic interview without using a diagnostic tool—is more useful than judgments based on structured diagnostic interviews in making correct diagnoses (Bruchmüller et al., 2011). However, considering that therapists are often confronted with multiple sources of information, as well as influenced by extraneous information (Bruchmüller & Meyer, 2009; Bruchmüller & Schneider, 2012; Meyer & Meyer, 2009; Wolkenstein, Bruchmüller,
Schmid, & Meyer, 2011), open clinical judgment is error-prone and can lead to diagnostic errors. For example, psychotherapists are more likely to diagnose attention-deficit/hyperactivity disorder (ADHD) in a boy than in a girl after reading case vignettes that described several ADHD symptoms but stated that other ADHD criteria were not fulfilled (Bruchmüller, Margraf, & Schneider, 2012). The authors interpret these results as an indication of therapists’ diagnostic decision-making apart from following diagnostic manuals. Other studies indicate that some diagnostic-irrelevant information, such as hallucinations (Meyer & Meyer, 2009) or reduced sleep (Bruchmüller & Meyer, 2009), decreases the possibility that psychotherapists will recognize bipolar disorder. Thus, especially prototypical symptoms are a double-edged sword in the field of diagnostic decision-making related to mental disorders. On the one hand, when more prototypical symptoms are present therapists are more likely to make correct diagnostic decisions (Horowitz, Post, de Sales, Wallis, & Siegelman, 1981). On the other hand, a recent study illustrated that therapists are also more likely to misclassify suicidal behavior as nonsuicidal self-directed violence when a patient with borderline personality disorder (BPD) was described in a case vignette compared to a patient with major depressive disorder (Cwik & Teismann, 2016). Taking these examples into account, structuring diagnostic reasoning by using diagnostic interviews is recommended. In line with this, Mitchell, Vaze, and Rao (2009) illustrated in their meta-analysis that diagnostic concordance between general practitioners’ diagnostic decisions and diagnostic decisions based on structured diagnostic interviews is relatively low with 47.3% of all cases. Our recent meta-analysis also indicated poor consensus between diagnoses based on open clinical judgment versus structured interviews (Cwik et al., 2017).

In order to obtain a better understanding of the reasons for the discordance between diagnostic decisions based on open clinical judgment and structured diagnostic interviews, it could be useful to investigate the processing of psychotherapists’ diagnostic decision-making, such as the order of information collection in daily practice.

Therapists aim to form their own picture of symptoms patients suffer from. Using an open clinical judgment approach, patients report symptoms according to their sense of relevance or in the order that the information comes to their mind. Thus, the order of reporting diagnostically relevant information usually varies between patients and is commonly not in line with the structure of diagnostic manuals. Consequently, psychotherapists using open clinical judgment must mentally refer to and organize the diagnostic criteria in question. Study results suggest an effect of the order of information on the integration of diagnostic information (see for review Hogarth & Einhorn, 1992), and that psychotherapists can be subject to confirmation bias (Margraf & Schneider, 2009; Schulz-Hardt, Frey, Luthgens, & Moscovici, 2000; Shemberg & Doherty, 1999). It is conceivable that the first information presented to the therapist may have a particularly strong influence on the diagnosis, and thus which information is presented first during a diagnostic assessment could have an impact to the final diagnostic outcome. There is evidence that such primary effects occur in diagnostic reasoning (Cunnington, Trunbull, Regehr, Marriott, & Norman, 1997; Rebitschek, Krems, & Jahn, 2015); for instance, the information that there is the swelling of an eyelid has the greatest impact on diagnostic judgments when reported first. Conceivably, this phenomenon could also partly explain misdiagnoses of mental disorders. However, studies also suggest that diagnostic reasoning stepwise in parts counteracts the primacy effect (Rebitschek, Bocklisch, Scholz, Krems, & Jahn, 2015) and can even lead to a recency effect (Hogarth & Einhorn, 1992). Furthermore, research indicates that besides the first also the last information has significant impact on diagnostic judgment (primacy and recency effects; Rebitschek, Bocklisch, et al., 2015). Investigations of Chapman, Bergus, and Elstein (1996) illustrated that a recency effect can also occur in single end-of-sequence probability judgments. Therefore, as both primacy and recency effect have been shown to be related to diagnostic reasoning of mental disorders, further investigations are needed. Thus, an investigation of order effects could add relevant knowledge about the reasons for misdiagnoses. Additionally, a better understanding of potential order effects could help to optimize diagnostic tools and approaches.

Based on our clinical experience in daily practice, patients sometimes seek psychotherapeutic help after reoccurrence of symptoms and some of them provide a pretreatment report from their previous therapist to communicate relevant information. Such information can be diagnostically useful and provide additional information at the beginning of therapy. However, the information could also be based on inaccurate diagnostic decisions. Given that therapists make false diagnostic assumptions near the mark (e.g., physical symptoms and worries in relation to a generalized anxiety disorder [GAD] are misinterpreted as panic attack symptoms and avoidance behavior related to panic disorder [PDI]), one could assume that such nearby misleading information could lead to misdiagnoses causing decoy effects—also known as attraction effect (see Connolly, Reb, & Kausel, 2013). As Kostopoulou, Russo, Keenan, Delaney, and Douiri (2012) showed, physicians are more likely to maintain their diagnostic decisions after having developed an initial diagnostic leaning, even though they might receive additional conflicting information. Similar to decoy effects, therapists trust pretreatment reports and develop an initial diagnostic leaning based on the information of the report. Especially in realistic choice situations, decoy effect has been shown as a relevant explanation for diagnostic errors in decision-making situations (Slaughter, Sinar, & Highhouse, 1999) and particularly in medical decision-making (see for a review Blumenthal-Barby & Krieger, 2015). However, in the case of first given information, a decoy effect appears to be more likely when the judging person has less knowledge in the field (Mantonakis, Rodero, Lesschaeve, & Hastie, 2009). However, to date studies on decoy and order effects related to diagnosing of mental disorders are rare.

