Short Paper

Heterogeneity in Effects of Automated Results Feedback After Online Depression Screening: Secondary Machine-Learning Based Analysis of the DISCOVER Trial

Matthias Klee[1], PhD; Byron C Jaeger[2], PhD; Franziska Sikorski[3], PhD; Bernd Löwe[3], MD; Sebastian Kohlmann[1,3], PhD

[1]Department of General Internal Medicine and Psychosomatics, Heidelberg University, Heidelberg, Germany
[2]Department of Biostatistics and Data Science, Wake Forest University School of Medicine, Winston-Salem, NC, United States
[3]Department of Psychosomatic Medicine & Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Corresponding Author:

Sebastian Kohlmann, PhD
Department of General Internal Medicine and Psychosomatics
Heidelberg University
Im Neuenheimer Feld 410
Heidelberg 69120
Germany
Phone: 49 62215632879
Email: sebastian.kohlmann@med.uni-heidelberg.de

Abstract

Background: Online depression screening tools may increase uptake of evidence-based care and consequently lead to symptom reduction. However, results of the DISCOVER trial suggested no effect of automated results feedback compared with no feedback after online depression screening on depressive symptom reduction six months after screening. Interpersonal variation in symptom representation, health care needs, and treatment preferences may nonetheless have led to differential response to feedback mode on an individual level.

Objective: The aim of this study was to examine heterogeneity of treatment effects (HTE), that is, differential responses to two feedback modes (tailored or nontailored) versus no feedback (control) following online depression screening.

Methods: We used causal forests, a machine learning method that applies recursive partitioning to estimate conditional average treatment effects (CATEs). In this secondary data analysis of the DISCOVER trial, eligible participants screened positive for at least moderate depression severity but had not been diagnosed or treated for depression in the preceding year. The primary outcome was heterogeneity in depression severity change, over a and six-month follow up period, measured with the Patient Health Questionnaire-9. Analysis comprised exploration of average treatment effects (ATE), HTE, operationalized with the area under the targeting operator characteristic curve (AUTOC), and differences in ATE when allocating feedback based on predicted CATE. We extracted top predictors of depression severity change, given feedback and explored high-CATE covariate profiles. Prior to analysis, data was split into training and test sets (1:1) to minimize the risk of overfitting and evaluate predictions in held-out test data.

Results: Data from 946 participants of the DISCOVER trial without missing data were analyzed. We did not detect HTE; no versus nontailored feedback, AUTOC –0.48 (95% CI –1.62 to 0.67, P=.41); no versus tailored feedback, AUTOC 0.06 (95% CI –1.21 to 1.33, P=.93); and no versus any feedback, AUTOC –0.20 (95% CI –1.30 to 0.89, P=.72). There was no evidence of alteration to the ATE in the test set when allocating feedback (tailored or nontailored) based on the predicted CATE. By examining covariate profiles, we observed a potentially detrimental role of control beliefs, given feedback compared with no feedback.

Conclusions: We applied causal forests to describe higher-level interactions among a broad range of predictors to detect HTE. In absence of evidence for HTE, treatment prioritization based on trained models did not improve ATEs. We did not find evidence of harm or benefit from providing tailored or nontailored feedback after online depression screening regarding depression severity change after six months. Future studies may test whether screening alone prompts behavioral activation and downstream depression severity reduction, considering the observed uniform changes across groups.

Trial Registration: ClinicalTrials.gov NCT04633096; https://clinicaltrials.gov/study/NCT04633096

International Registered Report Identifier (IRRID): RR2-10.1016/j.invent.2021.100435

JMIR AI 2025;4:e70001; doi: 10.2196/70001

Keywords: heterogeneity of treatment effects; treatment heterogeneity; causal random forest; random forest; depression; depression screening; DISCOVER; feedback

Introduction

Online depression screening may foster early identification of individuals with depressive symptoms and reinforce help-seeking behavior [1,2]. However, recently published results of the DISCOVER trial (ClinicalTrials.gov, NCT04633096) suggested no effect of automated results feedback compared with no feedback after online depression screening on depressive symptom reduction six months after screening [1].

Randomized controlled trials (RCT) are the gold-standard method for evaluating intervention efficacy. Still, evidence-based medicine incorporates the notion of differential responses to interventions on a person-level, potentially masking harms or benefits at the group-level [3]. Thus, it is important to examine individual-level responses to feedback.

