Original Paper

Potential Harms of Feedback After Web-Based Depression Screening: Secondary Analysis of Negative Effects in the Randomized Controlled DISCOVER Trial

Franziska Sikorski[1], MSc; Bernd Löwe[1], Prof Dr; Anne Daubmann[2], MSc; Sebastian Kohlmann[1,3], Prof Dr

[1]Department of Psychosomatic Medicine and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[2]Department of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[3]Department of General Internal Medicine and Psychosomatics, University Medical Centre Heidelberg, Heidelberg, Germany

Corresponding Author:
Sebastian Kohlmann, Prof Dr
Department of Psychosomatic Medicine and Psychotherapy
University Medical Center Hamburg-Eppendorf
Martinistr. 52
Hamburg, 20246
Germany
Phone: 49 6221 56 32878
Email: s.kohlmann@uke.de

Abstract

Background: Web-based depression screening followed by automated feedback of results is frequently used and promoted by mental health care providers. However, criticism points to potential associated harms. Systematic empirical evidence on postulated negative effects is missing.

Objective: We aimed to examine whether automated feedback after web-based depression screening is associated with misdiagnosis, mistreatment, deterioration in depression severity, deterioration in emotional response to symptoms, and deterioration in suicidal ideation at 1 and 6 months after screening.

Methods: This is a secondary analysis of the German-wide, web-based, randomized controlled DISCOVER trial. Affected but undiagnosed individuals screening positive for depression (9-item Patient Health Questionnaire [PHQ-9] ≥10 points) were randomized 1:1:1 to receive nontailored feedback, tailored feedback, or no feedback on their screening result. Misdiagnosis and mistreatment were operationalized as having received a depression diagnosis by a health professional and as having started guideline-based depression treatment since screening (self-report), respectively, while not having met the *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition) (*DSM-V*) criteria of a major depressive disorder at baseline (Structured Clinical Interview for *DSM-V* Disorders). Deterioration in depression severity was defined as a pre-post change of ≥4.4 points in the PHQ-9, deterioration in emotional response to symptoms as a pre-post change of ≥3.1 points in a composite scale of the Brief Illness Perception Questionnaire, and deterioration in suicidal ideation as a pre-post change of ≥1 point in the PHQ-9 suicide item. Outcome rates were compared between each feedback arm and the no feedback arm in terms of relative risks (RRs).

Results: In the per protocol sample of 948 participants (n=685, 72% female; mean age of 37.3, SD 14.1 years), there was no difference in rates of misdiagnosis (ranging from 3.5% to 4.9% across all study arms), mistreatment (7.2%-8.3%), deterioration in depression severity (2%-6.8%), deterioration in emotional response (0.7%-2.9%), and deterioration in suicidal ideation at 6 months (6.8%-13.1%) between the feedback arms and the no feedback arm (RRs ranging from 0.46 to 1.96; *P* values ≥.13). The rate for deterioration in suicidal ideation at 1 month was increased in the nontailored feedback arm (RR 1.92; *P*=.01) but not in the tailored feedback arm (RR 1.26; *P*=.43), with rates of 12.3%, 8.1%, and 6.4% in the nontailored, tailored, and no feedback arms, respectively. All but 1 sensitivity analyses as well as subgroup analyses for false-positive screens supported the findings.

Conclusions: The results indicate that feedback after web-based depression screening is not associated with negative effects such as misdiagnosis, mistreatment, and deterioration in depression severity or in emotional response to symptoms. However, it cannot be ruled out that nontailored feedback may increase the risk of deterioration in suicidal ideation. Robust prospective research on negative effects and particularly suicidal ideation is needed and should inform current practice.

KEYWORDS

Introduction

Background

Depressive disorders, although being among the most disabling and most prevalent disorders worldwide [1], often remain undetected and therefore untreated [2]. In the last decades, depression screening has been increasingly discussed as promising to reach those affected but undetected at an early stage. In addition to population-level screening in routine clinical care, as, for example, recommended in the United States [3], advocates also speak out in favor of screening for depression on the web [4]. For many affected individuals, the web is already the favored source for information on mental health [5,6]. Furthermore, so-called web-based depression tests are widely promoted by mental health–related institutions and frequently used by those seeking diagnostic advice [7]. The rationale of web-based depression tests typically involves administering symptom-based screening questionnaires and providing individuals with direct feedback on screening results, sometimes supplemented by direct links or referrals to services. The feedback is thought to empower affected individuals to better act on their symptoms [8] and to seek diagnostic consultation and, if necessary, appropriate care. As such, it might improve early detection and management of depression.