Thus, this study aimed to investigate the association between the order in which symptoms are presented and the correctness of diagnostic decisions. We expected that participants who receive a version of case vignettes describing the core symptoms of the disorder at the beginning of the vignette would be significantly more likely to make a correct diagnostic decision (primacy effect). Based on the assumption that a misleading information order (core symptoms at the end of the vignette) give therapists the impression of a more complex case, we expected a lower level of confidence in psychotherapists’ diagnostic decisions and a higher estimation of the severity of diagnosed conditions.
Second, aim of this study was to investigate the effect of incongruent pretreatment reports on diagnostic decisions by psychotherapists upon reading case vignettes of various cases with different mental disorders. Based on the results that clinicians are more likely to maintain their diagnostic decisions after having developed an initial diagnostic leaning (Kostopoulou et al., 2012), we hypothesized that therapists who receive an incongruent pretreatment report develop an initial diagnostic leaning and are biased by it. We expected that therapists with a pretreatment report are significantly more likely to make false diagnostic decisions than participants who did not receive such a report. Additionally, we expected that the receipt of an incongruent pretreatment report creates the impression of a more complex case and thus results in less confidence in psychotherapists' diagnostic decisions and in a higher estimation of the severity of diagnosed conditions.

2 METHODS

2.1 Sample

2.1.1 Response rate

For the recruitment of our sample, we used the official email lists of the psychotherapists' chambers of all federal states of Germany. The email addresses of 1,490 psychotherapists were chosen randomly. Two hundred sixty five participants responded to the invitation (response rate: 17.23%). Participants included 11 psychiatrists as well as 28 psychotherapy trainees and 226 licensed psychotherapists. Both trainees and licensed psychotherapists were clinical psychologist graduates (diploma or Master's degree). It is possible that several of the contacted therapists did not receive our invitation because their email addresses were no longer in use or they were on holiday during the period of the study based on their notices of absence. For the following analyses, we included \( n = 120 \) (45.28%) questionnaires where both the demographic questionnaire and the diagnostic decision questionnaire related to the case vignettes had been completed.

A dropout analysis revealed that participants who dropped out after reading case vignette 1 where significantly more likely to work in a research facility \( \chi^2[1, n = 197] = 5.230, P = .037 \), whereas participants who dropped out after reading case vignette 2 were more likely to have another therapeutic orientation than a psychoanalytic orientation \( \chi^2[1, n = 149] = 4.972, P = .033 \), and participants who dropped out after reading case vignette 1 where significantly less confident with DSM-5 \( \chi^2[127] = -1.194, P = .012 \). Related to all other demographic variables no significant differences were found (see Figure 1).

2.1.2 Participants' characteristics

On average, the age of participants was 45.78 years \( (SD = 12.81) \), 94 (78.3%) were female, and the average number of years of work experience was 14.46 years \( (SD = 10.14) \). Highest academic level data were as follows: 104 participants had a Master of Science or diploma (94.2%) and 6 participants (5.0%) had another level (e.g., state examination). Of all participants, 14 (11.7%) had an academic title (PhD or MD). In relation to therapeutic training, 101 (84.8%) were licensed therapists, 17 (14.2%) were trainees, and 2 (1.7%) gave no answer. Participants partly had multiple therapeutic orientations in cognitive behavioral therapy (70.8%), followed by psychoanalytic therapy (42.5%), Rogerian therapy (11.7%), and systemic therapy (8.3%). Some participants were trained in 2 (21.7%) or 3 (5.8%) therapeutic orientations. Several participants had even further therapeutic trainings. Of all 120 participants, 41 (34.2%) had one, 14 (11.7%) had two trainings, 2 (1.7%) had three, and 2 (1.7%) had four further therapeutic trainings (e.g., Trauma Focused Therapy, Schema Therapy, or Hypnotherapy). The working environment of most therapists was their own private practice \( n = 103; 85.8\% \), followed by inpatient unit \( n = 13; 10.8\% \), research facility \( n = 6; 5.0\% \), day unit \( n = 6; 5.0\% \), and others \( n = 15; 12.5\% \). Ninety-six participants (80.0%) engaged a supervisor, whereof 35 (29.2%) rarely, 52 (43.3%) commonly, and 8 (6.7%) always engaged a supervisor (1 answer missing). In relation to the question whether the therapists decline to treat patients with specialized mental disorders, 59 (49.2%) decline at least one disorder, whereas 25 (20.8%) were specialized for at least one disorder. The familiarity with diagnostic systems was assessed on a scale from 1 = "not familiar" to 4 = "very familiar". Participants were most familiar with the 10th edition of the International Classification of Diseases and Related Health Problems (ICD-10; WHO, 1992) \( M = 3.40; SD = 0.57 \), followed by the 4th edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR; APA, 2000) \( M = 1.97; SD = 0.96 \) and the 5th edition of DSM (DSM-5; APA, 2013) \( M = 1.42; SD = 0.62 \). Finally, participants were asked for an estimation of their diagnostic qualities in relation to other therapists, on a scale ranging from 0 = "worse" to 100 = "superior" \( M = 58.81; SD = 15.62 \).