Machine-learning (ML) methods have previously been adapted to examine such heterogeneity of treatment effects (HTE) [4-6]. ML can circumvent the issue of multiple testing through cross-validation and by design, account for higher level interactions [7]. ML approaches to estimate HTE have been successfully applied in health care literature, especially in cardiology and psychiatry [5,8].

The aim of this paper was to investigate the presence of HTE in response to feedback (no feedback ie, control group vs nontailored or tailored, ie, intervention), following online depression screening. We tested for the presence of HTE, based on person-level characteristics at baseline, with heterogeneity in depression severity change at six months as the primary outcome. The efficacy of allocating feedback to individuals with more favorable predicted conditional treatment response was examined in exploratory analysis.

Methods

Study Sample and Design

DISCOVER is a three-armed RCT examining change in depression severity six months after online screening with tailored, nontailored, or no feedback (control). The study was advertised as a study on stress and psychological well-being [1]. Recruitment strategies involved print and online advertisements on social media platforms and in a nationwide online access survey panel in Germany [1]. Eligible participants were 18 years or older, proficient in German, and screened positive for at least moderate depression severity (Patient Health Questionnaire-9, PHQ-9 ≥10) [1]. Participants with missing data, or those with a diagnosis of or treatment for depression in the previous year were excluded.

Feedback comprised depression screening results, a recommendation to consult a mental health care professional or general practitioner, and further information regarding depression and related treatment based on national guidelines [1,9]. For the tailored feedback group, feedback was adapted according to participants' symptom profiles, treatment preferences, and available guideline-recommended options [1,9].

Ethical Considerations

Review and approval was provided by the Ethics Committee of the Hamburg Medical Chamber (PV7039) [9]. Participants provided online informed consent covering secondary data analyses [1]. Participants received a € 5 (US $5.85) voucher for compensation upon each completed follow-up. Data was deidentified prior to analysis.

Main Outcomes and Measures

The primary outcome was heterogeneity in depression severity change six months after online screening. Depression severity was measured with the PHQ-9 [10]. Predictors involved baseline depression (PHQ-9), anxiety (Generalized Anxiety Disorder Scale-7; GAD-7 [11]), and somatic symptom severity (Somatic Symptom Scale-8; SSS-8 [12]), health-related quality of life (EuroQoL-5 Dimension-5 Level visual analogue scale [EQ-5D-5 L VAS]) [13], illness beliefs (Brief Illness Perception Questionnaire; B-IPQ [14,15]), patient history, depression-related risk factors, and sociodemographic characteristics (Table S1 in Multimedia Appendix 1).

Statistical Analysis

Causal Forests (CF) [5,6] have previously been used to investigate HTE [8] (Box S1 in Multimedia Appendix 1). CFs estimate conditional average treatment effects (CATE), which approximate individual-level treatment effects (ITE). ITE cannot be inferred directly since only one potential outcome is realized per participant. Thus, CATE are more granular than average treatment effects (ATE) but less granular than ITE.

Two CFs were trained based on either training (tau-forest) or test data (eval-forest), with a random split (1:1). CFs were trained with 2000 trees, a sample fraction of 0.5, a minimum node size of 5, and mtry = 30. ATE and CATE were computed by contrasting intervention (nontailored [1], tailored [2], or any [1/2] feedback) to no feedback (control) with depression severity change as the outcome and a propensity score for treatment allocation (P=.50) [4,5].

To assess the presence of HTE discretely, tau-forest predictions of CATE for the test data were grouped into quartiles. Then, ATE was estimated in each quartile using the eval-forest. To assess the presence of HTE continuously, we computed the area under the targeting operator characteristic curve (AUTOC) and tested for the presence of HTE with a significance test for AUTOC (H_1:AUTOC≠0) [4]. Significance of AUTOC was tested two-sided, with bootstrapped standard errors (n=200 bootstrap replicates).

For a comprehensive overview of model evaluation and sample code, see Box S2 in Multimedia Appendix 1, Sverdrup, Petukhova [4] and Klee [16]. Analyses were conducted using R (version 4.3.1; R Foundation for Statistical Computing) using the *grf* package [17].

