

Original Article

Cite this article: Cohen, S., Zantvoord, J. B., de Boer, T., Mattila, T., van Wingen, G., & Denys, D. (2025). Does early non-improvement predict treatment failure in pharmacotherapy for obsessive-compulsive disorder? A diagnostic test accuracy meta-analysis with individual participant data. *Psychological Medicine*, 55, e237, 1–7
<https://doi.org/10.1017/S0033291725101335>

Received: 06 May 2025

Revised: 04 July 2025

Accepted: 12 July 2025

Keywords:






antidepressants; diagnostic test accuracy; individual participant data meta-analysis; obsessive-compulsive disorder; response prediction; selective serotonin-reuptake inhibitors

Corresponding author:

Sem Cohen;

Email: s.e.cohen@amsterdamumc.nl

Does early non-improvement predict treatment failure in pharmacotherapy for obsessive-compulsive disorder? A diagnostic test accuracy meta-analysis with individual participant data

Sem Cohen^{1,2} , Jasper Brian Zantvoord^{1,2} , Ton de Boer³, Taina Mattila³ , Guido van Wingen^{1,2}  and Damiaan Denys^{1,2} 

¹Department of Psychiatry, Amsterdam UMC, Amsterdam, The Netherlands; ²Has nog differing departments, Amsterdam Neuroscience Research Institute, Amsterdam, The Netherlands and ³Medicines Evaluation Board, Utrecht, The Netherlands

Abstract

Background. In the treatment of obsessive-compulsive disorder (OCD) with antidepressant medication, the earliest reliable indication of treatment failure remains uncertain. We investigated if non-improvement following 4 weeks of treatment predicts nonresponse at the end of the trial.

Methods. We conducted a random-effects bivariate diagnostic accuracy study using individual patient data from industry-sponsored short-term trials of adults with OCD receiving selective serotonin reuptake inhibitors or clomipramine, submitted for marketing approval. The primary outcome was accuracy of non-improvement (<25% reduction on the Yale–Brown Obsessive Compulsive Scale [YBOCS] after 4 weeks) in predicting nonresponse (<35% YBOCS reduction at trial endpoint [10–13 weeks]). Secondary outcomes were accuracy of non-improvement after 6 weeks, nonresponse after 8 weeks, and inclusion of Clinical Global Impression Scale – Improvement in definitions of improvement and response. We performed meta-regressions for sex, age, severity, trial duration, dosing regimen, and compound.

Results. In 11 studies totaling 1,753 patients, non-improvement at week 4 predicted subsequent nonresponse (positive predictive value, PPV) in 86% of cases (95% confidence interval [CI] = 83–88%). Sensitivity was 78%, specificity was 70%, and the negative predictive value was 60%. Secondary outcomes showed similar PPV after 6 weeks and a PPV of 93% for nonresponse after 8 weeks. Predictive accuracy was significantly higher in men relative to women ($\beta = -0.64$, 95% CI = -1.12 to -0.16 , $p = 0.0089$).

Conclusions. Patients with OCD who do not improve after 4 weeks of antidepressants will likely not respond to short-term treatment. Thus, a change in strategy should be considered after 4 weeks without treatment benefits.

Introduction

Patients with obsessive-compulsive disorder (OCD) fail to respond adequately to first-line antidepressant treatment, such as selective serotonin reuptake inhibitors (SSRIs) or clomipramine, in ~40–60% of cases. In case of insufficient response, clinical guidelines recommend dose escalation or switching to an alternative drug. Treatment efficacy is typically evaluated after 12 weeks of treatment, which is long compared to other psychiatric conditions, such as major depressive disorder and schizophrenia (Carolyn Rodriguez, Bloch, Sachs, & Williams, 2015; Fineberg, Brown, Reghunandanan, & Pampaloni, 2012; Koran, Hanna, Hollander, Nestadt, & Simpson, 2007). American Psychiatric Association guidelines recommend continuing for at least 8–12 weeks after reaching the maximum recommended dose before considering a patient a nonresponder (Koran et al., 2007), although the evidence for this timeframe remains debatable (Issari, Jakubovski, Bartley, Pittenger, & Bloch, 2016; Pittenger & Bloch, 2014). Identifying nonresponse sooner in the treatment process would benefit both patients and clinicians by enabling earlier adjustments in treatment strategies. If early indicators of treatment failure can be reliably assessed, clinicians may more promptly revise the intervention – such as by switching medication or introducing psychotherapy – thereby reducing treatment delays and minimizing patients' experiences of failure.

In OCD, some relatively small studies have investigated whether symptom change after 4 weeks can predict eventual response to pharmacotherapy. However, definitions of nonresponse differed across studies, and most studies did not use a priori definitions of nonresponse, making these exploratory in nature (Ackerman, Greenland, & Bystritsky, 1996; Brar, Sidana, Chauhan, & Bajaj, 2022; da Conceição Costa et al., 2013; Krompinger et al., 2017). Methodological differences

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (<http://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is used to distribute the re-used or adapted article and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.



in trial settings, patient populations, and chosen response times have limited the clinical utility of these studies. Nevertheless, these preliminary results do suggest that early response evaluation on an individual level might inform clinical practice, even though, on a group level, pharmacotherapy in OCD might take longer to show clinically relevant differences compared to placebo.

