

Tilburg University

On the use of positive test strategies when diagnosing mental disorders

Van Der Heijden, Paul T.; Cefo, Irma; Witteman, Cilia L.m.; Grootens, Koen P.

Published in:
Comprehensive Psychiatry

DOI:
[10.1016/j.comppsy.2022.152325](https://doi.org/10.1016/j.comppsy.2022.152325)

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
Van Der Heijden, P. T., Cefo, I., Witteman, C. L. M., & Grootens, K. P. (2022). On the use of positive test strategies when diagnosing mental disorders. *Comprehensive Psychiatry*, 116, Article 152325. <https://doi.org/10.1016/j.comppsy.2022.152325>

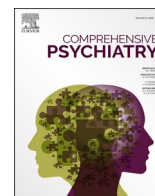
General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



On the use of positive test strategies when diagnosing mental disorders

Paul T. van der Heijden^{a,b,*}, Irma Cefo^a, Cilia L.M. Witteman^b, Koen P. Grootens^{a,c}

^a Reinier van Arkel Mental Health Institute, 's-Hertogenbosch, the Netherlands

^b Behavioural Science Institute, Radboud University Nijmegen, Nijmegen, the Netherlands

^c Tranzo, Tilburg School of Social and Behavioral Sciences, Tilburg University, Tilburg, the Netherlands

ARTICLE INFO

Keywords:

Psychiatric diagnosis
Bias
Confirmation bias
Positive test strategy
Decision making
Heuristics

ABSTRACT

Background: Despite the adverse impact diagnostic errors can have, clinical interviewing and decision-making in psychiatric practice have received relatively little empirical attention. When diagnosing patients, clinicians tend to fall back on a specific (heuristic) rule of thumb, the positive test strategy, a confirmatory approach that increases the risk of confirmation bias.

Method and results: A group of 83 clinical psychologists and psychiatrists was asked to give their diagnostic hypotheses about two vignettes. We found them to self-generate significantly (i.e., $p < .01$; $d = 1.57$) more confirming than disconfirming questions to test their initial diagnostic impressions, with supervisors considering significantly more differential diagnoses than the less experienced post-grads/residents. When offered a list of 100 potentially relevant diagnostic queries, the supervisors selected fewer confirming and proportionally more disconfirming themes.

Conclusions: Our results demonstrate that irrespective of clinical experience mental-health clinicians indeed tend to use a confirmatory thinking style that contrasts with the stricter principle of falsification. More field-based research on this topic is needed, as well as studies probing whether a systematized diagnostic approach is feasible in psychiatric practice and increases diagnostic accuracy and patient satisfaction.

1. Introduction

1.1. Diagnostic decision-making

Despite the detrimental impact diagnostic errors may have on patients, in terms of their management, treatment and potential for recovery, relatively little empirical research has been conducted on how clinical decision-making in psychiatric practice proceeds [1]. Without doubt, the diagnostic process is complex and based on mostly subjective, and sometimes incomplete and inconsistent information, past and current observations and (hetero)anamnestic data obtained during psychiatric interviews. Standardized manuals for diagnosing mental disorders, such as the DSM-5 [2] and ICD-11 [3], may improve diagnostic reliability, but, due to its 'inexact' nature, the process remains prone to observation and interpretation errors [4,5].

Having interviewed 31 attending psychiatrists about clinical decision-making in daily practice, Bhugra and colleagues [6] concluded

that, in addition to clinical experience, chance factors such as time, financial resources, and available treatment options considerably influenced their decisions. Such random or uncontrollable factors are then potential sources of errors and bias in any subsequent diagnostic process. An even larger source of diagnostic errors may lie in the ways doctors think, reason, solve problems, and arrive at decisions [7], while they seldom receive meaningful feedback on their decisions, which may lead to missed opportunities to correct any inadvertent mistakes or adopt a different decision strategy [8].