2.2 Design

Participants were randomized within each of two conditions (pretreatment report condition and order-of-symptom-description-condition). As can be seen in Table 1 and Figure 2, within the pretreatment report condition, participants received either a pretreatment report prior to the three case vignettes \( n = 60 \) or only three case vignettes \( n = 60 \). Within the order-of-symptom-description-condition, participants received either three case vignettes with descriptions of core symptoms at the beginning of the vignettes \( n = 60 \) or at the end of the vignettes \( n = 60 \). Thus, the design resulted in a \( 2 \times 2 \) between-subjects design \( n = 120 \).

2.3 Procedure

All participants received an email request to take part in the online survey with a link to the survey questionnaire. The survey questionnaire was composed with SoSci Survey (Leiner, 2014) and made available to participants on www.soscisurvey.com. After clicking on the link, participants received a short description of the survey questionnaire and the expected amount of time to complete it (35 min). Those who wished to participate received assurance of confidentiality and were asked to give their informed consent. Then the survey automatically started. The procedure of the survey questionnaire can be seen in Figure 2.

As can be seen in Figure 2, participants allocated to the pretreatment report condition first received a pretreatment report. Participants
who were not in this condition received the first case vignette immediately. Participants in the pretreatment report condition could proceed with the survey after reading the report. According to Slaughter, Kausel, and Quiñones (2011), a decoy effect occurs when additional information (e.g., a pretreatment report) lead to changes in choices while the decoy is added to only one of the two possible choice conditions. This is particularly the case when the additional information is highly attractive for the deciding person (Connolly et al., 2013). Thus, adding decoys is an effective method to study decision-making processes (Choplin & Hummel, 2005). On average, participants spent 3.94 min ($SD = 13.87$) on each pretreatment report whereupon the case vignette was presented. Participants in the pretreatment report condition spent $M = 6.69$ min ($SD = 10.28$) on each case vignette, whereas participants in the condition without receiving a pretreatment report

TABLE 1  Between-subjects design ($2 \times 2$) with information about sample size within each condition ($N = 120$)

<table>
<thead>
<tr>
<th>2: Order-of-symptom-description condition</th>
<th>2A Core symptoms at the beginning</th>
<th>2B Core symptoms at the end</th>
<th>Participants in the order-of-symptom-description conditions (2A + 2B: $N = 120$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Pretreatment report condition</td>
<td>1A Pretreatment report</td>
<td>1A-2B ($n = 31$)</td>
<td>$n = 61$</td>
</tr>
<tr>
<td></td>
<td>1B No pretreatment report</td>
<td>1B-2B ($n = 29$)</td>
<td>$n = 59$</td>
</tr>
<tr>
<td>Participants in the pre-treatment report conditions (1A + 1B: $N = 120$)</td>
<td>$n = 60$</td>
<td>$n = 60$</td>
<td></td>
</tr>
</tbody>
</table>

Note. The 120 participants in the two pretreatment report conditions were the same participants as in both order-of-symptom-description conditions.
spent \( M = 7.69 \text{ min} (SD = 14.55) \) on each case vignette. After reading a case vignette, participants stated whether they would give at least one diagnosis for the case. They then automatically proceeded to the diagnostic questionnaire related to the particular case vignette, where they were asked to make at least one diagnosis. Subsequently, the next pre-treatment report and case vignette were automatically presented.

The Ethics Committee of Ruhr-Universität Bochum approved the study.

2.4 | Materials

2.4.1 | Background questionnaire

The demographic section of the assessment asked participants for their gender and age. In the second section, job-related information was assessed, including highest academic degree obtained, academic title, therapeutic approach, professional experience (in years), working environment, familiarity with diagnostic manuals, use of supervision (1 = never to 4 = always), exclusion of patients with specific mental disorders, specialization in specific mental disorders, and number and orientation of advanced therapeutic trainings.

2.4.2 | Pretreatment reports

For the design of the pretreatment reports, we used an original template used at the Mental Health Research and Treatment Center of the Ruhr-Universität Bochum. Pretreatment reports included information about patient’s treatment, the length of therapy, time frame, number of therapeutic sessions, explicit naming of previously given diagnosis/diagnoses with the corresponding ICD-10 code, description of the complaints at the beginning of the therapy, and description of the therapeutic process, including information about therapy content and the used therapeutic rationale (e.g., Dialectical behavior therapy (DBT); see Figure 3).

All reports were incongruent with each of the following case vignettes. The assignment of pretreatment reports with corresponding case vignettes was the same for each participant. Two pretreatment reports provided information that matched the symptom described in the vignettes, but reported discordant diagnoses. The first one was the pretreatment report presented prior to the GAD case vignette. Here, the somatic symptoms and worries of a patient with GAD were affiliated to PD and hypochondriasis in the pretreatment report. The second pretreatment report presented prior to the PD case vignette, provided information that matched the symptom described in the vignette, but were incongruent with the case vignette. In this report, somatic symptoms of a patient with PD were affiliated to GAD. These combinations were chosen because each of these diagnoses should be differentially diagnostically clarified according to DSM (see APA, 2013, pp. 213 and 314). Thus, the misdiagnoses named in the pretreatment report seemed to be misleading. Nevertheless, it also does not seem unlikely that in daily practice these disorders are incorrectly diagnosed related to the symptoms reported in the case vignette. With respect to the BPD/PTSD case, we presented a pretreatment report that only reported BPD as given diagnosis. Doing so, we expected that such misleading pretreatment reports affect therapists to neglect PTSD symptomatology.
2.4.3 | Case vignettes

Contrary to other studies that use case vignettes for the investigation of diagnostic accuracy (e.g., Bruchmüller et al., 2012; Meyer & Meyer, 2009), we decided to send three case vignettes with descriptions of three different mental disorders. The main reason for this procedure was that we aimed to investigate decoy and information order effects more generally in relation to different disorders and comorbidity. Additionally, to investigate whether these effects occur consequently, the usage of more than one vignette was necessary.