In the current diagnostic accuracy study, we aimed to investigate whether lack of improvement after 4 weeks of treatment could predict nonresponse at the end of the trial, on an individual patient level. In a meta-analytic framework, we utilized crude data from short-term, double-blind, placebo-controlled pharmacotherapy trials for OCD in adults. We hypothesized that early non-improvement would be predictive of nonresponse and explored the predictive accuracy across different patient and study subgroups.

Methods

We prepublished a protocol outlining data extraction and substantiating our choice of outcomes and covariates (link: osf.io/whm7r, also available in [Supplementary Material](#)). Our study was done according to the Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy guidance, while taking into account that we did not perform a systematic review (McInnes *et al.*, 2018).

Outcomes

The primary outcome was the diagnostic accuracy of the index test (early non-improvement after 4 weeks) for predicting the reference test (nonresponse at trial end). Early non-improvement was defined as <25% reduction on the Yale–Brown Obsessive Compulsive Scale (YBOCS), reflecting the minimal clinically noticeable improvement (minimal important difference), making it a relevant cutoff (M. Cervin *et al.*, 2025; Cohen *et al.*, 2024). Nonresponse at the end of the trial was defined as symptom reduction of <35%, a common threshold in OCD trials (Nakatani *et al.*, 2005; Ramakrishnan *et al.*, 2024).

Secondary outcomes included diagnostic accuracy of (1) <25% response after 6 weeks, (2) <35% response at 8 weeks, and (3) <25% YBOCS-change or Clinical Global Impression Scale – Improvement (CGI-I) > 3 (no minimal improvement) after 4 weeks, each predicting <35% YBOCS-change or CGI >2 (no substantive improvement) at the end of the trial (Busner & Targum, 2007). Including CGI-I provided a more stringent measure, requiring both symptom reduction on YBOCS and observed clinical improvement (Mataix-Cols *et al.*, 2022).

Trial selection and data extraction

We analyzed individual patient-level data from double-blind, randomized, placebo-controlled trials submitted to the Dutch Medicines Evaluation Board (MEB) assessing the efficacy of SSRIs for adult OCD patients diagnosed using Diagnostic and Statistical Manual of Mental Disorders-III, III-R, or IV criteria. Confidentiality agreements between the trial sponsors and the MEB prevented us from specifying the compounds that were used. Only patients randomized to active treatment (SSRIs or clomipramine as active comparator) were included. Baseline and biweekly symptom scores, patient-level data (age, sex, and symptom severity), trial duration (10 vs. 12–13 weeks), and dosing schedule (fixed vs. flexible) were extracted. Age was dichotomized at 40 years, and severity was categorized as mild (YBOCS ≤ 21), moderate (22–29), or severe

(≥30), per validated cutoffs (Matti Cervin, Consortium, & Mataix-Cols, 2022). Missing data were imputed using multiple imputation by chained equations per trial.

Risk of bias analysis

We evaluated the risk of bias and applicability of the results with the quality assessment tool for diagnostic accuracy studies (QUADAS) 2 tool (Whiting *et al.*, 2011). Two assessors (S.E.C. and J.B.Z.) independently assessed the four key domains: patient selection, index test, reference standard, and flow/timing. Discrepancies were resolved by consensus and/or a third reviewer.

Data analysis

To account for clustering within trials and between-trial differences, we used a two-stage meta-analytic approach for analyzing crude patient data (Abo-Zaid *et al.*, 2013; Burke, Ensor, & Riley, 2017). In the first stage, we created a confusion matrix subdividing patients into true negatives, false positives, false negatives, and true positives. In the second stage, we conducted a bivariate random-effect meta-analysis of sensitivity/specificity and positive predictive value (PPV)/negative predictive value (NPV), accounting for inverse correlation (Leeftang, Deeks, Rutjes, Reitsma, & Bossuyt, 2012; Reitsma *et al.*, 2005).

Study estimates with 95% confidence intervals (CIs) were plotted in a summary receiver operating curve (sROC), and the area under the curve (AUC) was calculated as an overall accuracy measure.

We focused on nonresponse rather than response, as identifying when to cease medication – if unlikely beneficial – is more clinically relevant than knowing when to continue, which is already standard practice shortly after treatment initiation. Early non-improvement indicated a positive index test, while early improvement indicated a negative index test. Similarly, subsequent nonresponse (at 12 weeks) implied a positive reference test, and subsequent response implied a negative reference test.

Meta-regression and sensitivity analysis

We performed exploratory meta-regressions for prespecified patient- and study-level variables, using the log odds ratio (OR, a univariate measure combining specificity and sensitivity) as outcome. In case of statistical significance, we conducted subgroup analyses for bivariate accuracy measures (sensitivity/specificity, PPV/NPV) and AUC of the sROC curve. We further conducted a meta-regression for influence of study compound on the accuracy.

We performed sensitivity analyses excluding patients who dropped out before week 10 or week 4, patients randomized to clomipramine, and outlying trials identified by visual inspection of the ROC curve.

Results

Included trials

Out of 14 double-blind, randomized, placebo-controlled trials in adult OCD that were submitted to the MEB by the sponsors, we included 11 trials. Two trials were excluded as they did not provide outcome data, and one trial had to be excluded as information on the assignment of intervention per patient was not available. After excluding the patients assigned to placebo ($n = 885$), we included a total of 1,753 patients (267 of whom were assigned to clomipramine and 1,486 to an SSRI). Three trials (27%) lasted 10 weeks, seven

trials (64%) lasted 12 weeks, and one trial (9%) lasted 13 weeks. Of the included patients, 947 were male (54%) and 806 (46%) were female. The mean age of included patients was 38.2 years (standard deviation [SD] = 12.2) and the mean YBOCS severity at baseline was 24.5 (SD = 5.1), which is indicative of moderate OCD severity. See Table 1 for an overview of the included trials. The mean attrition after 10 weeks was 21.9%. For an overview of attrition after 10 weeks per trial, please refer to Supplementary Table 1.