1.2. Heuristics in diagnostic decision-making

In order to make decisions in complex and uncertain contexts, people may fall back on heuristic decision rules. Heuristics can be described as straightforward but imperfect rules of thumb that enable people to reach decisions quickly and easily [9]. Such heuristic decisions ideally precede more thorough and balanced decisions that typically require more time

Abbreviations: ADHD, Attention-deficit/hyperactivity disorder; APD, Antisocial personality disorder; ASD, Autism spectrum disorder; CQ, Confirmatory question; DQ, Disconfirmatory question; DSM-5, Diagnostic and Statistical Manual of Mental Disorders - fifth edition; ICD-11, International Classification of Diseases - 11th edition; MDD, Major depressive disorder; NQ, Neutral question; PD, Personality disorder; SUD, Substance use disorder; TQ, Total questions.

* Corresponding author at: Reinier van Arkel Mental Health Institute, 's-Hertogenbosch, the Netherlands.

E-mail address: P.T.vander.Heijden@reiniervanarkel.nl (P.T. van der Heijden).

<https://doi.org/10.1016/j.comppsy.2022.152325>

Available online 13 May 2022

0010-440X/© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

and effort, but in practice this latter approach is often not adopted until a first heuristic decision is deemed insecure. Heuristic rules imply a strong bias towards confirming initial beliefs [9], precluding spontaneous or more well-considered decisions.

A typical example of a heuristic decision approach is the positive test strategy that is characterized by the exclusive tendency to confirm a hypothesis instead of trying to falsify it. Although heuristics may have some use in everyday clinical decision-making (by affording a quick diagnosis through pattern recognition, for instance) [1], a positive test strategy may also result in incomplete or even erroneous diagnoses when quick (and possibly imprecise) first impressions drive the diagnostic process in the wrong direction.

1.3. Confirmation bias

It has not yet been empirically established how often a positive test strategy is used in daily psychiatric practice, what circumstances prompt its use, and whether and when the approach leads to confirmation bias, where its aim is to confirm an initial impression or diagnosis with resulting ignorance or rejection of conflicting or disconfirming information. Mendel and colleagues [10] did investigate the use of positive test strategies in psychiatrists and medical students in an experimental decision task. Presenting their participants with a clinical case and two diagnostic options to choose from (i.e., Alzheimer's disease or depression), and asking them whether they were in need of additional information in order to reach their diagnosis and treatment plan, the authors found that 13% of the psychiatrists and 25% of the medical students showed confirmation bias in their information search. Moreover, they demonstrated that this bias led to poorer diagnostic accuracy. Arguably, since in daily psychiatric practice cases tend to be more complex and multiple hypotheses may apply, clinicians might resort to using a positive test strategy even more frequently. Therefore, and building on the findings reported by Mendel and colleagues, we created a diagnostic context that is more reflective of everyday mental-health practice to test confirmation bias.

We decided to ask psychologists and psychiatrists to formulate differential diagnostic hypotheses and probe their subsequent diagnostic strategies based on two vignettes of patients showing signs and symptoms indicative of multiple DSM-5 diagnoses. After having judged the first vignette, the participants were instructed to write down up to ten questions with which they would be assessing their diagnostic impressions. With the second vignette they were invited to select a maximum of 15 questions from a list presenting 100 theoretically relevant diagnostic questions. Compared to Mendel et al.'s study, we expect more respondents to ask more confirming questions (i.e., questions that are likely to corroborate the diagnosis they deemed most likely) than disconfirming questions (i.e., questions that might counter this assumption) in both conditions, which would point at an overriding use of a positive test strategy. At the same time, we anticipate this tendency to be stronger when participants need to formulate their own questions than when they are enabled to derive their queries from the longlist, where they may be less likely to be (mis)guided by subjective stereotypes and tunnel vision [11–13].

Additionally, Mendel and colleagues demonstrated that the medical students they evaluated were more affected by confirmation bias than the more experienced psychiatrists. However, there is also empirical evidence suggesting that professional experience has limited influence on the decision strategies and quality of the decisions of mental-health professionals [14,15]. Experienced clinicians may get overconfident about their diagnostic judgements [16], possibly resulting in a less critical attitude towards their own impressions. We will also be looking into this issue, anticipating that clinical experience will not be related to the more frequent use of positive test strategies.