The first vignette described a patient fulfilling GAD, based on a DSM-III case description (Spitzer, Gibbon, Skodol, Williams, & First, 1991). For this case vignette we shortened the original description, attuned to German culture and to DSM-5 criteria. To ensure that participating therapists who use ICD-10 diagnostic criteria in their daily routine would be able to diagnose GAD, we additionally included all required symptoms based on ICD-10. The second vignette described a patient fulfilling all general criteria for a personality disorder according to DSM-5 as well as seven required criteria for BPD. The case vignette was based on a BPD case description of Zaudig, Wittchen,
and Sass (2000). Additionally, all required symptoms of PTSD according to DSM-5 and ICD-10 were added. We decided to include the criteria of PTSD to investigate effects that cause misdiagnoses in comorbidity considering that this could be of high relevance for daily practice. These two disorders (BPD and PTSD) were combined due to their high rates of comorbidity (Grant et al., 2008; Pabst et al., 2012). Notably, an integration of the disorders has been discussed in the field (Eichelman, 2010; Hodges, 2003). Studies have shown that the especially complex symptomatology blur boundaries between the two disorders (Cloitre, Garvet, Weiss, Carlson, & Bryant, 2014). The third case vignette was a patient fulfilling all symptoms of a PD according to DSM-5 and ICD-10. This case vignette was created for the current study by the authors.

To validate the diagnostic criteria of the underlying disorders, seven licensed psychotherapists of the Mental Health Research and Treatment Center of the Ruhr-Universität Bochum reviewed the vignettes. All psychotherapists diagnosed the disorders in each vignette correctly without an additional (comorbid) diagnosis.

As an example, we present the GAD case vignette (according to Spitzer et al., 1991) in the following. The underlined part of the vignette was either at the beginning of the case vignette (as in this example) or at the end of the case vignette (after the last passage):

"Mrs. L., a 44-year-old married administrative assistant contacts you to receive outpatient psychotherapeutic treatment. She tells you that she has completed psychotherapy to address this problem some time ago. She complains about dizziness, sweaty hands, palpitations, and ringing in her ears. At night Mrs. L. stares at the ceiling and ruminates. She is always worried about the health of her parents. Her father had a myocardial infarction 2 years ago, but feels healthy today. She also worries about being a "good mother," whether her husband could leave her (there was no indication that she is dissatisfied with the marriage), and whether her colleagues accept her. Despite her better judgment that there are no problems concerning these worries, she is not able to suppress these thoughts and the inner tension. Mrs. L. also experienced that her mouth and her throat were dry, and that she was periodically uncontrollably shaking as well as having a persistent feeling of "being always on the go." This influenced her ability to concentrate. These feelings have continuously occurred for the last 2 years and were not restricted to limited periods of time. Because of these problems, Mrs. L. contacted her family doctor, a neurologist, a neurosurgeon, a chiropractor, and an ENT specialist. She was put on a hypoglycemic diet, received physical therapy because of a pinched nerve, and she was informed that she has got a "problem with her inner ear." Mrs. L. has had only limited social contacts in recent years because of her nervous symptoms. Although she had to flee from work several times because of her symptoms, she still works at the road traffic office where she has been working since the beginning of her career. She tends to conceal her symptoms from her husband and the children. For them she wants to appear "perfect." She reports that her nervousness problems are not related to her family."

2.4.4 | Diagnostic survey

After reading a vignette, participants were asked to give a diagnostic assessment of the vignette. Participants were instructed to use their clinical judgment without reference to auxiliary materials (e.g., DSM or ICD). For their diagnostic appraisal, they could choose up to three diagnoses out of 19 listed diagnoses and were asked to indicate whether each selected disorder was present with clinical or subclinical intensity. The diagnoses used in the present study were derived from another study testing a similar research question (see Cwik, Papen, Lemke, & Margraf, 2016). As that previous study only used 12 diagnoses, seven additional diagnoses were added. Thus, there were a total of 19 diagnoses, and all of them represented high prevalence rates according to the DSM-5. Each participant received the list of disorders in the same order after each vignette. They were also asked to indicate their confidence in each diagnosis chosen, using a scale ranging from 0 = "absolutely no confidence" to 100 = "complete confidence". This type of confidence rating has already been used in former studies (e.g., Meyer & Meyer, 2009; Schmidt, Salas, Bemert, & Schachtschneider, 2005). Finally, they were asked for their assessment of severity of diagnosed conditions for each disorder with clinical intensity.

For the order-of-symptom-description between-subjects manipulation, the core symptoms (criterion A in DSM-IV) were described either at the beginning or the end of the vignette. In the GAD case vignette, participants either received the criterion "expressive anxiety and worry" at the beginning or at the end of the vignette. The vignette of the BPD and comorbid PTSD was varied so that participants received a version with either the BPD criterion at the beginning of the vignette, followed by the PTSD criterion naming the concrete traumatic event, or a version with the PTSD criterion at the beginning, starting with the traumatic event and followed by the BPD description. In the PD vignette, participants received either a version describing the criterion "recurrent unexpected panic attacks" at the very beginning or at the end of the vignette. At the end, participants were debriefed on the study, receiving information about the aims of the study and the experimental manipulation.