After 4 weeks, 1,220 patients (70%) were positive for the index test of non-improvement (i.e. <25 percent YBOCS reduction compared to baseline), while 534 (30%) were negative for the index test (i.e. patients did improve more than 25% compared to baseline). A total of 1,107 patients (63%) met the criteria for the reference test (i.e. nonresponse [<35% YBOCS reduction compared to baseline] at the end of the trial [10–13 weeks]), while 647 (37%) patients responded (i.e. negative reference test). For an overview of classifications of index test and reference test per trial, please see Supplementary Table 2.

Dosing regimen

Three trials (27%) were fixed-dose studies, eight (73%) were flexible-dose studies. All trials used a dosing regimen where patients were titrated to a minimally effective dose within 1 week, after which patients were either uptitrated according to a fixed-dose regimen, or guided by adverse events or effect according to a flexible-dose regimen. Individual patient dosing information was not available for flexible-dose studies. For an overview of dose ranges in fluoxetine equivalent, see Table 1.

Quality assessment

Regarding patient selection, all studies except one were conducted on the North-American continent, causing “some concerns” in applicability according to the QUADAS tool. As the index tests and reference tests were all prespecified, blinded, and clinically relevant, namely the YBOCS, we found no concerns in Domains 2 and 3 for risk of bias and applicability. Three studies out of 11 had an attrition rate of more than 25% introducing possible bias in flow and timing of the study, as not all patients had experienced the

reference test. Three studies had a 10-week duration instead of a 12-week duration as was our prespecified duration. See Figure 1 for an overview of risk of bias, and Supplementary Table 3 for risk of bias per trial.

Main outcome

The main outcome was with which accuracy early non-improvement (i.e. <25% improvement after 4 weeks) could predict ultimate non-response (i.e. <35% improvement at the end of the trial). The results showed that non-improvement after 4 weeks predicted nonresponse at the end of the trial with a PPV (correctly identifying nonresponse) of 86%, a sensitivity of 78%, a specificity of 70%, and an NPV of 60%. The AUC was 0.82. See Figure 2 for the sROC figure with specificity and sensitivity per trial, and Table 2 for an overview of primary and secondary outcomes, including 95% CIs.

Secondary outcomes

Non-improvement (<25% improvement) after 6 weeks had the following diagnostic accuracy measures for predicting nonresponse at the end of the study: AUC of 0.86, sensitivity of 85%, specificity of 73%, PPV of 84%, and NPV of 74% (see Supplementary Figure 1 for the SROC curve). Nonresponse (<35% improvement) after 8 weeks of treatment showed the following diagnostic accuracy for predicting nonresponse at the end of the study: AUC of 0.91, sensitivity of 84%, specificity of 85%, PPV of 93%, and NPV of 70% (see Supplementary Figure 2). Early non-improvement and ultimate nonresponse according to a more stringent criterion, by combining the CGI with YBOCS, yielded an accuracy that was comparable to the primary outcome (see Supplementary Figure 3).

Meta-regression

Meta-regression on the main outcome for patient- and study-level variables showed that prediction accuracy was significantly higher in male patients compared to female patients (β -log OR = 0.64, 95% CI = 0.16–1.12, $p = 0.0089$; see Table 3). We then performed additional analyses for accuracy measures. For male patients, all accuracy measures were higher than for female patients (see

Table 1. Trial characteristics

Trial	N	Age (mean, SD)	Sex (% female/male)	Active treatments	Trial duration (weeks)	Dosing regimen	Baseline severity (YBOCS, SD)	Mean dose, end, fluoxetine eq.
1	266	36.8 (11)	56/44	SSRI	13	Fixed-dose finding	23.8 (5.5)	20/40 /60 mg
2	80	35.3 (11)	53/47	SSRI	10	Flexible	23.3 (4.8)	50–60 mg
3	73	36.6 (10)	51/59	SSRI	10	Flexible	22.7 (6.1)	50 mg
4	250	40.6 (14)	25/75	SSRI	12	Fixed-dose finding	24.2 (5.9)	20/40 /60 mg
5	164	38.7 (12)	44/56	SSRI, clomipramine	12	Flexible	23.8 (4.8)	50 mg
6	299	38.0 (12)	51/49	SSRI, clomipramine	12	Flexible	25.4 (5.1)	40 mg
7	44	37.0 (11)	22/78	SSRI	10	Flexible	23.3 (4.8)	56 mg
8	76	39.3 (13)	61/39	SSRI	12	Flexible	27.0 (3.7)	?
9	239	40.1 (13)	43/57	SSRI	12	Fixed-dose finding	23.9 (5.2)	20/40 /60 mg
10	177	37.7 (12)	55/45	SSRI, clomipramine	12	Flexible	25.5 (3.8)	?
11	86	35.3 (11)	43/57	SSRI	12	Flexible	25.2 (3.8)	53 mg

Abbreviations: mg, milligram; SD, standard deviation; SSRI, selective serotonin reuptake inhibitor; YBOCS, Yale–Brown obsessive-compulsive scale.