2. Material and methods

Drawing from the DSM-5 case book, we modelled two vignettes of patients with characteristics of multiple disorders, based on the DSM-5 case book (the exact representation of the vignettes is available on request from the first author). Vignette 1 concerns a male patient presenting with signs and symptoms of an antisocial personality disorder (APD) and substance use disorder (SUD). The vignette states that the patient himself wants to know if he has attention-deficit/hyperactivity disorder (ADHD) because his sons have also been diagnosed with the disorder. Vignette 2 involves a letter of referral from a general practitioner concerning a male adolescent with suspected autism spectrum disorder (ASD). The young man presents with signs and symptoms fitting the following DSM-5 diagnoses: ADHD, ASD, major depressive disorder (MDD), personality disorder (PD). He also has psychotic symptoms. He is experiencing problems at school and in social functioning. Vignette 2 finally states that ASD runs in his family.

Supervisors, all attending psychiatrists and clinical psychologists, and psychiatry residents and postgraduates in clinical psychology (PsyDs; the latter two referred to as juniors) consulting and training in a mental-health hospital experienced in providing well-established training programs in both disciplines in the south of The Netherlands, were asked to participate in the study. Recruitment was via emails to eligible candidates via their professional email account, residents were asked to participate after courses at the hospital. The survey was completed in a single group meeting in the hospital and took about 30 min to complete. Eligible participants were informed that the survey was designed to investigate how mental-health clinicians generally come to a diagnosis. They were assured that their identity would remain anonymous. After having obtained informed consent, a research assistant unknown to the participants provided them with the patient vignettes and paper and pencil, requesting them to write down their hypotheses for the three DSM-5 diagnoses they deemed most likely. Next, they were instructed to indicate additional questions they would be seeking an answer to in order to arrive at their final diagnosis. They were also informed that no conferring was allowed. Having received the first vignette (Condition 1), the respondents were asked to write down any questions (with a maximum of 10) that they judged relevant for their diagnostic decisions. With the second vignette (Condition 2), they were instructed to select a maximum of 15 questions from a list comprising 100 potentially relevant questions to guide the diagnostic process, which were formulated based on DSM-5 criteria for multiple diagnostic classifications and semi-structured interviews. (The list can be obtained from the first author).

Independently from each other, the researchers (PvdH, IC, KG) first categorized each question drafted by the participants for Vignette 1 (Condition 1) and selected from the list for Vignette 2 (Condition 2) as a confirmatory question (CQ), a disconfirmatory question (DQ), or a neutral question (NQ). An example of a CQ in case of a suspected narcissistic PD is: "Do you long for admiration?," where a DQ may read: "Do you feel uncomfortable in social situations?" and an NQ: "Can you tell something about your elementary school period?." Over 80% all questions in both conditions had been categorized identically by all three researchers. Subsequently, the remaining questions were discussed and categorized on the basis of consensus. For the purpose of analyses, the number of CQs and DQs for each of the two vignettes was determined for each respondent.

The mean numbers of CQs and DQs per vignette were calculated and the difference between these means was tested with a one-sided *t*-test. In addition, we calculated the ratio between the number of CQs and the total number of questions (TQ) per participant, and similarly the DQ/TQ ratio, and expressed both as a proportion (a number between 0 and 1). Since TQs per participant differed for the two conditions (Vignette 1: TQ ≤ 10 questions; Vignette 2: ≤ 15), we contrasted the mean CQ/TQ and DQ/TQ proportions for Vignette 1 with those obtained for Vignette 2. We set the significance level at $p = .01$ and calculated effect sizes

Table 1

The participants' mean (*M*) ages and years of professional experience, with standard deviations (*SD*), as a function of professional status.

| Discipline and professional status | n | Age | | Experience | |
|------------------------------------|----|----------|-----------|------------|-----------|
| | | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Medical intern | 2 | 22.5 | 2.1 | 0 | |
| Psychiatry resident (MSc.) | 23 | 30.6 | 4.7 | 2.8 | 2.2 |
| Psychiatrist/Supervisor (M.D.) | 25 | 45.9 | 8.7 | 17.0 | 8.9 |
| Psychologist (MSc.) | 31 | 29.9 | 5.8 | 3.3 | 3.8 |
| Clinical psychologist (PsyD.) | 5 | 52 | 10.8 | 21.2 | 14.0 |

because of multiple tests.