2.5 | Statistical analysis

First, we examined whether participants in each condition across the two groups (pretreatment report vs. no pretreatment report; core symptom at the beginning of the vignette vs. core symptom at the end of the vignette) varied in demographic variables. To control for significant differences and trends (\(\text{P} < .10\)) in demographic variables, propensity scores were calculated (Austin, 2011; Bartak et al., 2009; Rosenbaum & Rubin, 1983). All variables with significant differences and trends between groups were used for the calculation of propensity scores for each condition, thus each participant had one propensity score for the information order condition and another propensity score for the pretreatment condition. The propensity scores 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; D’Agostino, 1998; Rosenbaum & Rubin, 1983; Weinberger, Maciejewski, McKee, Reutenauer, & Mazure, 2009). The propensity scores were then included in the corresponding regression analyses for controlling significant group differences.

For the investigation of the diagnostic decisions of participants, we first rated whether participants gave the so-called "aimed diagnosis." This was the case when participants made the GAD diagnosis in the GAD vignette, the BPD and PTSD diagnoses in the BPD and PTSD vignette, and the PD diagnosis in the PD vignette. Next, we investigated the rates of participants who made the aimed and at least one other “false comorbid diagnosis” (e.g., GAD diagnosis and a major depressive disorder diagnosis in the GAD vignette). Last, we investigated whether participants made a "fully correct diagnostic decision" (e.g., GAD diagnosis and no other diagnosis in the GAD vignette).

In order to examine whether the order of information about symptoms and the presentation of a pretreatment report is associated with diagnostic decisions, we conducted multiple binary logistic regression analyses with only fully correct diagnostic decisions as dependent variables (yes vs. no). To examine the association between conditions and diagnostic confidence and severity of diagnosed conditions, linear regression analyses were conducted. Subsequently, results of all regression analyses were Bonferroni-corrected for multiple comparisons. We chose logistic and linear regression analyses as analyzing methods, which enabled us to control for group differences by including propensity scores and each noninteresting condition. Data analyses were conducted using SPSS version 22.0 for Mac (IBM Corporation, 2013).

3 | RESULTS

First, we examined whether participants in each condition (across the two groups) varied significantly with respect to differences in demographic variables. In the case of the pre-treatment report condition, work environment (research unit) varied between groups: There was a trend toward a higher proportion of researchers in the group that did not receive the pre-treatment report ($X^2[1, N = 120] = 2.95, P = .086$). Regarding the information order condition, participants who received all vignettes describing the core symptoms at the beginning were significantly more often men ($X^2[1, N = 120] = 7.070, P = .008$) and were significantly less familiar with ICD-10 ($U(60,60) = 1285, P = .002$). In addition, there was a trend that these participants had earned a PhD ($X^2[1, N = 120] = 2.911, P = .088$), were working in a research unit ($X^2[1, N = 120] = 2.807, P = .094$), had fewer further therapeutic trainings ($t(120) = -1.805, P = .074$), and were less likely to be specialized in treating trauma-related disorders ($X^2[1, N = 120] = 2.727, P = .099$). Thus, we used these variables for the calculation of propensity scores. Furthermore, regarding the “information-order condition,” an additional propensity score was calculated, solely based on variables revealing significant group differences for further analyses. In all other variables, no significant differences were found between participants.

At least one false diagnosis without giving the aimed diagnosis was given by 62 (51.7%) participants in the GAD case vignette, and by 75 (62.5%) in the BPD and PTSD case vignette, and by 50 (41.7%) in the PD case vignette. Across all conditions, participants made the aimed diagnosis in each case vignette in fewer than half of the cases: GAD was diagnosed by 58 (48.4%) participants in the GAD case vignette, BPD and PTSD were both diagnosed by 45 (37.5%) participants in the BPD and PTSD case vignette, and PD was diagnosed by 70 (58.3%) participants in the PD case vignette. The number of psychotherapists who made the aimed and at least one false comorbid diagnosis was relatively low: Of all participants, 23 (19.2%) made a false comorbid diagnosis in the GAD case vignette, 45 (37.5%) in the BPD and PTSD case vignette, and 28 (23.3%) in the PD case vignette. Finally, 35 (29.2%) participants made a fully correct diagnostic decision in the GAD case, none of the participants made a fully correct diagnostic decision in the BPD and PTSD case vignette, and 42 (35.0%) participants made a fully correct diagnostic decision in the PD case vignette.

In relation to the confidence with given diagnoses, participants were least confident in the GAD case vignette ($n = 116; M = 60.60, SD = 23.58$), followed by the PD case vignette ($n = 116; M = 72.09, SD = 22.55$), and the BPD and PTSD case vignette ($n = 117; M = 74.94, SD = 21.66$). Analyses of the severity of the symptoms at each indicated that participants rated the lowest severity in the PD case vignette ($n = 116; M = 67.30, SD = 15.51$), followed by the GAD case vignette ($n = 113; M = 70.78, SD = 13.08$), and the BPD and PTSD case vignette ($n = 117; M = 89.34, SD = 8.43$).

Additionally, regarding the confidence with given diagnoses and the estimation of severity of described diagnoses, there was no significant association between receiving a pretreatment report and a higher expectation of the severity of the described disorders.

As expected, the groups did also not differ in the level of confidence with the given diagnoses and the expectation of the severity of described diagnoses (see Table 3).