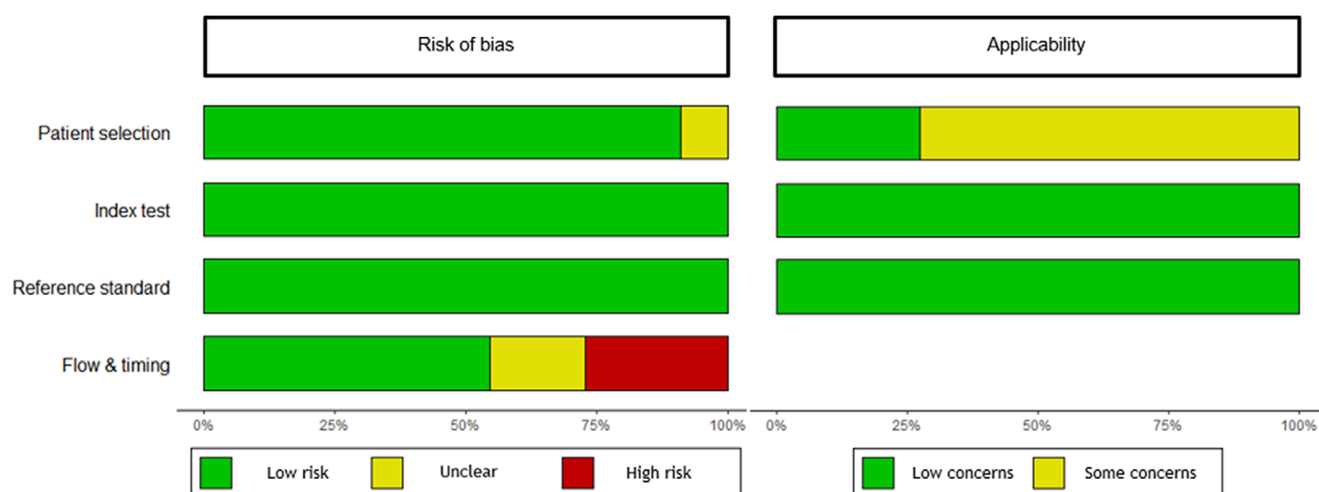


Figure 1. Risk of bias and concerns regarding applicability.

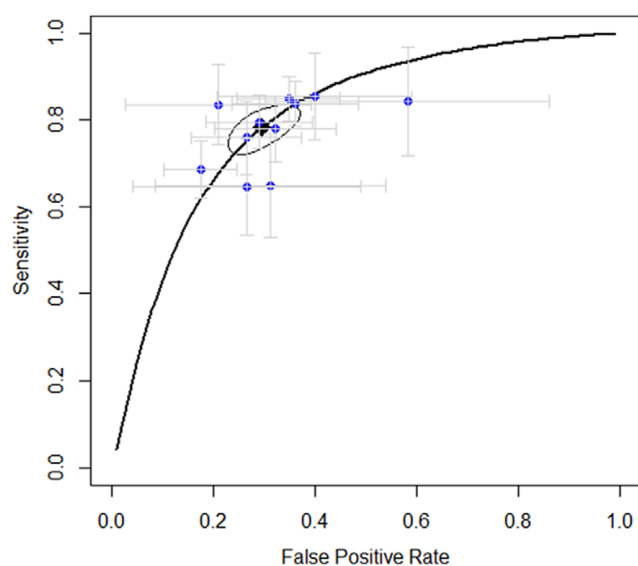


Figure 2. Summary of the receiver operating curve.

Table 3). An exploratory meta-regression for prediction at 6 weeks did not show significant differences for gender. Age, illness severity, compounds, dosing schedule, and trial duration did not significantly impact accuracy (see [Supplementary Table 4](#) for outcomes for all meta-regression analyses).

Sensitivity analyses

Sensitivity analysis on the main outcome after excluding patients who dropped out before week 4 yielded an AUC of 0.81, with a sensitivity of 77% (71–83), a specificity of 73% (67–88), PPV of 86%

Table 2. Primary and secondary outcomes

Outcome	AUC	Sensitivity	CI	Specificity	CI	PPV	CI	NPV	CI
Non-improvement after 4 weeks	0.82	78	73–83	70	64–76	86	83–88	60	56–64
Non-improvement after 6 weeks	0.86	85	81–88	73	67–78	84	81–87	74	70–77
Nonresponse after 8 weeks	0.91	84	81–87	85	80–89	93	91–95	70	66–73
Non-improvement after 4 weeks, including CGI-I	0.78	78	70–83	71	63–77	88	83–91	53	46–59

Abbreviations: AUC, area under the curve; CGI-I, Clinical Global Impression Scale – Improvement; CI, 95% confidence interval; NPV, negative predictive value; PPV, positive predictive value.

Table 3. Subgroup analysis for sex (bracketed numbers are 95% confidence intervals)

Subgroup	Partial AUC	Sensitivity	Specificity	PPV	NPV
Male	83	81 (77–84)	73 (66–79)	89 (85–93)	60 (55–66)
Female	70	74 (66–80)	68 (61–74)	82 (77–86)	56 (51–62)

Abbreviations: AUC, area under the curve; NPV, negative predictive value; PPV, positive predictive value.

(83–89), and NPV of 59% (55–62). After excluding patients who dropped out before week 10, AUC was 0.81, sensitivity was 75% (68–80), specificity was 75% (68–80), NPV was 59% (54–63), and PPV was 86% (83–89). Excluding patients who were assigned to clomipramine yielded the following accuracy: AUC of 0.79, sensitivity of 79% (74–83), specificity of 68 (62–73), PPV of 86% (82–88), and NPV of 57% (53–68).