However, feedback after web-based depression screening has been proposed and implemented without due consideration of its appropriateness, that is, without evaluating its effectiveness against the background of potential negative effects. While assessing negative effects is recommended for all clinical interventions, it is particularly important in screening interventions, as in these per definition a substantial amount of participants will not benefit [9]. In addition, in web-based contexts there is no experienced health care stuff available to monitor participants who might need support [10]. As such, the balance between harms and benefits of feedback after web-based depression screening could easily lean toward harms. Indeed, there is no empirical evidence of positive effects on targeted patient-related outcomes: 2 randomized controlled trials, 1 by Batterham and colleagues [11] and the DISCOVER trial recently conducted by our research team [12], do not indicate that feedback of depression-screening results promotes the uptake of evidence-based depression care or reduces depression severity. Negative effects, if present, would therefore likely be generated without creating substantial health benefits.

Evidence regarding negative effects, however, is scarce, with the current scientific debate being mainly reflected by opinion papers. The first area of negative effects of depression screening, discussed in both medical and web-based contexts, relates to inadequate management and care for individuals who receive false-positive feedback. Critics particularly point to the risk of increased rates of misdiagnosis and mistreatment, which refers to the allocation of depression diagnosis and treatment by health care professionals to individuals who screen positive but do not meet diagnostic criteria for a depressive disorder. This, again, is assumed to lead to unnecessary iatrogenic effects such as adverse medication and psychotherapy side effects in healthy individuals, societal costs, and waste of limited health care resources resulting in potential undertreatment of more severe cases [4,13,14]. A second area of concern relates to negative psychological effects to the feedback of screening results. In the field of breast cancer screening, for example, it is well established that screening and particularly false-positive results may be associated with increased anxiety and distress [15]. With regard to depression-screening feedback, it is similarly assumed that feedback-induced labeling, resembling a clinical diagnosis, might induce anxiety, distress, stigma, or nocebo effects as, for example, deterioration of symptoms [4,14,16]. These effects could be amplified by the fact that, in contrast to medical settings, in web-based depression screening, the "diagnosis" would be delivered without a health professional who could provide emotional support or advice on further steps [17]. Indeed, in qualitative studies on web-based mental health screening, some participants describe having been discouraged, shocked, or concerned by the feedback they received [8,18]. Furthermore, 1 observational study found that screening procedures including referrals to in-person care had a higher likelihood of subsequent web-based searches for suicidal intent, potentially suggesting a deterioration of suicidal ideation [7]. In contrast, in our recently conducted DISCOVER trial on feedback after web-based depression screening, less than 1% of participants qualitatively reported any negative effect attributed to trial participation when asked via telephone 6 months after screening, with no indication for an association of negative effects with the recommendation to seek diagnostic advice (see the study by Kohlmann et al [12] for previously published results). However, systematic and large-scale quantitative research on the discussed potential negative effects is outstanding.

Objectives

In this study, we addressed this lack of evidence by analyzing data from our recently conducted randomized controlled DISCOVER trial on feedback after web-based depression screening [12]. In extension to efficacy findings and qualitative reports on negative effects published previously [12], this secondary analysis quantitatively evaluates the potential negative effects discussed in the literature outlined previously (regarding misdiagnosis, mistreatment, and psychological negative effects). Specifically, we aimed to examine whether feedback after

web-based depression screening is associated with increased misdiagnosis and mistreatment 6 months after screening, as well as deterioration in depression severity, deterioration in emotional response to symptoms, and deterioration in suicidal ideation 1 and 6 months after screening.

Methods

Study Design and Participants

The DISCOVER trial [19] and this secondary analysis [20] were preregistered. We conducted small deviations from the preregistration: we added the outcomes misdiagnosis and emotional response to symptoms, as we deemed this of clinical interest. Furthermore, we added sensitivity analyses based on logistic regression models. The detailed study protocol [18] and main results of the trial [12] have been described previously. Data collection was conducted on the web and in the German language between January 12, 2021, and September 30, 2022; this secondary data analysis was conducted between May 3, 2023, and December 23, 2023.

DISCOVER was an investigator-initiated, observer-blinded, 3-armed, randomized controlled trial that compared automated feedback with no feedback after web-based depression screening. After being screened for depression with the digitized 9-item Patient Health Questionnaire (PHQ-9 [21]), eligible participants were randomized to receive no feedback, nontailored feedback, or tailored feedback on their screening result (1:1:1 allocation ratio). Assessments were set at baseline and 1- and 6-month follow-ups. In this secondary analysis, we compared rates of misdiagnosis and mistreatment (at 6 months) as well as deterioration in depression severity, deterioration in emotional response to symptoms, and deterioration in suicidal ideation (at 1 and 6 months) between each feedback arm and the no feedback arm.

Participants were 18 years or older with at least moderate depression severity (PHQ-9 ≥10) but not diagnosed with or treated for depression within the last year. Additional eligibility criteria were having sufficient web-based literacy and German language proficiency, providing contact details, and giving web-based informed consent.

Study Procedures

The study was promoted as being on "stress and psychological well-being" on a publicly accessible study website [22] from January 2021 to January 2022. The aim of evaluating web-based depression screening was not explicitly communicated, but interested participants were informed that some of them will get feedback on a part of their answers. Traditional and social media campaigns as well as print advertisements in public areas of several German cities were used to approach interested individuals across Germany. To reach a sample that strives for representativeness of the German population with respect to age and gender, a marketing company further advertised the study via a nationwide web-based access survey panel.