3. Results

A total of 86 participants, of whom 73% were self-identified women, completed our survey. Details on age, professional status and discipline, and clinical experience are listed in Table 1. We like to point out once again that the initial diagnostic hypotheses the participants offered were not the subject of our investigations. Rather, they were the starting point for our research into the subsequent decision process.

Having studied Vignette 1, the respondents recorded the following diagnostic classifications as most likely: Narcissistic PD (27%), Cluster B PD (22%), Antisocial PD (17%), Antisocial / Narcissistic PD (14%), ADHD (4%), and 'other' (17%). In this latter category, diagnoses mainly concerned bipolar disorder / hypomania and SUD, but Intellectual Disability and paranoid disorder were also mentioned. The most likely diagnoses they documented for Vignette 2 (at least once) were ASD (74%), ADHD (6%), Schizophrenia (4%), Evasive PD (2%), Mild Intellectual Disability (2%), MDD (2%) and 'Don't know' (2%).

Table 2 shows the mean number of CQs and DQs per vignette as well as the proportions of CQs and DQs in relation to its TQ. The analysis yielded no significant difference between the mean CQ/TQ ratios for the two conditions ($T = 0.39; p = .70$). As to the DQ/TQ ratio, we did find a significant difference: in Condition 2 relatively more DQs had been selected than had been put forward in Condition 1 ($T = -5.63; p < .001$; Cohen's $d = -0.82$).

To investigate the effect of years of professional experience on decision strategies, we contrasted the datasets of the juniors ($n = 56$) having a mean 2.95 years of experience ($SD = 3.17$) to those of the supervisors ($n = 30$) whose mean experience amounted to 17.7 years ($SD = 9.76$). The difference in experience proved to have a large effect ($T = -8.05; p < .01$; Cohen's $d = -2.03$). As Table 3 shows, in Condition 1 (Vignette 1) the supervisors offered more differential diagnostic considerations than the less experienced juniors, while in Condition 2 (Vignette 2) they asked fewer CQs and more DQs than the juniors did, with correspondingly lower CQ/TQ and higher DQ/TQ ratios.

Table 2

The mean numbers (*M*) and standard deviations (*SD*) of confirmatory (CQ) and disconfirmatory questions (DQ) and their proportions relative to the maximum number of questions (TQ) per vignette.

| Vignette 1 | | | | | | | | | | | | | |
|------------|-----------|----------|-----------|----------|----------|----------|----------|-----------|----------|-----------|----------|----------|----------|
| CQ | | DQ | | | | | CQ/TQ | | DQ/TQ | | | | |
| <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>T</i> | <i>p</i> | <i>d</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>T</i> | <i>p</i> | <i>d</i> |
| 5.69 | 2.75 | 2.10 | 1.70 | 8.73 | <0.01 | 1.57 | 0.61 | 0.26 | 0.23 | 0.18 | 11.14 | < 0.01 | 1.70 |
| Vignette 2 | | | | | | | | | | | | | |
| CQ | | DQ | | | | | CQ/TQ | | DQ/TQ | | | | |
| <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>T</i> | <i>p</i> | <i>d</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>T</i> | <i>p</i> | <i>d</i> |
| 8.45 | 3.23 | 5.50 | 3.28 | 4.25 | <0.001 | 0.91 | 0.60 | 0.23 | 0.40 | 0.23 | 5.70 | < 0.01 | 0.87 |

Note: $d =$ Cohen's d (effect size).