Discussion

In this diagnostic test accuracy meta-analysis of pharmacotherapy in adults with OCD, we aimed to investigate whether early non-improvement (<25% YBOCS reduction after 4 weeks) predicts subsequent antidepressant nonresponse (<35% YBOCS reduction at trial endpoint). We found that early non-improvement predicts subsequent nonresponse with a PPV of 86% (95% CI = 83–88%), meaning that after 1 month of treatment, out of every 100 patients who do not show at least minimal symptom improvement, 86 will not respond at the end of the study. Thus, early non-improvers are likely to become eventual nonresponders to that particular drug.

A lower NPV, of 60%, illustrates that if patients do show improvement at 4 weeks, this does not guarantee response. A sensitivity of 78% indicates that, of the 100 patients who ultimately did not respond to treatment, 78 were correctly identified as non-improvers after 4 weeks. In other words, non-improvement effectively captures most of the patients who will not go on to respond. Conversely, a specificity of 70% means that out of 100 patients who did respond by the end of the trial, 70 were correctly identified as showing early improvement at 4 weeks. Thus, some responders may still be incorrectly predicted to be nonresponders.

In clinical practice, this means that a lack of symptomatic improvement after 4 weeks of antidepressant treatment could serve as an early indicator that a patient is less likely to benefit from continued treatment. In such cases, early reevaluation could be considered within a framework of shared decision-making. At the same time, overall predictive accuracy further improves at 6 weeks rather than 4 weeks, and is highest when nonresponse at 8 weeks is used as the index timepoint. These findings suggest that the timing of treatment decisions should be individualized and guided by the clinical context and patient preferences.

Accuracy was significantly higher in male compared to female patients at 4 weeks, but not at 6 weeks. This may seem surprising as a sex difference has not been found in earlier diagnostic accuracy studies in OCD, or meta-analyses in other psychiatric classifications (de Vries et al., 2019; Samara et al., 2015). Sex has not been identified as a modifier or predictor of treatment outcome in OCD (Cohen et al., 2025; Mathes, Morabito, & Schmidt, 2019). Differences in pharmacokinetics between men and women, where female patients may have a relatively high exposure to an antidepressant dose, do not provide a clear explanation, as this may mean that women receive higher serum levels sooner, and so the clinical effect should be more pronounced at 4 weeks (Kokras, Dalla, & Papadopoulou-Daifoti, 2011). These findings may suggest that the timing of treatment response differs between male and female patients. Predictive accuracy did not differ significantly between compounds, even though the pharmacokinetics of compounds do vary, especially in fluoxetine, which has a relatively long half-life (Hiemke & Härtter, 2000). Similarly, age and illness severity do not affect accuracy. It should be noted that these meta-regression findings are exploratory, since we did not use internal or external validation for linear regression.

In disorders other than OCD, secondary analyses of short-term trials and meta-analyses have found that a lack of early response can predict later treatment success (Welten et al., 2016). In schizophrenia, for instance, a meta-analysis of diagnostic test accuracy studies found that nonresponse to antipsychotic medication after 2 weeks predicted unsuccessful treatment with a PPV of 90% (Samara et al., 2015). The diagnostic accuracy of non-improvement in OCD is thus somewhat lower than in schizophrenia, but it is comparable to the diagnostic accuracy in antimanic medication in bipolar disorder and higher than in pharmacotherapeutic treatment of MDD (de Vries et al., 2019; Samara et al., 2015; Welten et al., 2016).

Our results contrast the commonly held assumption that response to antidepressants is delayed in OCD, and that, consequently, a minimum of 12 weeks of treatment is necessary before judging an antidepressant trial as unsuccessful (Paxos, 2022). Our results add to findings from an aggregate data meta-analysis, which analyzed outcomes at interim time points of OCD trials in order to examine the trajectory of medication response in OCD (Issari et al., 2016). The authors found a significant difference compared to placebo after only 2 weeks of treatment. On the group level, 75% of symptom improvement due to SSRIs

compared to placebo is established within the first 6 weeks. Our study shows that this effect is not only apparent on group-level data but also that early evaluation of response can influence clinical decision-making on an individual patient level. These findings are corroborated by smaller, explorative predictive studies in OCD (Ackerman et al., 1996; da Conceição Costa et al., 2013).

Clinical relevance of response prediction does not solely depend on prediction accuracy, but also on a priori chance of successful treatment, and on the presence of possible treatment alternatives. For instance, in our study sample, the a priori chance of nonresponse at the end of the trial is 63%. After a positive index test, this chance is 86%, which means a more than 60% increase in the chance of nonresponse ($[(37-14\%)/37\%]$). It should be added that a high a priori chance of nonresponse increases the PPV due to the dependence of predictive values on the prevalence of the reference test. As such, in a population with a lower chance of nonresponse, the PPV might decrease. In the pharmacological treatment of MDD, research has shown that the chance of response decreases with each subsequent medication step (Rush et al., 2004). To our knowledge, in OCD, there is only one double-blind study exploring switching in case of nonresponse, finding a favorable effect of switching with a response rate of 50% (Denys, van Meegen, van der Wee, & Westenberg, 2004). However, this study is relatively small and considers switching from SSRI to serotonin–norepinephrine reuptake inhibitors and vice versa in case of initial nonresponse, without a non-switching control group. In other words, we cannot definitely state that *switching* medications increases the chance of response, compared to continuing the current medication. The same can be stated for the addition of psychotherapy, pharmacological augmentation strategies, or further dosage escalation (Levy, Arush, Carmi, Wetzler, & Zohar, 2024). This emphasizes the need for *switch studies* in OCD treatments (Wang et al., 2022). Yet, such studies remain scarce, partly due to a lack of public funding and the dominance of industry-sponsored trials, which limits variability in design and may affect generalizability (Ioannidis et al., 2025).