After completing baseline assessment and screening, eligible participants were automatically randomized by random permuted blocks randomization stratified for baseline depression severity (moderate: PHQ-9 score 10-14 points; severe: PHQ-9 score ≥15

points) and allocated 1:1:1 to 1 of the 3 study arms. Double entries identified based on personal data by a privacy-preserving record linkage service [23] were automatically reallocated to their former study arm. Research staff were masked to allocation at any time until breaking the blind. Due to the design, participants could not be masked but were kept unaware of trial hypotheses to minimize expectancy bias.

Web-based follow-up assessments were set at 1 month and 6 months after randomization, with up to 10 automatic email reminders being sent to participants in case of incomplete surveys. Two to five days and 6 months after randomization, participants were contacted via telephone for complementary diagnostic interviews, with calls being repeated at different hours during daytime and evening in case participants were not reached (see the study protocol by Sikorski et al [18] for more detailed information on data collection procedures).

Web-Based Depression Screening and Feedback of Screening Results

Participants underwent depression screening as part of the baseline survey using the digitized PHQ-9 [21,24] (see Multimedia Appendix 1 for the layout of the digitized version). The use of the PHQ-9 for web-based depression screening is justified by several reasons. First, at the standard cutoff value of ≥10 points, the paper-pencil PHQ-9 demonstrates high discriminatory performance for detecting a major depressive disorder. Based on a recent individual participant data meta-analysis of studies with a semistructured interview reference standard, pooled PHQ-9 sensitivity and specificity (95% CI) were 0.85 (0.79-0.89) and 0.85 (0.82-0.87), respectively [25]. Second, preliminary evidence suggests that psychometric characteristics for the PHQ-9 are comparable for the digitized version [26,27]. Third, the PHQ-9 is recommended for depression screening by the US Preventive Services Task Force and the German National Clinical Practice Guideline for Unipolar Depression [3,28].

After completing the baseline survey, all participants were thanked for participating in the study and received information on follow-up procedures. Participants of the feedback arms were additionally offered feedback on their screening result by clicking on a "next" button (Figure 1 [12]). Both nontailored and tailored feedback comprised four sections: (1) the depression screening result indicating the presence of "significant depressive symptoms," (2) a note to seek diagnostic consultation by a health professional together with a link to make an appointment within the next 2 weeks, (3) brief general information on depression, and (4) information on depression treatment based on the German National Clinical Practice Guideline for Unipolar Depression [28]. Notably, in the German health care system depression care is available and covered by the social health insurance. Information was extended by direct links to referenced health or social services (eg, web-based therapies covered by the health insurance and self-help groups), and the feedback form could be downloaded in a file that included all hyperlinks. In extension to the nontailored feedback, the information in the tailored feedback intervention was personalized to participants' characteristics (eg, "You have indicated that you had l*ow spirits, sleep disturbances,* and *loss*

of energy during the past two weeks."). In addition, after being provided with the screening result (section 1) but before receiving further information (sections 2-4), participants were asked whether they think that their symptoms were indications of depression and whether they worried about the symptoms. According to the participants' answers, the following 3 feedback sections were arranged in a differing order, phrased slightly

differently, and extended by information tailored to participants' risk profile (eg, "Depression *in pregnancy* is common."). The feedback was developed in a multistage process involving patient representatives [29,30] and a digital graphic agency to adapt the material to the possibilities of web-based presentation. Illustrations of the complete nontailored and tailored feedback versions can be found in Multimedia Appendices 2 and 3.