4. Discussion

In line with Mendel et al.'s conclusion that clinicians are inclined to try and confirm their initial considerations [10], we found that most of the clinicians in our study indeed mostly applied a positive test strategy, reflecting a confirmation bias. When diagnosing vignettes of patients, they all generated more confirming than disconfirming questions to verify their diagnostic hypotheses, independent of years of professional experience. Importantly, when selecting questions from a longlist describing relevant diagnostic queries, however, they applied a more disconfirming style that might disprove their initial diagnostic hypotheses, with the supervisors selecting significantly more disconfirming questions than the less experienced juniors.

Since the use of a predominantly positive test strategy may lead to diagnostic errors [10], this latter finding is promising and substantiates previous empirical findings that systematization of the clinical decision-making process by means of pre-formulated diagnostic criteria enhances diagnostic reliability, reducing the risk of errors [e.g., 13, 17]. However and disconcertingly, when diagnosing mental disorders in psychiatric practice, clinicians do not always adhere to classification recommendations [18,19].

As predicted based on a meta-analysis failing to find that professional experience improved diagnostic decision-making [15], the relationship

Table 3

Differences in the number of classifications mentioned and the application of a positive test strategy.

| Vignette 1 | Junior | | Supervisor | | <i>T</i> | <i>p</i> | <i>d</i> |
|------------|----------|-----------|------------|-----------|----------|----------|----------|
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | | | |
| DD | 0.30 | 0.74 | 1.57 | 1.70 | -3.89 | < 0.001 | -0.97 |
| CQ | 5.70 | 2.63 | 5.66 | 3.00 | 0.073 | 0.94 | |
| DQ | 2.16 | 1.73 | 1.93 | 1.65 | 0.61 | 0.55 | |
| TQ | 9.33 | 1.30 | 9.21 | 2.13 | 0.29 | 0.77 | |
| pp CQ | 0.63 | 0.25 | 0.61 | 0.27 | 0.36 | 0.72 | |
| pp DQ | 0.24 | 0.184 | 0.21 | 0.17 | 0.59 | 0.56 | |
| Vignette 2 | Junior | | Supervisor | | <i>T</i> | <i>p</i> | <i>d</i> |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | | | |
| CQ | 9.11 | 3.23 | 7.15 | 2.84 | 2.79 | < 0.01 | 0.65 |
| DQ | 4.94 | 3.23 | 6.59 | 3.17 | -2.19 | 0.03 | |
| TQ | 14.07 | 1.90 | 13.74 | 1.85 | 0.76 | 0.45 | |
| pp CQ | 0.65 | 0.22 | 0.52 | 0.21 | 2.42 | 0.02 | 0.60 |
| pp DQ | 0.34 | 0.23 | 0.48 | 0.21 | -2.76 | < 0.01 | -0.64 |

Note: $d =$ Cohen's d (effect size); DD = differential diagnostic considerations (number of possible classifications mentioned); CQ = confirmatory questions; DQ = disconfirmatory questions; TQ = total questions; pp. = proportion relative to TQ.

between the use of a positive test strategy and years of professional experience in our study was weak. Only after our participants had been presented with a comprehensive checklist, did we find a significant difference in favor of the supervisors, whose experience exceeded that of the juniors sixfold. It has been suggested that it is hard to learn from experience in a context of complex and uncertain information in the absence of gold standards and systematic feedback on diagnostic decisions [20].

We think that with our vignette design mimicking mental-health practice in the Netherlands more closely, the relatively large number of well-trained professionals participating, and their diversity in terms of discipline and clinical experience, our findings are highly relevant. However, we were not able to monitor the participants' thought processes as in a 'thinking-out-loud' format. Although we did gain insight into the decision strategies that were adopted, our design prevented us from determining the consequences of these strategies for the correctness of the diagnostic decisions. Moreover, diagnosing mental disorders is a multifaceted process requiring numerous skills, such as gaining the trust of patients and maintaining a conducive therapeutic relationship, facets we did not address in our study. Also, although we and others have shown that the use of symptom checklists can improve diagnostic accuracy in an empirical setting, we do not know whether and how it will affect the patients in terms of quality and perceived quality of care.