Our study has significant strengths, including the use of a unique dataset from multiple large published and unpublished clinical trials, to conduct a robust prediction meta-analysis, with a uniform analysis across trials according to a preregistered protocol. We employed a priori defined index and reference tests, thus avoiding post-hoc adjustments or data-driven modifications that can inflate predictive accuracy estimates.

However, there are several limitations. First, placebo-controlled settings may not accurately reflect real-world clinical practice because the presence of a placebo arm can alter treatment expectations, potentially causing hesitancy among patients and assessors in reporting improvements (Salanti et al., 2018). This may explain our conservative response rates (37%) compared to open-label studies (Brar et al., 2022; Tükel, Bozkurt, Polat, Genç, & Atli, 2006). Real-life settings could either heighten sensitivity to improvement, due to established patient–doctor relationships, or reduce careful symptom tracking due to clinical time constraints and broader treatment goals. Second, patients might not experience the full therapeutic dose throughout the entire 12-week study period, suggesting additional improvement beyond 12 weeks remains possible. Although early responses correlate with long-term outcomes, evidence remains mixed regarding further symptom reduction post-12 weeks. Short-term treatment response is strongly correlated with long-term outcomes, yet it remains unclear whether and to what extent patients with OCD continue to improve beyond the 12-week mark (Bloch et al., 2013; Greist et al., 1995). Third, although all included studies

reached therapeutic doses by 2 weeks, the initial dose titration implies that the effective duration at an adequate dose by week 4 may be shorter. Mitigating this, we have included a secondary outcome at 6 and 8 weeks, in order to give a more nuanced view of improvement trajectories. Fourth, our study included three finding studies with multiple fixed-dose arms. Although all patients who were included used a dose within the registered dose range according to the summary of product characteristics for each compound, differences in dose may still have impacted the results. Finally, our study focused solely on symptomatic improvement and did not encompass functional or quality-of-life scales.

Despite these limitations, this study challenges the current notion that in OCD, the success of a medication trial can only be assessed after 10–12 weeks of treatment. While initial improvement does not guarantee subsequent response, we showed that non-improvement after 4 weeks reliably predicts subsequent nonresponse. This work advances the integration of early stop criteria into shared decision-making, which could decrease the exposure and burden of treatments that are unlikely to help patients with OCD.

Supplementary material. The supplementary material for this article can be found at <http://doi.org/10.1017/S0033291725101335>.

Data availability statement. The data are proprietary and may not be shared outside the Medical Evaluation Board.

Funding statement. This study was not sponsored.

Competing interests. The authors declare none.