Figure 1. Illustrations of no feedback, nontailored feedback (first screen), and tailored feedback (first screen; reprinted from the study by Kohlmann et al [12]).
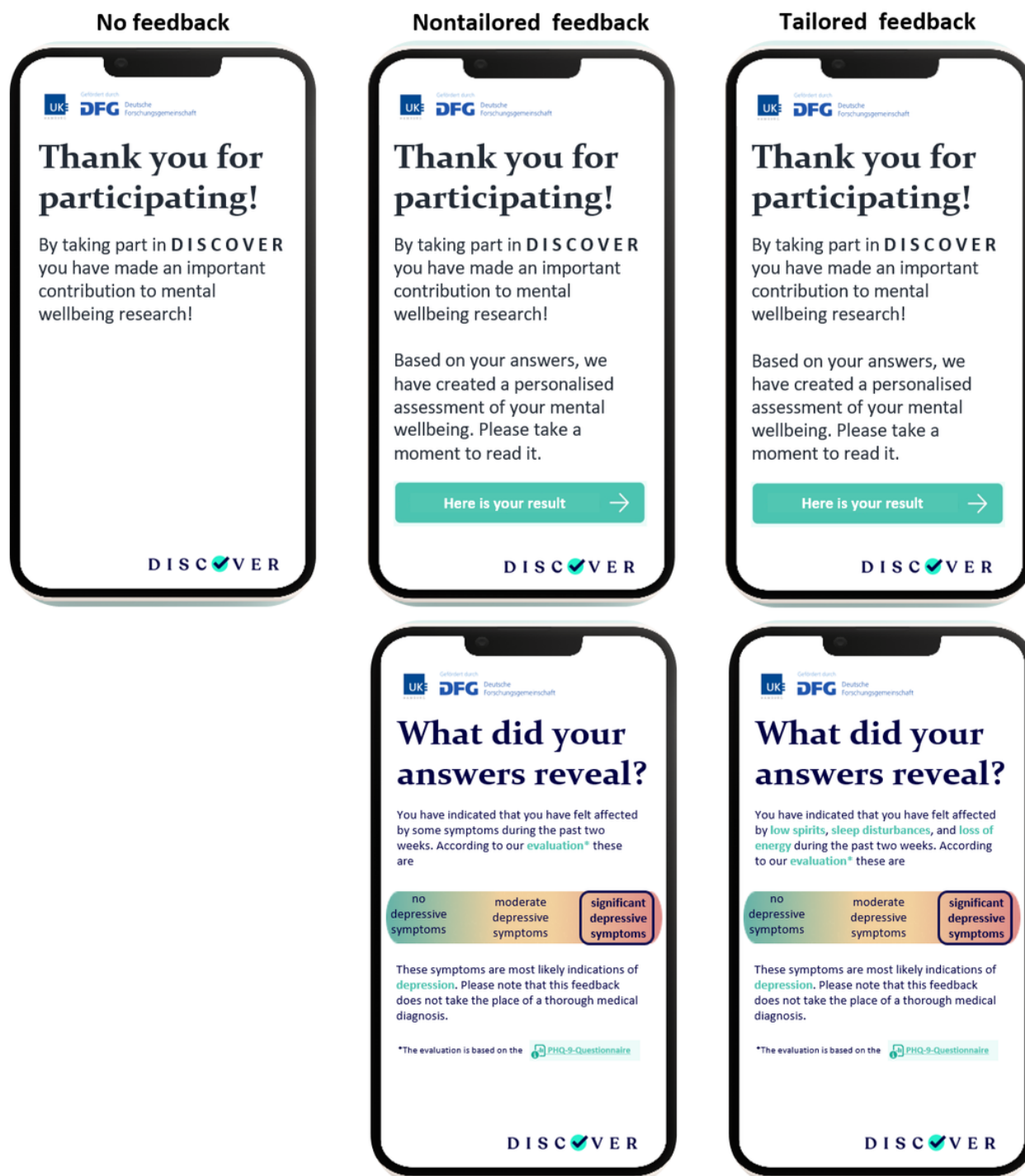
Due to ethical considerations, all participants who have indicated elevated suicidal ideation (PHQ-9 suicide item ≥2; *more than half the days*) were shown a screen providing advice to urgently seek help and relevant information on available help services (eg, general practitioner, local psychiatric emergency units, and the national emergency number; Multimedia Appendix 4).

Measures

Depression diagnosis by a health professional was assessed at 6 months with the question: "Have you been diagnosed with depression or burnout in the last six months?"

Guideline-based depression treatment, that is, pharmacotherapy with antidepressant medication or psychotherapy recommended by the German National Clinical Practice Guideline for Unipolar Depression [28], was assessed at 6 months with the questions: "Have you started any psychotherapy or similar treatment in the last 6 months [which]?") and "Have you started taking medication to treat depression or other complaints such as sleep problems, anxiety or stress [which ones]?" Participants could choose from guideline-based treatment options or give open answers. In case of open answers, these were checked for guideline conformity independently by 2 of the authors (SK and FS).

Criteria for a major depressive disorder at baseline were assessed with the depression-related modules of the Structured Clinical Interview for *DSM-V* Disorders (SCID-5-CV) [31] 2-5 days after screening. The interviewers (MSc psychology students) were trained and supervised by the project leader, who is an experienced psychotherapist. Participants who did not meet the criteria for a major depressive disorder according to the SCID were considered false-positive screens.

Depression severity was assessed with the PHQ-9 at 1 and 6 months after screening. In accordance with the *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition) (*DSM-V*) diagnostic criteria, the PHQ-9 assesses 9 depressive symptoms each rated in terms of frequency during the past 2 weeks (0-3; *not at all* to *nearly every day*), resulting in a total score ranging from 0 to 27, with a higher score indicating higher depression severity. The PHQ-9 is among the most frequently used self-report depression questionnaires, has good psychometric properties, and is sensitive to change [21,32].

Suicidal ideation was assessed with the PHQ-9 suicide item (item 9): "Over the last two weeks, how often have you been bothered by thoughts that you would be better off dead or of hurting yourself in some way?" rated from 0 to 3 (*not at all, several days, more than half the days,* and *nearly every day*).