Results

Baseline Characteristics

After visual inspection of missingness patterns, 19 participants were removed, assuming missingness at random (Table S2 in Multimedia Appendix 1). In total, 946 participants were eligible for analysis. Participants were aged 18 to 79 years, mean 37.20 (SD 13.98) (Table 1).

Table 1. Baseline characteristics of participants in the analytic data set (N=946).

Characteristic	No feedback (n=318)	Nontailored feedback (n=313)	*P* value[a]	Tailored feedback (n=315)	*P* value[a]
Age in years, mean (SD)	36.4 (13.7)	38.2 (13.8)	.11	37.0 (14.4)	.62
Gender, n (%)			.85[b]		.73[b]
Women	232 (73.0)	222 (70.9)	—	221 (70.2)	—
Men	83 (26.1)	88 (28.1)	—	91 (28.9)	—
Other	3 (0.9)	3 (1.0)	—	3 (1.0)	—
Education, n (%)			.66		.55
<10 years	55 (17.3)	59 (18.8)	—	47 (14.9)	—
≥10 years	94 (29.6)	99 (31.6)	—	104 (33.0)	—
University entrance qualification	169 (53.1)	155 (49.5)	—	164 (52.1)	—
Depression severity (PHQ-9[c]), mean (SD)	14.8 (4.03)	14.7 (4.09)	.78	14.8 (3.82)	.94
Somatic symptom severity (SSS-8[d]), mean (SD)	14.6 (5.23)	14.5 (5.13)	.90	14.3 (5.32)	.57
Anxiety severity (GAD-7[e]), mean (SD)	12.0 (4.32)	12.5 (4.23)	.19	12.0 (4.29)	.94

[a] *P* values for pairwise comparisons with the ‘no feedback’ group. Continuous characteristics were compared with Student’s *t*-test, categorical characteristics were compared with χ^2 tests.

[b] χ^2 approximation may be incorrect due to small cell size. Analysis based on men and women only replicated findings (*P*=.63 for nontailored feedback vs no feedback, *P*=.48 for tailored feedback vs no feedback).

[c] PHQ-9=Patient Health Questionnaire-9.

[d] SSS-8=Somatic Symptom Scale-8.

[e] GAD-7=Generalized Anxiety Disorder Scale-7.

Average Treatment Effect

We did not find evidence of non-zero ATE in either tau- or eval-forests, suggesting no benefit of providing any form of feedback compared with the control (no feedback) (Table 2).

Table 2. Average treatment effects for pairwise comparison of feedback conditions.

Comparison	Tau-forest[a]	Eval-forest[b]
	ATE[c] (SE)	ATE (SE)
No feedback versus nontailored feedback	0.04 (0.54)	–0.24 (0.55)
No feedback versus tailored feedback	–0.38 (0.57)	–0.16 (0.57)
No feedback versus any feedback	0.07 (0.48)	–0.48 (0.49)

[a] Tau-forest is based on training data.

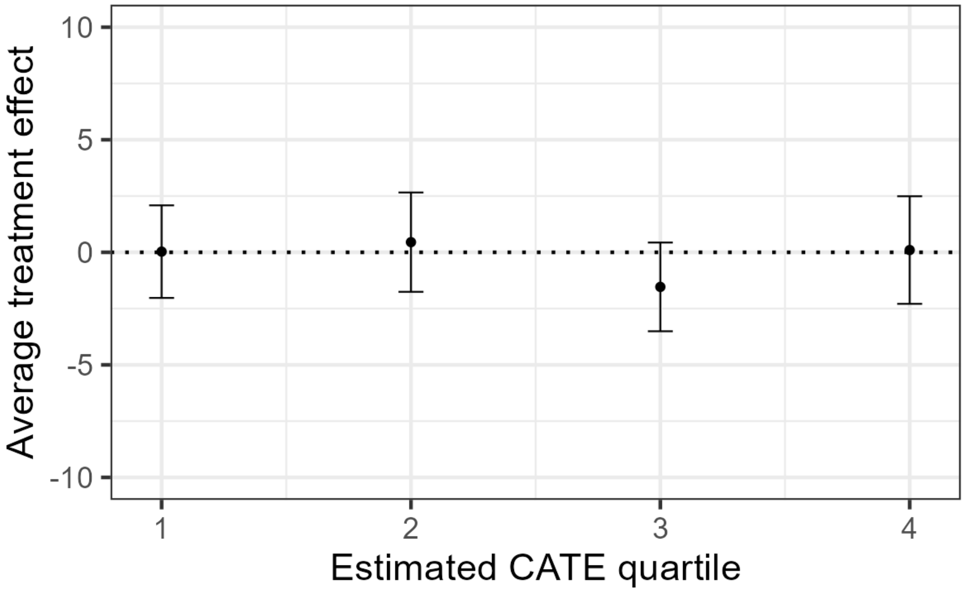
[b] Eval-forest is based on test data.