As can be seen in Table 4, significant associations were observed regarding the order of symptom descriptions. Psychotherapists who received a GAD case vignette with the description of the GAD core symptoms at the end of the vignette detected GAD significantly more often than those receiving a vignette describing the core symptoms at the beginning (GAD: odds ratio $OR = 2.89, P = .017$), and thus, they made significantly more fully correct diagnostic decisions. In the PD case, participants who received a vignette with the core symptoms at the end also made significantly more fully correct diagnostic decisions (PD: $OR = 2.65, P = .024$). Both results would remain significant, even if the calculated propensity score would have been calculated solely based on variables that significantly differed ($P < .05$) between groups (GAD: $OR = 2.61$, $P = .025$; PD: $OR = 2.44$, $P = .034$). Admittedly, in that case, the results would not persist Bonferroni-corrections for multiple comparisons. In the BPD and PTSD case, all participants made a false diagnosis. Thus, it was not possible to investigate whether there was an association between the order of symptom descriptions and false diagnoses or the correctness of diagnostic decisions.
TABLE 2  Frequency of diagnostic decisions for the pretreatment report and order-of-symptom-description conditions (N = 120)

<table>
<thead>
<tr>
<th>Aimed diagnosis/diagnoses</th>
<th>Pretreatment report condition N (%)</th>
<th>Order-of-symptom-description condition N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>With report</td>
<td>No report</td>
</tr>
<tr>
<td>GAD</td>
<td>33 (54.1)</td>
<td>25 (42.4)</td>
</tr>
<tr>
<td>BPD/PTSD</td>
<td>24 (39.3)</td>
<td>21 (35.6)</td>
</tr>
<tr>
<td>PD</td>
<td>37 (60.7)</td>
<td>33 (55.9)</td>
</tr>
<tr>
<td>False comorbid diagnosis/diagnoses</td>
<td>GAD</td>
<td>13 (21.3)</td>
</tr>
<tr>
<td></td>
<td>24 (39.3)</td>
<td>21 (35.6)</td>
</tr>
<tr>
<td></td>
<td>15 (24.6)</td>
<td>13 (22.0)</td>
</tr>
<tr>
<td>Fully correct diagnostic decision</td>
<td>GAD</td>
<td>20 (32.8)</td>
</tr>
<tr>
<td></td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>22 (36.1)</td>
<td>20 (33.9)</td>
</tr>
</tbody>
</table>

Note. BPD = borderline personality disorder; GAD = generalized anxiety disorder; PD = panic disorder; PTSD = posttraumatic stress disorder.

TABLE 3  Results of the linear regression models associating the independent variables (pretreatment-report and information order conditions) with the dependent variable (confidence ratings and severity ratings of diagnoses)

<table>
<thead>
<tr>
<th></th>
<th>95%-CI for B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
</tr>
<tr>
<td>GAD vignette</td>
<td></td>
</tr>
<tr>
<td>Confidence ratings (0–100)</td>
<td></td>
</tr>
<tr>
<td>Pretreatment report</td>
<td>-9.396</td>
</tr>
<tr>
<td>Information order</td>
<td>-4.277</td>
</tr>
<tr>
<td>Severity ratings of diagnosis (1–8)</td>
<td></td>
</tr>
<tr>
<td>Pretreatment report</td>
<td>-1.680</td>
</tr>
<tr>
<td>Information order</td>
<td>-4.011</td>
</tr>
<tr>
<td>BPD/PTSD vignette</td>
<td></td>
</tr>
<tr>
<td>Confidence ratings (0–100)</td>
<td></td>
</tr>
<tr>
<td>Pretreatment report</td>
<td>-11.158</td>
</tr>
<tr>
<td>Information order</td>
<td>-8.814</td>
</tr>
<tr>
<td>Severity ratings of diagnosis (1–8)</td>
<td></td>
</tr>
<tr>
<td>Pretreatment report</td>
<td>-4.011</td>
</tr>
<tr>
<td>Information order</td>
<td>-7.366</td>
</tr>
<tr>
<td>PD vignette</td>
<td></td>
</tr>
<tr>
<td>Confidence ratings (0–100)</td>
<td></td>
</tr>
<tr>
<td>Pretreatment report</td>
<td>-8.973</td>
</tr>
<tr>
<td>Information order</td>
<td>-5.551</td>
</tr>
<tr>
<td>Severity ratings of diagnosis (1–8)</td>
<td></td>
</tr>
<tr>
<td>Pretreatment report</td>
<td>-2.711</td>
</tr>
<tr>
<td>Information order</td>
<td>-4.917</td>
</tr>
</tbody>
</table>

Note. BPD = borderline personality disorder; GAD = generalized anxiety disorder; PD = panic disorder; PTSD = posttraumatic stress disorder.

TABLE 4  Results of the logistic regression analyses associating the independent variables (pretreatment-report and information order conditions) with the dependent variable (fully correct diagnostic decision).

<table>
<thead>
<tr>
<th></th>
<th>95%-CI for OR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
</tr>
<tr>
<td>GAD vignette</td>
<td></td>
</tr>
<tr>
<td>Pretreatment report</td>
<td>-0.102</td>
</tr>
<tr>
<td>Information order</td>
<td>1.060</td>
</tr>
<tr>
<td>PD vignette</td>
<td></td>
</tr>
<tr>
<td>Pretreatment report</td>
<td>-0.102</td>
</tr>
<tr>
<td>Information order</td>
<td>0.975</td>
</tr>
</tbody>
</table>

Note. GAD = generalized anxiety disorder; PD = panic disorder.
In relation to the pretreatment report condition, there was no significant association between receiving a report and fully correct diagnostic decisions (GAD: OR = 0.634, $P = .256$; PD: OR = 0.903, $P = .792$; see Table 4). Psychotherapists who received a pretreatment report were not significantly more likely to make a fully correct diagnostic decision. In relation to the third case vignette (BPD and PTSD), we were not able to investigate the associations between the condition and the diagnostic decisions, because none of the 120 participants identified BPD and PTSD correctly.