Emotional response to depressive symptoms was assessed with a composite scale based on 2 items of the Brief Illness Perception Questionnaire (Brief IPQ) that cover emotional representations of depressive symptoms: "How concerned are you about your symptoms?" and "How much do your symptoms affect you emotionally? (eg, do they make you angry, scared, upset or depressed)?" The items were assessed directly after the PHQ-9 and were scored on a Likert scale ranging from 0 (*not at all*) to 10 (*extremely*). Item scores were pooled for the composite scale, resulting in a total scale ranging from 0 to 10. The respective items of the Brief IPQ showed good psychometric properties [33].

Outcomes

Participants were classified as misdiagnosed or mistreated if they reported having received a depression diagnosis by a health professional or guideline-based depression treatment while not having met the criteria for a major depressive disorder at baseline (SCID), that is, while being screened false positive.

Deterioration in depression severity was defined as a pre-post change score of at least 4.4 points in the PHQ-9. The cutoff is based on the reliable change index (RCI), a psychometric criterion to evaluate whether a change in symptoms is considered statistically reliable, that is, not attributable to measurement error [34]. The RCI was calculated using the PHQ-9 SD from the current sample ($SD_{baseline}=4$), the reliability coefficient from the PHQ-9 validation study ($r_{tt}=0.84$) [21], and a 95% CI. The resulting RCI of 4.4 points is comparable with cutoffs found in comparable research [32,35].

Deterioration in emotional response to depressive symptoms was defined as a pre-post change score of at least 3.1 points in the relating composite scale. The RCI was calculated using the SD of this composite scale ($SD_{baseline}=1.9$), the pooled reliability coefficients from the Brief IPQ validation study ($r_{tt}=0.66$), and a 95% CI.

Deterioration in suicidal ideation was defined as the pre-post change score of at least 1 point in the PHQ-9 suicide item.

Sample

We performed this secondary analysis in the per protocol sample that included 89% (948/1178) of randomized participants who had at least one postbaseline value of one of the outcomes and no major protocol violation. Major protocol deviations were predefined as not receiving or adhering to the intervention (ie, feedback not opened, feedback reading time 15 seconds, or no download of feedback form), multiple participation (post hoc data check or self-report), reports of not having answered the survey seriously, baseline survey completion time less than 2 minutes, and provision of an invalid email address. We preferred per protocol over intention-to-treat (ITT) analysis, as the second is likely to underestimate the risk of an event by inflating the denominator with participants who have provided invalid data or have never received the intervention. Whereas this is conservative in efficacy evaluations, in the current case of a risk evaluation we consider it more appropriate to prevent failing to detect a risk than overestimating it [36].

In addition, we performed sensitivity analyses in ITT sample, both with and without missing data imputation. We used 2 strategies for imputing data: assuming that all dropouts were deteriorators, considering this to be the most conservative estimate (worst case); and assuming that all dropouts were nondeteriorators, considering this to be the most optimistic estimate (best case).

Statistical Analyses

We compared the rates of negative effects between study arms in terms of relative risks (RRs). The RR estimates how much higher (or lower) the probability of negative effects is for participants in each feedback arm compared with the no feedback arm. To directly estimate the RR with 95% CIs, we applied generalized linear models with a log link and robust sandwich variance estimator using modified log-Poisson regressions [37]. We chose this approach over alternative models as it is suited as well in case of frequent outcomes and suffers

least from convergence problems [38,39]. To test for differential effects in the subgroup of false-positive screened participants, we ran another series of models additionally including false-positive screens and the false-positive screen by study arm interaction term. We set the significance level at α=.05 and did not correct for multiple testing for 2 reasons: the trial was not powered for this secondary analysis, and as already mentioned, in case of negative effects we consider it more important to prevent the inflation of a type II error (ie, failing to detect negative effects in case they exist) instead of the type I error. As some negative effects turned out to be rare in the study data, we also estimated odds ratios based on logistic regression models as post hoc sensitivity analyses. We performed analyses with SPSS software (version 27; IBM Corp).

Ethical Considerations

The data used in this secondary analysis are derived from the DISCOVER RCT, which has been reviewed and approved by the ethics committee of the Hamburg Medical Chamber (# PV7039). The study and reporting of this manuscript followed appropriate CONSORT (Consolidated Standards of Reporting Trials) guidelines, including the harms and the eHealth statement [9,40-44] (Multimedia Appendix 5). Participants received detailed study information, including information on their ability to withdraw from the study at any time and without giving reasons. Web-based informed consent, covering the use of data for secondary analyses, was obtained from all participants via checkboxes. Participation was compensated with Amazon vouchers worth up to €15 (US $17.11; one €5, US $5.70, voucher per complete follow-up assessment). The data were deidentified (pseudonymized) after the completion of data collection.