[c] ATE: Average treatment effect.

Heterogeneity in Treatment Effects

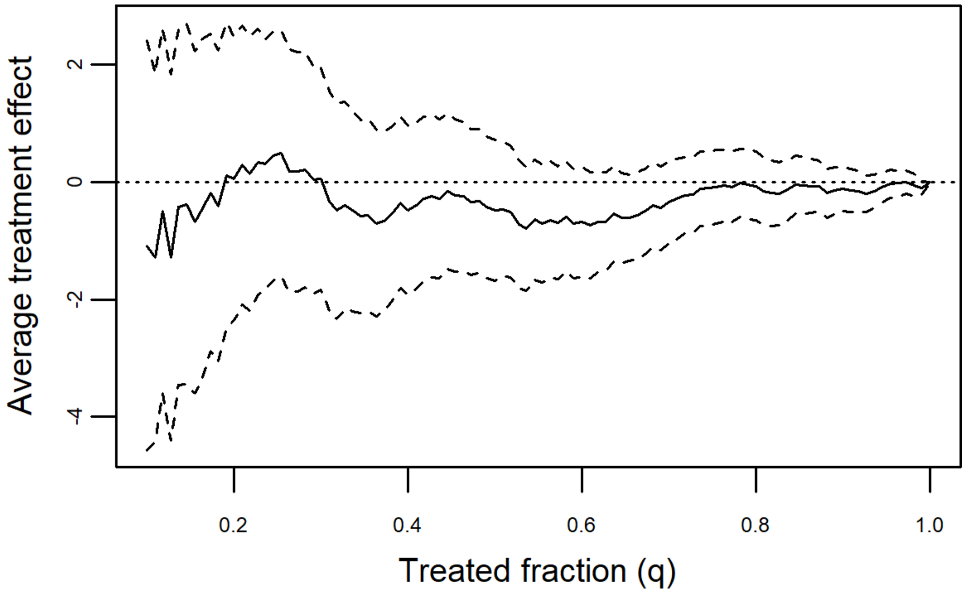
There was no evidence of HTE when comparing nontailored (Figure 1) or any feedback with control (Figure S4 in Multimedia Appendix 1). However, there was a lower (ie, more favorable) ATE when comparing tailored feedback with control among participants with predicted CATE in the second most favorable quartile regarding depression severity change (Figure S1 in Multimedia Appendix 1).

Figure 1. Average treatment effects in participant groups reflecting quartiles of predicted CATE from lowest (L) to highest (R). CATE was predicted in test data with the tau-forest. ATE was estimated within quartiles with the evaluation forest (based on test data). Positive values indicate less favorable ATE. ATE: average treatment effect; CATE: conditional average treatment effect.



AUTOC estimates (Figures S2 and S5 in Multimedia Appendix 1) did not suggest the presence of HTE in any comparison: AUTOC −0.48 95% CI −1.62 to 0.67, *P*=.41, nontailored; AUTOC 0.06, 95% CI −1.21 to 1.33, *P*=.93 tailored; AUTOC −0.20 (95% CI −1.30 to 0.89, *P*=.72 any feedback vs control). Allocating feedback based on predicted CATE did not substantially alter ATE (Figure 2). This is consistent with the near-zero AUTOC estimates, and indicates limited potential for altering the effects of feedback mode regarding depression severity change through targeted allocation.

Figure 2. Targeting operator characteristic curve plot. The dashed lines are pointwise 95% confidence intervals conditional on the estimated CATE function, (ie, the tau-forest based on the training data). The y-lab illustrates the benefit of providing feedback only to a fraction of participants based on their CATE (ie, treatment priority score), over treating everyone (difference in average treatment effects; ie, PHQ-9 change six months after screening). The x lab illustrates the fraction treated from highest (L) to lowest (R) CATE. Positive values indicate less favorable ATE. ATE: average treatment effect; CATE: conditional average treatment effect.