4 | DISCUSSION

Considering the structure of diagnostic classification systems—assessing core criteria first and subsequently proceeding to verify or falsify a diagnosis by assessing the other criteria required—diagnosticians might also follow this structure when making diagnoses. Therefore, we expected that the correctness of diagnostic decisions would be associated with the presentation of the core symptoms of mental disorders at the beginning of the case vignettes. If so, results could have been seen as an indication for a primacy effect in clinical psychological diagnosing. According to the results of our study, this hypothesis has to be rejected. Rather, the data showed significant associations in the opposite direction in two case vignettes, indicating a recency effect. Additionally, as we expected, the order of symptom description was also related to psychotherapists’ confidence with diagnostic decisions or their estimation of disorder severity. This hypothesis has also to be rejected. However, studies in diagnostic reasoning indicate that—depending on the presentation rate—both a primacy effect and a recency effect occur when the order of symptom presentation is varied (Lange, Thomas, Buttaccio, Illingworth, & Davelaar, 2012; Lange, Thomas, & Davelaar, 2012; Rebitschek, Bocklisch, et al., 2015). Nevertheless, there is also evidence for a recency effect in medical decision-making (Bergus, Chapman, Levy, Ely, & Oppliger, 1998; Croskerry, 2003; Lawson & Daniel, 2011), consistent with our findings.

As mentioned previously, a stepwise diagnostic process may reduce the primacy effect in diagnostic reasoning. Further, Hogarth and Einhorn (1992) asserted that diagnosticians are more likely to be subject to the recency effect when they have to process and decide stepwise a sequence of information. Several other studies lend support to this assertion. For example, Kerstholt and Jackson (1998) showed that judicial decisions are more likely to be associated with a recency effect when the defendant’s probability of guilt is considered in a stepwise manner, and multiple studies demonstrate a recency effect in diagnostic reasoning when it occurs stepwise (Bergus et al., 1998; Rebitschek, Bocklisch, et al., 2015).

In addition, it appears that the likelihood of a recency effect increases in relation to the complexity and length of the information source (Hogarth & Einhorn, 1992). This assumption has been well-evidenced. For example, Canic and Pachur (2014) demonstrated that longer sequences of information increase the probability of a recency effect. Furthermore, Bergus et al. (1998) reported an increased likelihood of considering clinical data when it was presented last. Compared to other studies in medical decision-making, the diagnosticians in this study engaged in more complex decision-making than participants in studies that reported, for example, a primacy effect in medical decision-making. Higher complexity and longer length of sequences could be the reason for the recency effect observed in this study. All three case vignettes depicted several symptoms and additional information about the patients. Aside from the description level, the diagnosticians evaluated whether each symptom is part of a hypothesized mental disorder and/or whether it is an indication for an exclusion of another disorder. Handling multiple levels of information makes the diagnostic process in relation to mental disorders complex and may increase the likelihood of false decisions. The nonsignificant association between the information order and the estimation of disorder severity and confidence with given diagnoses as was expected, as merely the order of information was manipulated, not the content of the vignettes.

Also contrary to our hypothesis, we did not observe significant associations between the receipt of incongruent pretreatment reports and misdiagnoses in any of the three case vignettes. Further, pretreatment reports did not affect the therapists’ confidence with diagnostic decisions. One could speculate that participating therapists were not distracted by the decoy diagnoses named in the pretreatment reports, and therefore did not include the information from these reports in their diagnostic judgment. Support for this assumption comes from Mantonakis et al. (2009), putting forward that information given first is more likely to be included when the judging person has less knowledge in the field, and that judging persons with less knowledge are less likely to compare initially given information and information given later. The participants in this study had a high level of knowledge of mental disorders and diagnostic criteria, as indicated by their average 14.46 (SD = 10.14) years of work experience and high familiarity with the diagnostic systems ICD-10, DSM-IV, and DSM-5.

Furthermore, Rebitschek, Bocklisch, et al. (2015) showed how stepwise diagnostic reasoning partially counteracts the primacy effect and can even lead to a recency effect (see experiments 1B, 2B, and 3). Given this procedural influence, one could speculate that stepwise diagnostic reasoning may also counteract decoy effects. In contrast, Hogarth and Einhorn (1992) showed that it can even lead to a recency effect. Additionally, a study on sequential diagnostic reasoning demonstrated that increased activation of the initially formed hypothesis when reasoning about multiple candidate hypotheses leads to a primary effect (Rebitschek, Krems, et al., 2015). Given these results, one could speculate that stepwise diagnostic reasoning also counteracts decoy effects. The diagnostic process in relation to mental disorders is characterized by a continual reconciliation of symptoms with criteria of several disorders. Hence, each symptom listed in a case vignette could create a new diagnostic hypothesis or be used to reject previous hypotheses. Such a stepwise nature of clinical psychological diagnosing could be one possible explanation why we did not observe the expected decoy effect. Thus, it is possible that such multiple hypotheses testing could lead therapists to desist from hypothesis caused by the decoy. Alternatively, it is also possible that a memory-based reasoning-effect could explain these results: Complex vignettes are associated with substantial workload, which could have either interfered with memorizing the pre-treatment report information or could have removed initial hypotheses that were built on the
information from the pretreatment report from memory. However, given our design, we are unable to fully explain why we did not find a decoy effect.

Additionally, recent social psychological findings illustrate that new information (e.g., counterevidence) can result in fast and long-lasting reinterpretations of previous assumptions (Cone & Ferguson, 2015), which may be related to diagnostic processes (Mann, Cone, & Ferguson, 2015). This is in line with a study on clinical psychologists, which illustrated that this specific professional group is likely to incorporate new information in their diagnostic reasoning (De Kwaadsteniet, Kim, & Yopchick, 2013). In our study, the information in the case vignettes could be seen as new information that induced a reinterpretation of information and consequently might have led to a rejection of diagnostic hypotheses.