References

- Abo-Zaid, G., Guo, B., Deeks, J. J., Debray, T. P., Steyerberg, E. W., Moons, K. G., & Riley, R. D. (2013). Individual participant data meta-analyses should not ignore clustering. *Journal of Clinical Epidemiology*, *66*(8), 865–873. <https://doi.org/10.1016/j.jclinepi.2012.12.017>.
- Ackerman, D. L., Greenland, S., & Bystritsky, A. (1996). Use of receiver-operator characteristic (ROC) curve analysis to evaluate predictors of response to clomipramine therapy. *Paper presented at the Psychopharmacology Bulletin*, *32*(1), 157–65. PMID: 8927667.
- Bloch, M. H., Green, C., Kichuk, S. A., Dombrowski, P. A., Wasylink, S., Billingslea, E., ... Pittenger, C. (2013). Long-term outcome in adults with obsessive-compulsive disorder. *Depression and Anxiety*, *30*(8), 716–722. <https://doi.org/10.1002/da.22103>.
- Brar, J., Sidana, A., Chauhan, N., & Bajaj, M. K. (2022). Early improvement as a predictor of treatment response in patients with obsessive-compulsive disorder: A 12-week randomized trial of sertraline and fluvoxamine. *Prim Care Companion CNS Disord*, *24*(2), 21m03065. <https://doi.org/10.4088/PCC.21m03065>.
- Burke, D. L., Ensor, J., & Riley, R. D. (2017). Meta-analysis using individual participant data: One-stage and two-stage approaches, and why they may differ. *Statistics in Medicine*, *36*(5), 855–875. <https://doi.org/10.1002/sim.7141>.
- Busner, J., & Targum, S. D. (2007). The clinical global impressions scale: Applying a research tool in clinical practice. *Psychiatry (Edgmont)*, *4*(7), 28–37.
- Carolyn Rodriguez, M., Bloch M, Sachs R, & Williams M. (2015). Anxiety & depression Association of America, OCD practice guideline. Retrieved February 7 2025, from <https://adaa.org/resources-professionals/practice-guide-lines-ocd>
- Cervin, M., Consortium, O. C. D. S. B., & Mataix-Cols, D. (2022). Empirical severity benchmarks for obsessive-compulsive disorder across the lifespan. *World Psychiatry*, *21*(2), 315–316. <https://doi.org/10.1002/wps.20984>.
- Cervin, M., Storch, E. A., Farrell, L., Andersson, E., Rück, C., Geller, D. A., & Mataix-Cols, D. (2025). Minimal clinically important difference in obsessive-compulsive disorder. *World Psychiatry*, *24*(1), 138–139. <https://doi.org/10.1002/wps.21287>.
- Cohen, S. E., Storosum, B. W., Zantvoord, J. B., Mattila, T. K., de Boer, A., & Denys, D. (2025). Individual patient data meta-analysis of placebo-controlled trials of selective serotonin reuptake inhibitors submitted for regulatory approval in adult obsessive-compulsive disorder. *The British Journal of Psychiatry*, 2025 May 15:1–8. <https://doi.org/10.1192/bjp.2025.87>.
- Cohen, S. E., Zantvoord, J. B., Mattila, T. K., Storosum, B. W. C., de Boer, A., & Denys, D. (2024). The minimal important difference in obsessive-compulsive disorder: An analysis of double-blind SSRI trials in adults. *European Psychiatry*, *67*(1), e53. <https://doi.org/10.1192/j.eurpsy.2024.1768>.
- da Conceição Costa, D. L., Shavitt, R. G., Castro Cesar, R. C., Joaquim, M. A., Borcato, S., Valério, C., ... Diniz, J. B. (2013). Can early improvement be an indicator of treatment response in obsessive-compulsive disorder? Implications for early-treatment decision-making. *Journal of Psychiatric Research*, *47*(11), 1700–1707. <https://doi.org/10.1016/j.jpsychires.2013.07.006>.
- de Vries, Y. A., Roest, A. M., Bos, E. H., Burgerhof, J. G. M., van Loo, H. M., & de Jonge, P. (2019). Predicting antidepressant response by monitoring early improvement of individual symptoms of depression: Individual patient data meta-analysis. *British Journal of Psychiatry*, *214*(1), 4–10. <https://doi.org/10.1192/bjp.2018.122>.
- Denys, D., van Megen, H. J., van der Wee, N., & Westenberg, H. G. (2004). A double-blind switch study of paroxetine and venlafaxine in obsessive-compulsive disorder. *The Journal of clinical psychiatry*, *65*(1), 37–43. <https://doi.org/10.4088/jcp.v65n0106>.
- Fineberg, N. A., Brown, A., Reghunandanan, S., & Pampaloni, I. (2012). Evidence-based pharmacotherapy of obsessive-compulsive disorder. *The International Journal of Neuropsychopharmacology*, *15*(8), 1173–1191. <https://doi.org/10.1017/s1461145711001829>.
- Greist, J. H., Jefferson, J. W., Kobak, K. A., Chouinard, G., DuBoff, E., Halaris, A., et al. (1995). A 1 year double-blind placebo-controlled fixed dose study of sertraline in the treatment of obsessive-compulsive disorder. *International Clinical Psychopharmacology*, *10*(2), 57–65. <https://doi.org/10.1097/00004850-199506000-00001>.
- Hiemke, C., & Härtter, S. (2000). Pharmacokinetics of selective serotonin reuptake inhibitors. *Pharmacology & Therapeutics*, *85*(1), 11–28. [https://doi.org/10.1016/S0163-7258\(99\)00048-0](https://doi.org/10.1016/S0163-7258(99)00048-0).
- Ioannidis, K., Huneke, N. T. M., Solly, J. E., Fusetto Veronesi, G., Tzagarakis, C., Parlatini, V., ... Chamberlain, S. R. (2025). Placebo and nocebo effects in gambling disorder pharmacological trials: A meta-analysis. *Acta Neuropsychiatrica*, *37*, e40. <https://doi.org/10.1017/neu.2024.52>.
- Issari, Y., Jakubovski, E., Bartley, C. A., Pittenger, C., & Bloch, M. H. (2016). Early onset of response with selective serotonin reuptake inhibitors in obsessive-compulsive disorder: A meta-analysis. *The Journal of Clinical Psychiatry*, *77*(5), e605–e611. <https://doi.org/10.4088/JCP.14r09758>.
- Kokras, N., Dalla, C., & Papadopoulou-Daifoti, Z. (2011). Sex differences in pharmacokinetics of antidepressants. *Expert Opinion on Drug Metabolism & Toxicology*, *7*(2), 213–226. <https://doi.org/10.1517/17425255.2011.544250>.
- Koran, L. M., Hanna, G. L., Hollander, E., Nestadt, G., & Simpson, H. B. (2007). Practice guideline for the treatment of patients with obsessive-compulsive disorder. *The American Journal of Psychiatry*, *164*(7 Suppl), 5–53.
- Krompinger, J. W., Monaghan, S. C., Gironde, C. M., Garner, L. E., Crosby, J. M., Brennan, B. P., & Elias, J. A. (2017). Early response is predictive of outcome in intensive behavioral treatment for obsessive compulsive disorder. *Journal of Obsessive-Compulsive and Related Disorders*, *15*, 57–63. <https://doi.org/10.1016/j.jocrd.2017.10.001>.
- Leeflang, M. M. G., Deeks, J. J., Rutjes, A. W. S., Reitsma, J. B., & Bossuyt, P. M. M. (2012). Bivariate meta-analysis of predictive values of diagnostic tests can be an alternative to bivariate meta-analysis of sensitivity and specificity. *Journal of Clinical Epidemiology*, *65*(10), 1088–1097. <https://doi.org/10.1016/j.jclinepi.2012.03.006>.
- Levy, D. M., Arush, O. B., Carmi, L., Wetzler, A. J., & Zohar, J. (2024). Off-label higher doses of serotonin reuptake inhibitors in the treatment of obsessive-compulsive disorder: Safety and tolerability. *Comprehensive Psychiatry*, *133*, 152486. <https://doi.org/10.1016/j.comppsy.2024.152486>.
- Mataix-Cols, D., Andersson, E., Aspvall, K., Boberg, J., Crowley, J. J., de Schipper, E., ... Cervin, M. (2022). Operational definitions of treatment response and remission in obsessive-compulsive disorder capture meaningful improvements in everyday life. *Psychotherapy and Psychosomatics*, *91*(6), 424–430. <https://doi.org/10.1159/000527115>.