Top Predictors of Harm or Benefit From Treatment

The most important predictors of the tau-forest (comparing nontailored feedback with control) were items assessing depression–related treatment control belief (B-IPQ), trouble relaxing (GAD-7), anxiety severity (GAD-7) and trouble sleeping (SSS-8). For tailored feedback compared with control, items denoting age, somatic symptom (SSS-8) and anxiety (GAD-7) severity and depression-related treatment control belief (B-IPQ) were most important. For any feedback compared with control, illness beliefs (B-IPQ) were most important: treatment and personal control, concern, and emotional response.

Sensitivity Analyses

Higher treatment control (nontailored feedback vs control) and personal control beliefs (any feedback vs control) were the only predictors significantly associated with a linear approximation of CATE (Table 3 and Tables S3 and S4 in Multimedia Appendix 1). Both predictors were associated with less favorable CATE estimates, suggesting less favorable depression severity change at follow-up, given feedback.

Table 3. Best linear projection for top predictors of the causal forest with training data.

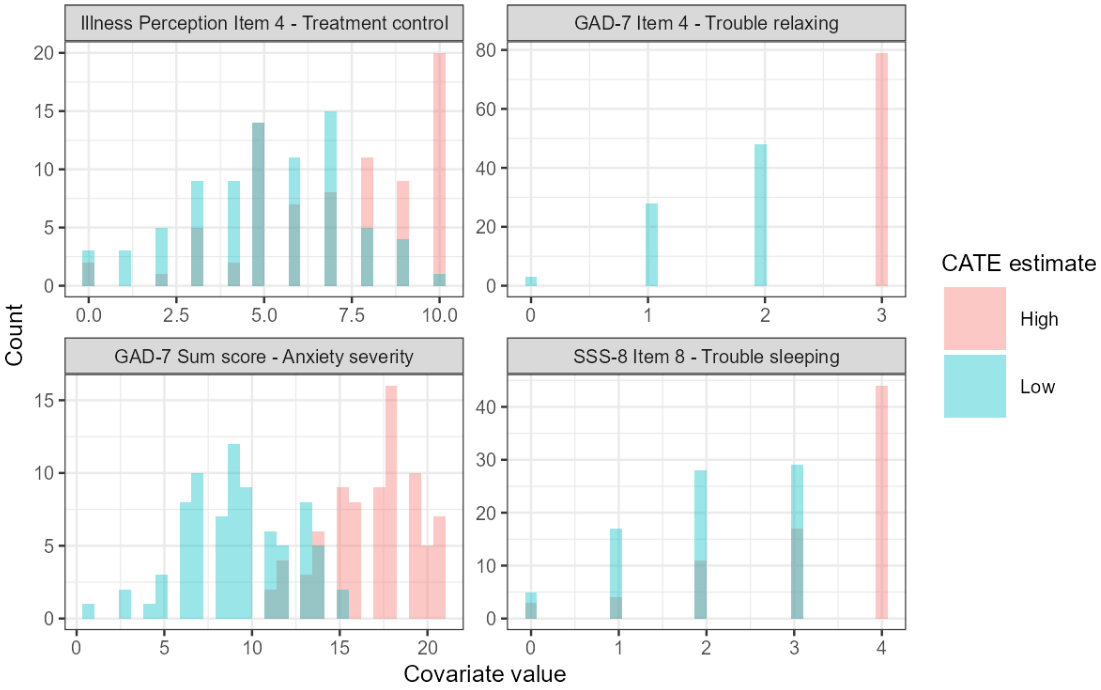
Term	Estimate	SE	*P* value
GAD-7[a] Item 4: How frequent did you feel impaired by the following symptoms during the past 2 weeks? – Trouble relaxing (0 not at all to 3 almost every day)	1.60	0.92	.08
Illness Perception Item 4: How much do you think a treatment can help with these complaints? (0 not at all to 10 extremely helpful)	0.47	0.21	.03
GAD-7[a] Sum score (0 to 21)	0.00	0.17	.99
SSS-8[b] Item 8: How strongly did you feel impaired by the following complaints during the past 7 days? – Trouble sleeping? (0 not at all to 4 very strongly)	0.61	0.50	.23

[a]GAD-7: Generalized Anxiety Disorder Scale-7.