In conclusion, a confirmatory working style contrasts with the stricter scientific practice of falsification. Posing that systematization in terms of pre-formulated diagnostic questions and a healthy amount of doubt will help improve diagnostic decision-making and boost diagnostic accuracy, we recommend further in-depth research into thinking processes, confirmatory diagnostics, confirmation bias, resultant diagnostic failures, and patient (dis)satisfaction in psychiatric settings and investigations examining whether a structured decision-making procedure prompting differential diagnoses has a positive effect on these parameters.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

None.

References

- [1] Crumlish N, Kelly BD. How psychiatrists think. *Adv Psychiatr Treat* 2009;15:72–9.
- [2] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Washington, DC: Author; 2013.
- [3] World Health Organization. *International Statistical Classification of Diseases and Related Health Problems*. 11th ed. 2020. <https://icd.who.int/>.
- [4] Paris J. *The Intelligent Clinician's Guide to the DSM-5*. New York: Oxford University Press; 2013.
- [5] Vanheule S, Desmet M, Meganck R, Inslegers R, Willemsen J, De Schryver M, et al. Reliability in psychiatric diagnosis with the DSM: old wine in new barrels. *Psychother Psychosom* 2014;83:313–4.
- [6] Bhugra D, Easter A, Mallaris Y, Gupta S. Clinical decision making in psychiatry by psychiatrists. *Acta Psychiatr Scand* 2011;124:403–11.
- [7] Croskerry P. A model for clinical decision-making in medicine. *Med Sci Educator* 2017;27:1–13.
- [8] Zarin D, Earls F. Diagnostic decision making in psychiatry. *Am J Psychiatry* 1993;150:197–206.
- [9] Kahneman D. *Thinking, fast and slow*. New York: Farrar, Straus and Giroux; 2011.
- [10] Mendel R, Traut-Mattausch E, Jonas E, Leucht S, Kane JM, Maino K, et al. Confirmation bias: why psychiatrists stick to wrong preliminary diagnoses. *Psychol Med* 2011;41:2651–9.
- [11] Garb HN. Clinical judgment and decision making. *Annu Rev Clin Psychol* 2005;1:67–89.
- [12] Ramirez Basco M, Bostic JQ, Davies D, Rush AJ, Witte B, Hendrickse W, et al. Methods to improve diagnostic accuracy in a community mental health setting. *Am J Psychiatry* 2000;157:1599–605.
- [13] Zimmerman M, Mattia JL. Differences between clinical and research practices in diagnosing borderline personality disorder. *Am J Psychiatry* 1999;156:1570–4.
- [14] Schulte-Mecklenbeck M, Spaanjaars NL, Wittman CLM. The (in)visibility of psychodiagnosticians' expertise. *J Behav Decis Mak* 2015;30:86–94.
- [15] Spengler PM, White MJ, Ægisdóttir S, Maugherman AS, Anderson LA, Cook RS, et al. The meta-analysis of clinical judgment project: effects of experience on judgment accuracy. *Couns Psychol* 2009;37:350–99.
- [16] Croskerry P, Norman G. Overconfidence in clinical decision making. *Am J Med* 2008;121:S24–9.
- [17] Westen D, Weinberger J. When clinical description becomes statistical prediction. *Am Psychol* 2004;59:595–613.
- [18] Reed GM, Correia JM, Esparza P, Saxena S, Maj M. The WPA-WHO global survey of psychiatrists' attitudes towards mental disorders classification. *World Psychiatry* 2011;10:118–31.
- [19] Zimmerman M, Gallione J. Psychiatrists' and non-psychiatrist physicians' reported use of the DSM-IV criteria for major depressive disorder. *J Clin Psychiatry* 2010;71:235–8.
- [20] Tracey TJ, Wampold BE, Lichtenberg JW, Goodyear RK. Expertise in psychotherapy: an elusive goal? *Am Psychol* 2014;69:218–29.