- Mathes, B. M., Morabito, D. M., & Schmidt, N. B. (2019). Epidemiological and clinical gender differences in OCD. *Current Psychiatry Reports*, **21**(5), 36. <https://doi.org/10.1007/s11920-019-1015-2>.
- McInnes, M. D. F., Moher, D., Thoms, B. D., McGrath, T. A., Bossuyt, P. M., Clifford, T., ... Willis, B. H. (2018). Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: The PRISMA-DTA statement. *JAMA*, **319**(4), 388–396. <https://doi.org/10.1001/jama.2017.19163>.
- Nakatani, E., Nakagawa, A., Nakao, T., Yoshizato, C., Nabeyama, M., Kudo, A., ... Kawamoto, M. (2005). A randomized controlled trial of Japanese patients with obsessive-compulsive disorder—effectiveness of behavior therapy and fluvoxamine. *Psychotherapy and Psychosomatics*, **74**(5), 269–276. <https://doi.org/10.1159/000086317>.
- Paxos, C. (2022). Moving beyond first-line treatment options for OCD. *Ment Health Clinician*, **12**(5), 300–308. <https://doi.org/10.9740/mhc.2022.10.300>.
- Pittenger, C., & Bloch, M. H. (2014). Pharmacological treatment of obsessive-compulsive disorder. *The Psychiatric Clinics of North America*, **37**(3), 375–391. <https://doi.org/10.1016/j.psc.2014.05.006>.
- Ramakrishnan, D., Farhat, L. C., Vattimo, E. F. Q., Levine, J. L. S., Johnson, J. A., Artukoglu, B. B., ... Bloch, M. H. (2024). An evaluation of treatment response and remission definitions in adult obsessive-compulsive disorder: A systematic review and individual-patient data meta-analysis. *Journal of Psychiatric Research*, **173**, 387–397. <https://doi.org/10.1016/j.jpsychires.2024.03.044>.
- Reitsma, J. B., Glas, A. S., Rutjes, A. W. S., Scholten, R. J. P. M., Bossuyt, P. M., & Zwinderman, A. H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology*, **58**(10), 982–990. <https://doi.org/10.1016/j.jclinepi.2005.02.022>.
- Rush, A. J., Fava, M., Wisniewski, S. R., Lavori, P. W., Trivedi, M. H., Sackeim, H. A., ... Niederehe, G. (2004). Sequenced treatment alternatives to relieve depression (STAR*D): Rationale and design. *Controlled Clinical Trials*, **25**(1), 119–142. [https://doi.org/10.1016/s0197-2456\(03\)00112-0](https://doi.org/10.1016/s0197-2456(03)00112-0).
- Salanti, G., Chaimani, A., Furukawa, T. A., Higgins, J. P. T., Ogawa, Y., Cipriani, A., & Egger, M. (2018). Impact of placebo arms on outcomes in antidepressant trials: Systematic review and meta-regression analysis. *International Journal of Epidemiology*, **47**(5), 1454–1464. <https://doi.org/10.1093/ije/dyy076>.
- Samara, M. T., Leucht, C., Leeflang, M. M., Anghelescu, I. G., Chung, Y. C., Crespo-Facorro, B., ... Leucht, S. (2015). Early improvement as a predictor of later response to antipsychotics in schizophrenia: A diagnostic test review. *The American Journal of Psychiatry*, **172**(7), 617–629. <https://doi.org/10.1176/appi.ajp.2015.14101329>.
- TÜKEL, R., BOZKURT, O., POLAT, A., GENÇ, A., & ATLLI, H. (2006). Clinical predictors of response to pharmacotherapy with selective serotonin reuptake inhibitors in obsessive-compulsive disorder. *Psychiatry and Clinical Neurosciences*, **60**(4), 404–409. <https://doi.org/10.1111/j.1440-1819.2005.01451.x-i1>.
- Wang, P., Gu, W., Gao, J., Wang, C., Fang, J., Hu, M., ... Wang, Z. (2022). Protocol for a pragmatic trial of pharmacotherapy options following unsatisfactory initial treatment in OCD (PROCEED). *Frontiers in Psychiatry*, **13**, 822976. <https://doi.org/10.3389/fpsy.2022.822976>.
- Welten, C. C., Koeter, M. W., Wohlfarth, T. D., Storosum, J. G., van den Brink, W., Gispens-de Wied, C. C., ... Denys, D. A. (2016). Early nonresponse in the antipsychotic treatment of acute mania: A criterion for reconsidering treatment? Results from an individual patient data meta-analysis. *The Journal of Clinical Psychiatry*, **77**(9), e1117–e1123. <https://doi.org/10.4088/JCP.15r10051>.
- Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., ... Bossuyt, P. M. (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, **155**(8), 529–536. <https://doi.org/10.7326/0003-4819-155-8-201110180-00009>.