Results

Overview

Of initially 5457 study participants, 4878 completed the screening questionnaire, and 1178 eligible participants were assigned to receive no feedback (n=391), nontailored feedback (n=393), or tailored feedback (n=394) on their depression screening result. Of the 787 participants randomized to receive any feedback, 95% (n=744) opened the feedback screen, of whom 62% (n=464) downloaded the PDF and 33% (n=248) interacted with the feedback by clicking at least 1 link or modal. Descriptively, there was no difference between the feedback engagement across feedback arms (see the study by Kohlmann et al [12] for the results per study arm). At 1 month, 976 participants provided follow-up data of outcome measures (loss to follow-up: 17%), of whom 909 were included in the per protocol analysis. At 6 months, 965 participants provided follow-up data of outcome measures (loss to follow-up: 18%), of whom 902 were included in the per protocol analysis. The numbers per study arm and analysis time points are shown in the CONSORT flowchart (Figure 2).

Relevant demographic and clinical characteristics of the per protocol sample were balanced across the 3 study arms (Table 1). The mean participant age was 37.3 (SD 14.1) years, 72% (685/948) of the participants were women, and 52% (488/948) of the participants had a high education level. At baseline, the average PHQ-9 depression severity score was 14.8 (SD 4.0), the average score in emotional response to depressive symptoms was 7 (SD 1.9), 48% (455/948) of the participants reported to experience suicidal ideation at least several days within the last 2 weeks, and 87% (820/948) of the participants thought that they currently experienced or maybe experienced depression. Out of 820 participants who were reached for diagnostic telephone interviews, 63% (n=514) met the criteria for a major depressive disorder according to the *DSM-V*. Conversely, 37% (n=306) of the participants were classified as false-positive screens. Characteristics of the per protocol sample are comparable with those of the ITT sample (Multimedia Appendix 6).

Figure 2. CONSORT flowchart (per protocol sample). PHQ-9: 9-item Patient Health Questionnaire-9; SCID: Structured Clinical Interview for DSM-V Disorders.
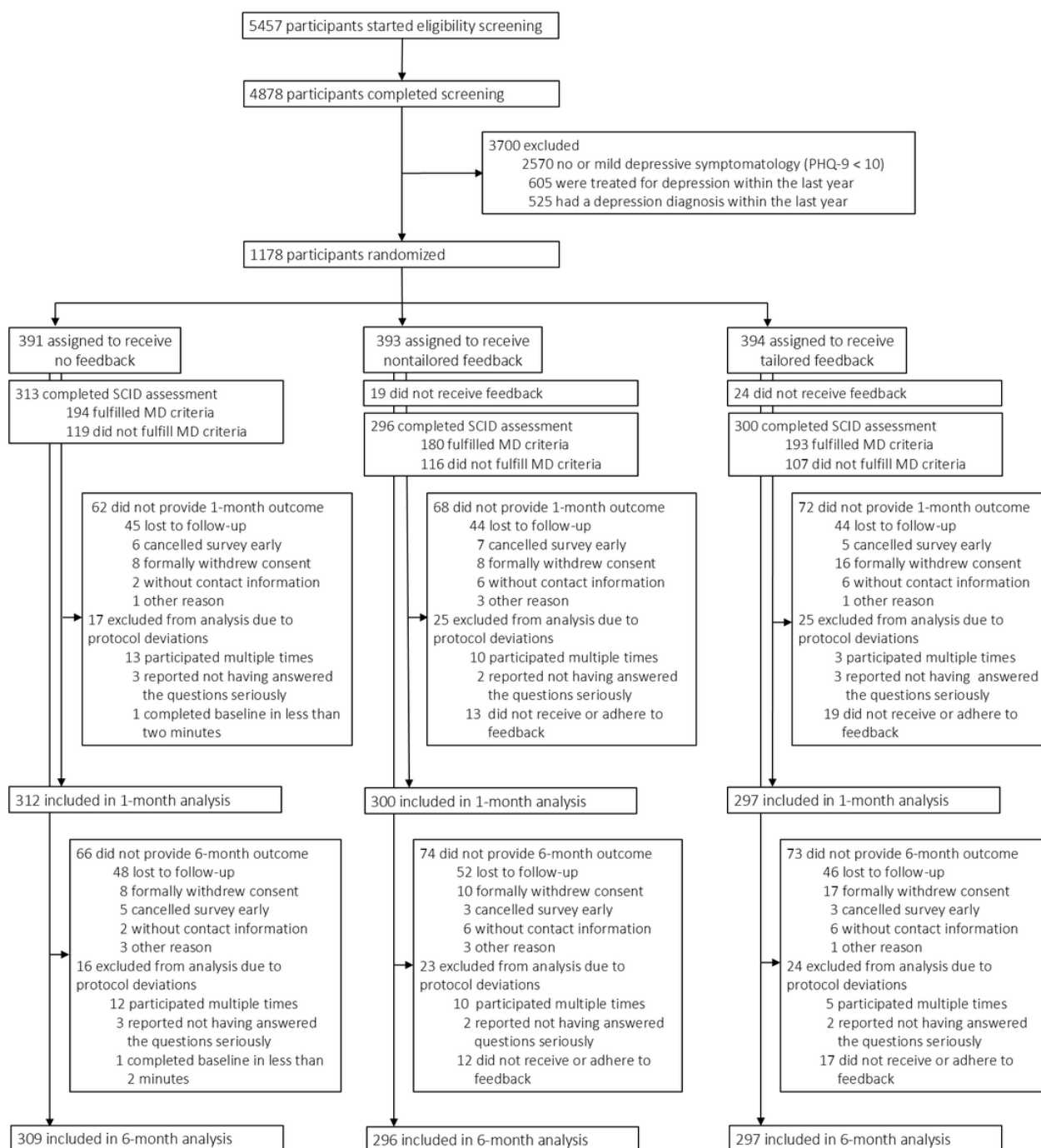
Table 1. Baseline demographic and clinical characteristics of the per protocol sample.