[b]SSS-8: Somatic Symptom Scale-8.

Covariate profiles of the most important predictors are depicted in Figure 3 and Figures S3 and S6 in Multimedia Appendix 1. Overall, a higher treatment control belief was more frequent in the highest (least favorable) CATE quartile. Findings are less clear for remaining top four most important predictors.

Figure 3. Covariate profiles for test data with high (upper 25%, magenta) or low (lower 25%, cyan) predicted CATE based on tau-forest. CATE: conditional average treatment effect; GAD-7: Generalized Anxiety Disorder Scale-7; SSS-8: Somatic Symptom Scale-8.



Examining model calibration, CATE functions estimated in tau-forests did not significantly contribute to predicting change in depression severity at follow-up above group-level mean prediction (Table S5 in Multimedia Appendix 1).

Discussion

Principal Results

By applying CF to the DISCOVER online RCT, we did not find evidence for HTE with feedback (tailored, nontailored, or any) regarding change in depression severity six months after screening. As such, no apparent subgroup with an altered response to any type of feedback mode was detected.

Limitations

First, the selection of participants with at least moderate depression severity increases the likelihood of regression to the mean, which may have impeded the investigation of HTE. Second, generalizability of findings is limited to individuals participating in an online study about psychological well-being, who may exhibit distinct severity trajectories. Third, follow-up time may have been too short to detect variation in severity change, given, for example, the latency of help seeking.

Comparison With Prior Work

Our findings are in line with previous results showing no significant average benefit of any feedback mode for change in depression severity [1]. Beyond average effects, we investigated within-group harms and benefits in accordance with an evidence-based medicine approach [3]. We complement previous research suggesting no between-group detriments [18] with results suggesting that there are no latent subgroups that vary in their response to feedback. In contrast to previous research, our findings are valid irrespective of a priori defined categorical operationalizations of harmful or beneficial events. We show that included predictors do not alter response to feedback, providing an approximated assessment of individual-level harms and benefits [1,18].

Critically, sensitivity analyses suggested limited calibration of trained models. However, when testing HTE with CF, accurate mean prediction of the primary outcome is a first step necessary to detect deviations from it (ie, HTE). Previous research illustrates the notorious difficulty of predicting future depression courses, even with updated analytic tools [19-21]. Still, by employing a nonparametric method, we provide HTE estimates, that can account for a broad range of heterogeneity mechanisms potentially underlying depressive symptom trajectories and their variation following automated results feedback [8].

Conclusions

Applying CF, we could examine a broad range of predictors to detect HTE. In the absence of evidence for HTE, treatment prioritization based on trained models did not improve ATEs. We did not find evidence for harm or benefit of providing feedback after online depression screening regarding depression severity change after six months. Future studies may test if screening alone prompts behavioral activation and downstream depression severity reduction, considering the observed uniform changes across groups [22].

Acknowledgments

We acknowledge financial support from the Open Access Publication Fund of UKE - Universitätsklinikum Hamburg-Eppendorf. We thank all participants who gave their consent to participate in DISCOVER and supported the study with data. We would like to further thank Margarita Nikolaeva, who supported proofreading, translation and copyediting of the manuscript. None of the sponsors had a role in the study design, data collection, data analysis, data interpretation, or writing. We did not use generative AI in any portion of the manuscript writing.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization – MK, SK
Data curation – SK, FS
Formal analysis – MK, BJ
Funding acquisition – SK, BL
Investigation – SK, FS
Methodology – MK
Project administration –SK, BL
Supervision – SK
Validation – MK
Visualization – MK
Writing – original draft – MK, SK
Writing – review & editing – MK, BJ, FS, BL, SK

Conflicts of Interest

MK, BJ and FS declare no competing interests.

SK reports research funding (no personal honoraria) from the German Research Foundation and the German Federal Ministry of Education and Research.