Conceivably, psychotherapists who received a pretreatment report in the present study integrated the information of the case vignette or consecutively compared the diagnoses named in the pretreatment report with their hypotheses. They then would proceed to evaluate the symptom description in the case vignette as incongruent with the pretreatment report and to reject counteracting hypotheses based on information named in the incongruent pretreatment reports. Accordingly, there was no significant association between diagnosticians' confidence with diagnostic decisions and the receipt of the report. This result suggests an independent evaluation of cases by psychotherapists, while taking into account that such diagnostic processing is not necessarily associated with the correctness of diagnostic decisions, as the data also indicates. Finally, it is possible that psychotherapists did not trust the pretreatment reports and thus were not distracted by the misleading diagnoses named in the reports. This would be in line with a study indicating that psychotherapists consider their own experience indispensable (Bruchmüller et al., 2011).

Simultaneously, the present study also shows that the receipt of a pretreatment report as an additional information source does not impact the diagnostic evaluation of psychotherapists concerning severity ratings. This was surprising to us, but could be explained by the fact that psychotherapists generally seem to conduct diagnostics independent of pretreatment reports. Future studies should additionally assess to what extent psychotherapists trust the given pretreatment reports.

More broadly, the results of this study illuminate again the problem of misdiagnoses in daily practice. In all three case vignettes, psychotherapists made aimed diagnoses in only almost a half of the cases, and false comorbid diagnoses in 65.0–100% of all cases. Indeed, fully correct diagnostic decisions were made at the most in 35.0% of the cases, with false diagnostic decisions as high as 100%, in the case of comorbid BPD and PTSD.

The high rate of misdiagnoses illustrates how susceptible psychotherapists are to diagnostic errors. Together with the results of the recency effect in relation to the order of information, our results suggest that an unstructured process in diagnosing mental disorders is highly prone to error. Diagnostic tools such as structured interviews may help to reduce diagnostic errors and give the user structure and decision support (Merten, Cwik, Margraf, & Schneider, 2017; Mitchell et al., 2009; Rettew, Lynch, Achenbach, Dumenci, & Ivanova, 2009).

5 | LIMITATIONS

The present study has some limitations. First, we used case vignettes to investigate diagnostic processes, which has inherent limitations. The case description in a vignette is only roughly comparable to obtaining information in a real therapeutic setting (Garb, 1998). The diagnostic information in vignettes is somewhat filtered. Patients' descriptions are usually not as structured as descriptions in vignettes. Additionally, diagnosticians are not able to pose additional inquiries when they are not confident with given information. Moreover, the present case vignettes presented participants with more or less simple cases that are not necessarily comparable to real cases, as Wolkenstein et al. (2011) have pointed out. However, case vignettes are a feasible tool for standardized investigations of diagnostic accuracy and are frequently used in such studies (e.g., Dalgleish, 2004; Flanagan & Blashfield, 2005; Meyer & Meyer, 2009; Wolkenstein et al., 2011).

Second, we only used cases with fulfilled diagnoses. As a consequence, it is not clear whether the observed associations are assignable to subclinical diagnoses (e.g., missing one criterion) that can take place in daily practice. Third, the exclusion rate of 54.72% of questionnaires was high. This could be the result of the relatively complex nature of the survey or the length of the questionnaire (35 min). Additionally, there were significant associations between dropouts and some demographic variables. Thus, it is unclear whether the results of this study may be affected by a selection bias. Fourth, for the calculation of propensity scores, we also included demographic variables whereby the groups only differed on the trend-level. The rationale for this approach was to control for all potential effects of group differences. However, if only significant (P < .05) demographic variables would have been used, all results would have remained significant but would not have survived the threshold for multiple comparisons. Taking this into consideration, even if our results regarding the recency effect are in line with former studies, it is possible that these results are affected by an accumulation of alpha errors. Thus, our results should be interpreted with caution and need to be replicated. Fifth, only German clinicians participated and, therefore, it is unclear whether these results can be generalized to therapists in other countries. Sixth, it should be considered that participants were instructed to judge without auxiliary materials, which possibly contradicted their day-by-day practice. Additionally, we did not assess participants' experience with open clinical judgment and thus, are unable to test whether the groups differed in this respect. However, participants did not differ significantly regarding years of work experience. Finally, the vignettes described constellations of symptoms based on the DSM-5, whereas German therapists are encouraged by the national health system to use ICD-10 criteria in providing treatment rationales for health insurance billing, and are thus more familiar with ICD-10 compared to DSM-IV and DSM-5.

CONCLUSION

This study makes important contributions that further advance our understanding of the influence of pre-treatment reports and order of information on diagnostic decisions. Surprisingly, the receipt of pre-
treatment reports was not related to misdiagnoses. Replication in other populations should be undertaken before generalizing this result. It remains unclear whether pre-treatment reports bias diagnostic decisions when they are received subsequent to a diagnostic interview, which may often occur in daily practice. In contrast, the results supported the hypothesis that information order has an effect on diagnosis. Considering the method of case vignettes and the limitations of this study, reported results should be seen as preliminary and thus should not be generalized for daily practice.

ACKNOWLEDGMENTS

We followed the TRENDS Statement (Des Jarlais, Lyles, Crepaz, & the TRENDS Group, 2004) for the manuscript preparation. We would like to thank Kristen Lavallee, Helen Copeland-Vollrath, and Simon E. Blackwell for critical reading and language editing of this manuscript. The authors have no conflict of interest to report.

ROLE OF FUNDING

This research was funded by an Alexander von Humboldt professorship, awarded to Jürgen Margraf.

REFERENCES


Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. Academic Medicine, 78(8), 775–780.