	Total sample (n=948)	Nontailored feedback (n=314)	Tailored feedback (n=307)	No feedback (n=327)
Age (years), mean (SD)	37.3 (14)	37.8 (14)	36.8 (14.3)	37.2 (14)
Sex, n (%)				
Female	685 (72)	223 (71)	219 (71)	243 (74)
Male	255 (27)	88 (28)	85 (28)	82 (25)
Divers	8 (0.8)	3 (1.0)	3 (1.0)	2 (0.6)
German mother tongue, n (%)	902 (95)	295 (94)	296 (96)	311 (95)
Migration background, n (%)	103 (11)	30 (10)	32 (10)	41 (13)
Being in a relationship, n (%)	445 (47)	162 (52)	143 (47)	140 (43)
Living with others, n (%)	631 (67)	217 (69)	202 (66)	212 (65)
Formal school education, n (%)				
Low (<10 years)	160 (17)	60 (19)	44 (15)	56 (17)
Middle (at least 10 years)	300 (32)	100 (32)	102 (33)	98 (30)
High (A-level or above)	488 (52)	154 (49)	161 (52)	173 (53)
Working, n (%)	691 (73)	230 (73)	232 (76)	229 (70)
Depression severity (PHQ-9[a]), mean (SD)	14.8 (4)	14.9 (4.2)	14.6 (3.8)	14.8 (4)
Emotional response to depressive symptoms (composite scale), mean (SD)	7 (1.9)	7 (1.9)	6.9 (1.7)	6.9 (2)
Quality of life (EQ-5D-5L VAS[b]), mean (SD)	57.6 (21.6)	57.2 (21.6)	58.2 (21.3)	57.4 (21.9)
Anxiety severity (GAD-7[c]), mean (SD)	12.1 (4.3)	12.5 (4.2)	11.8 (4.3)	12 (4.3)
Somatic symptom severity (SSS-8[d]), mean (SD)	14.4 (5.2)	14.5 (5.1)	14.2 (5.2)	14.4 (5.2)
Depression risk factors[e] (n), mean (SD)	6 (2.4)	6.1 (2.5)	5.9 (2.3)	6.1 (2.5)
Frequency of suicidal ideation within last 2 weeks (PHQ-9 item 9), n (%)				
None	493 (52)	161 (51)	165 (54)	167 (51)
Several days	305 (32)	113 (36)	94 (31)	98 (30)
More than half the days	86 (9)	23 (7)	26 (9)	37 (11)
Nearly every day	64 (7)	17 (5)	22 (7)	25 (8)
Self-identifying as experiencing depression, n (%)				
No	128 (14)	32 (10)	51 (17)	45 (14)
Maybe	432 (46)	160 (51)	141 (46)	131 (40)
Yes	388 (41)	122 (39)	115 (38)	151 (46)
Meeting criteria for major depressive disorder (SCID[f])	514 (63)[g]	161 (61)[h]	172 (64)[i]	181 (62)[j]

[a]PHQ-9: Patient Health Questionnaire-9 (0-27).

[b]VAS: visual analogue scale (0-100).

[c]GAD-7: Generalized Anxiety Disorder-7 (0-21).

[d]SSS-8: Somatic Symptom Scale (0-32).

[e]Risk factors included self-reported anxiety, addiction, traumatic life events, persistent physical symptoms, mood swings, chronic physical condition, lack of social support, mental comorbidity, mental comorbidity in family, history of suicide, current pregnancy, postnatal phase, menopause, and premenstrual syndrome.

[f]SCID: Structured Clinical Interview for *DSM-V* Disorders.

[g]A total of 128 cases with missing data.

[h]A total of 37 cases with missing data.

[i]A total of 51 cases with missing data.

[j]A total of 40 cases with missing data.