BL reports research funding (no personal honoraria) from the German Research Foundation, the German Federal Ministry of Education and Research, the German Innovation Committee at the Joint Federal Committee, the European Commission's Horizon 2020 Framework Programme, the European Joint Programme for Rare Diseases (EJP), the Ministry of Science, Research and Equality of the Free and Hanseatic City of Hamburg, Germany, and the Foundation Psychosomatics of Spinal

Diseases, Stuttgart, Germany. He received remunerations for several scientific book articles from various book publishers, from the Norddeutscher Rundfunk (NDR) for interviews in medical knowledge programs on public television, and as a committee member from Aarhus University, Denmark. He received travel expenses from the European Association of Psychosomatic Medicine (EAPM), and accommodation and meals from the Societatea de Medicina Biopsyhosociala, Romania, for a presentation at the EAPM Academy at the Conferința Națională de Psihosomatică, Cluj-Napoca, Romania, October 2023. He received a travel grant for a lecture on the occasion of the presentation of the Alison Creed Award at the EAPM Conference in Lausanne, 12-15 June 2024. He received remuneration and travel expenses for lecture at the Lindauer Psychotherapiewochen, April 2024. He is President of the German College of Psychosomatic Medicine (DKPM) (unpaid) since March 2024 and was a member of the Board of the European Association of Psychosomatic Medicine (EAPM) (unpaid) until 2022. He is member of the EIFFEL Study Oversight Committee (unpaid).

Multimedia Appendix 1

Supplementary material containing a detailed list of included predictors, descriptive characteristics of participants with or without missing data, results for comparisons of no feedback with tailored or any feedback, sensitivity analyses of calibration and more detailed description of causal forests to detect heterogeneity of treatment effects, and the model evaluation strategy.
[DOCX File (Microsoft Word File), 278 KB-Multimedia Appendix 1]

References

1. Kohlmann S, Sikorski F, König HH, Schütt M, Zapf A, Löwe B. The efficacy of automated feedback after internet-based depression screening (DISCOVER): an observer-masked, three-armed, randomised controlled trial in Germany. Lancet Digit Health. Jul 2024;6(7):e446-e457. [doi: 10.1016/S2589-7500(24)00070-0] [Medline: 38906611]
2. Leventhal H, Phillips LA, Burns E. The Common-Sense Model of self-regulation (CSM): a dynamic framework for understanding illness self-management. J Behav Med. Dec 2016;39(6):935-946. [doi: 10.1007/s10865-016-9782-2] [Medline: 27515801]
3. Kravitz RL, Duan N, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. Milbank Q. 2004;82(4):661-687. [doi: 10.1111/j.0887-378X.2004.00327.x] [Medline: 15595946]
4. Sverdrup E, Petukhova M, Wager S. Estimating treatment effect heterogeneity in psychiatry: a review and tutorial with causal forests. Int J Methods Psychiatr Res. Jun 2025;34(2):e70015. [doi: 10.1002/mpr.70015] [Medline: 40178041]
5. Athey S, Tibshirani J, Wager S. Generalized random forests. Ann Statist. 2019;47(2):1148-1178. [doi: 10.1214/18-AOS1709]
6. Wager S, Athey S. Estimation and inference of heterogeneous treatment effects using random forests. J Am Stat Assoc. Jul 3, 2018;113(523):1228-1242. [doi: 10.1080/01621459.2017.1319839]
7. Breiman L. Random Forests. Mach Learn. Oct 2001;45(1):5-32. [doi: 10.1023/A:1010933404324]
8. Inoue K, Adomi M, Efthimiou O, et al. Machine learning approaches to evaluate heterogeneous treatment effects in randomized controlled trials: a scoping review. J Clin Epidemiol. Dec 2024;176(111538):111538. [doi: 10.1016/j.jclinepi.2024.111538] [Medline: 39305940]
9. Sikorski F, König HH, Wegscheider K, Zapf A, Löwe B, Kohlmann S. The efficacy of automated feedback after internet-based depression screening: study protocol of the German, three-armed, randomised controlled trial DISCOVER. Internet Interv. Sep 2021;25:100435. [doi: 10.1016/j.invent.2021.100435] [Medline: 34401394]
10. Kroenke K, Spitzer RL, Williams JBW, Löwe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom scales: a systematic review. Gen Hosp Psychiatry. 2010;32(4):345-359. [doi: 10.1016/j.genhosppsych.2010.03.006] [Medline: 20633738]
11. Löwe B, Decker O, Müller S, et al. Validation and standardization of the Generalized Anxiety Disorder screener (GAD-7) in the general population. Med Care. Mar 2008;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]
12. Gierk B, Kohlmann S, Kroenke K, et al. The somatic symptom scale-8 (SSS-8): a brief measure of somatic symptom burden. JAMA Intern Med. Mar 2014;174(3):399-407. [doi: 10.1001/jamainternmed.2013.12179] [Medline: 24276929]
13. Günther OH, Roick C, Angermeyer MC, König HH. The responsiveness of EQ-5D utility scores in patients with depression: a comparison with instruments measuring quality of life, psychopathology and social functioning. J Affect Disord. Jan 2008;105(1-3):81-91. [doi: 10.1016/j.jad.2007.04.018] [Medline: 17532051]
14. Broadbent E, Petrie KJ, Main J, Weinman J. The brief illness perception questionnaire. J Psychosom Res. Jun 2006;60(6):631-637. [doi: 10.1016/j.jpsychores.2005.10.020] [Medline: 16731240]
15. Broadbent E, Wilkes C, Koschwanez H, Weinman J, Norton S, Petrie KJ. A systematic review and meta-analysis of the Brief Illness Perception questionnaire. Psychol Health. 2015;30(11):1361-1385. [doi: 10.1080/08870446.2015.1070851] [Medline: 26181764]

16. Klee M. Primary analysis code repository for ‘Heterogeneity in effects of automated results feedback after online depression screening: a secondary machine-learning based analysis of the DISCOVER trial. GitHub. 2024. URL: https://github.com/makleelux/discover_hte [Accessed 2025-08-15]
17. Tibshirani J, Athey S, Sverdrup E, Wager S. Grf: generalized random forests. 2.3.2 ed. GitHub. 2024. [Accessed 2025-08-15]
18. Sikorski F, Löwe B, Daubmann A, Kohlmann S. Potential harms of feedback after web-based depression screening: secondary analysis of negative effects in the randomized controlled DISCOVER trial. J Med Internet Res. Apr 30, 2025;27:e59476. [doi: 10.2196/59476] [Medline: 40305104]
19. Meehan AJ, Lewis SJ, Fazel S, et al. Clinical prediction models in psychiatry: a systematic review of two decades of progress and challenges. Mol Psychiatry. Jun 2022;27(6):2700-2708. [doi: 10.1038/s41380-022-01528-4] [Medline: 35365801]
20. Dinga R, Marquand AF, Veltman DJ, et al. Predicting the naturalistic course of depression from a wide range of clinical, psychological, and biological data: a machine learning approach. Transl Psychiatry. Nov 5, 2018;8(1):241. [doi: 10.1038/s41398-018-0289-1] [Medline: 30397196]
21. Moriarty AS, Meader N, Snell KI, et al. Prognostic models for predicting relapse or recurrence of major depressive disorder in adults. Cochrane Database Syst Rev. May 6, 2021;5(5):CD013491. [doi: 10.1002/14651858.CD013491.pub2] [Medline: 33956992]
22. Sikorski F, Löwe B, Kohlmann S. How adults with suspected depressive disorder experience online depression screening: a qualitative interview study. Internet Interv. Dec 2023;34(100685):100685. [doi: 10.1016/j.invent.2023.100685] [Medline: 37954006]

Abbreviations

CF: Causal Forests
ATE: Average treatment effects
AUTOC: Area under the targeting operator characteristic curve
B-IPQ: Brief Illness Perception Questionnaire
CATE: conditional average treatment effects
GAD-7: Generalized Anxiety Disorder Scale-7
HTE: Heterogeneity of treatment effects
ITE: Individual level treatment effects
ML: Machine learning
PHQ-9: Patient Health Questionnaire-9
RCT: Randomized controlled trial
SSS-8: Somatic Symptom Scale-8
TOC: Targeting operator characteristic (curve)

Edited by Fida Dankar; peer-reviewed by Connor T Jerzak, Xin Liu; submitted 12.12.2024; final revised version received 14.04.2025; accepted 24.06.2025; published 21.08.2025

Please cite as:
Klee M, Jaeger BC, Sikorski F, Löwe B, Kohlmann S
Heterogeneity in Effects of Automated Results Feedback After Online Depression Screening: Secondary Machine-Learning Based Analysis of the DISCOVER Trial
JMIR AI 2025;4:e70001
URL: https://ai.jmir.org/2025/1/e70001
doi: 10.2196/70001