XSL•FO

RenderX

Negative Effects Outcomes

Misdiagnosis rates 6 months after randomization were not higher after nontailored (RR 1.30, 95% CI 0.59-2.86; *P*=.51) or tailored feedback (RR 1.09, 95% CI 0.48-2.46; *P*=.84) as compared with no feedback, with rates of 4.9%, 4.1%, and 3.5% in the nontailored, the tailored, and the no feedback arm, respectively. Mistreatment rates 6 months after randomization were not higher after nontailored (RR 0.87, 95% CI 0.49-1.56; *P*=.65) or tailored feedback (RR 0.95, 95% CI 0.54-1.67; *P*=.86) either, with rates of 7.2%, 7.7%, and 8.3% in the nontailored, the tailored, and the no feedback arm. Rates of deterioration in depression severity were not higher after nontailored (1 month: RR 1.96, 95% CI 0.89-4.34; *P*=.10; 6 months: RR 0.60, 95% CI 0.3-1.19; *P*=.14) or tailored feedback (1 month: RR 0.70, 95% CI 0.25-1.94; *P*=.49; 6 months: RR 0.74, 95% CI 0.39-1.41; *P*=.37), with rates of 5.7%, 2.0%, and 2.9% at 1 month and 4.1%, 5.1%, and 6.8% at 6 months in the nontailored, tailored, and no feedback study arms. Rates of deterioration in emotional response to depressive symptoms were not higher after nontailored (1 month: RR 1.18, 95% CI 0.43-3.21; *P*=.75; 6 months: RR 0.46, 95% CI 0.14-1.49; *P*=.20) or tailored feedback (1 month: RR 0.23, 95% CI 0.06-1.42; *P*=.13; 6 months: RR 0.70, 95% CI 0.25-1.94; *P*=.49) either, with rates of 2.7%, 0.7%, and 2.3% at 1 month and 1.4%, 2.0%, and 2.9% at 6 months. Rates of deterioration in suicidal ideation were not higher after nontailored (RR 1.12, 95% CI 0.69-1.8; *P*=.66) or tailored feedback (RR 1.40, 95% CI 0.39-1.41; *P*=.15) at 6 months, with rates of 10.5%, 13.1%, and 9.4%. At 1 month, however, the rate of deterioration in suicidal ideation was almost 2-fold higher in the nontailored (RR 1.92, 95% CI 1.14-3.24; *P*=.01) but not in the tailored feedback arm (RR 1.26, 95% CI 0.25-1.94; *P*=.43), as compared with no feedback. Rates in the nontailored, the tailored, and the no feedback arms were 12.3%, 8.1%, and 6.4%. Absolute frequencies and rates for all negative effects per study arm and time point are shown in Table 2. Relative risks with corresponding 95% CIs are illustrated in Figure 3.

Results did not differ for the subgroup of false positives (*P*~interaction~ ranging between .29 and .80). Sensitivity analyses based on logistic regression models, as well as those in the ITT sample with the full analysis set and with missing data imputation based on the best-case scenario, showed comparable results. In the ITT analysis based on the worst case scenario, however, the RR for deterioration in suicidal ideation in the nontailored feedback arm at 1 month was not higher than that in the no feedback arm (RR 1.26, 95% CI 0.99-1.61; *P*=.07; Multimedia Appendix 7). Based on post hoc analyses, baseline demographic and clinical characteristics of all participants deteriorated in any outcome at any time point were comparable with the total sample (Multimedia Appendix 8).

Table 2. Absolute frequencies and rates of negative effects per study arm and time point.

	Participants	Nontailored feedback	Participants	Tailored feedback	Participants	No feedback
Misdiagnosis (6 months)[a], n (%)	263	13 (4.9)	267	11 (4.1.)	290	11 (3.5)
Mistreatment (6 months)[a], n (%)	263	19 (7.2)	267	21 (7.7)	290	24 (8.3)
Psychotherapy[a]	263	13 (4.9)	267	17 (6.4)	290	18 (6.2)
Medication[a]	263	9 (3.4)	267	6 (2.2)	290	8 (2.8)
Deterioration in depression severity, n (%)						
1 month	300	17 (5.7)	297	6 (2.0)	312	9 (2.9)
6 months	296	12 (4.1)	297	15 (5.1)	309	21 (6.8)
Deterioration in suicidal ideation, n (%)						
1 month	300	37 (12.3)[b]	297	24 (8.1)	312	20 (6.4)
6 months	296	31 (10.5)	297	39 (13.1)	309	29 (9.4)
Deterioration in emotional response, n (%)						
1 month	299	8 (2.7)	296	2 (0.7)	308	7 (2.3)
6 months	294	4 (1.4)	299	6 (2)	307	9 (2.9)

[a]Participants who completed both the follow-up assessment and the Structured Clinical Interview for *DSM-V* Disorders depression module at baseline.

[b]Significantly increased relative risk as compared with no feedback, with *P*<.05.

Figure 3. Relative risks (95% CIs) for all negative effects at all time points in the nontailored and tailored feedback arms as compared with no feedback.
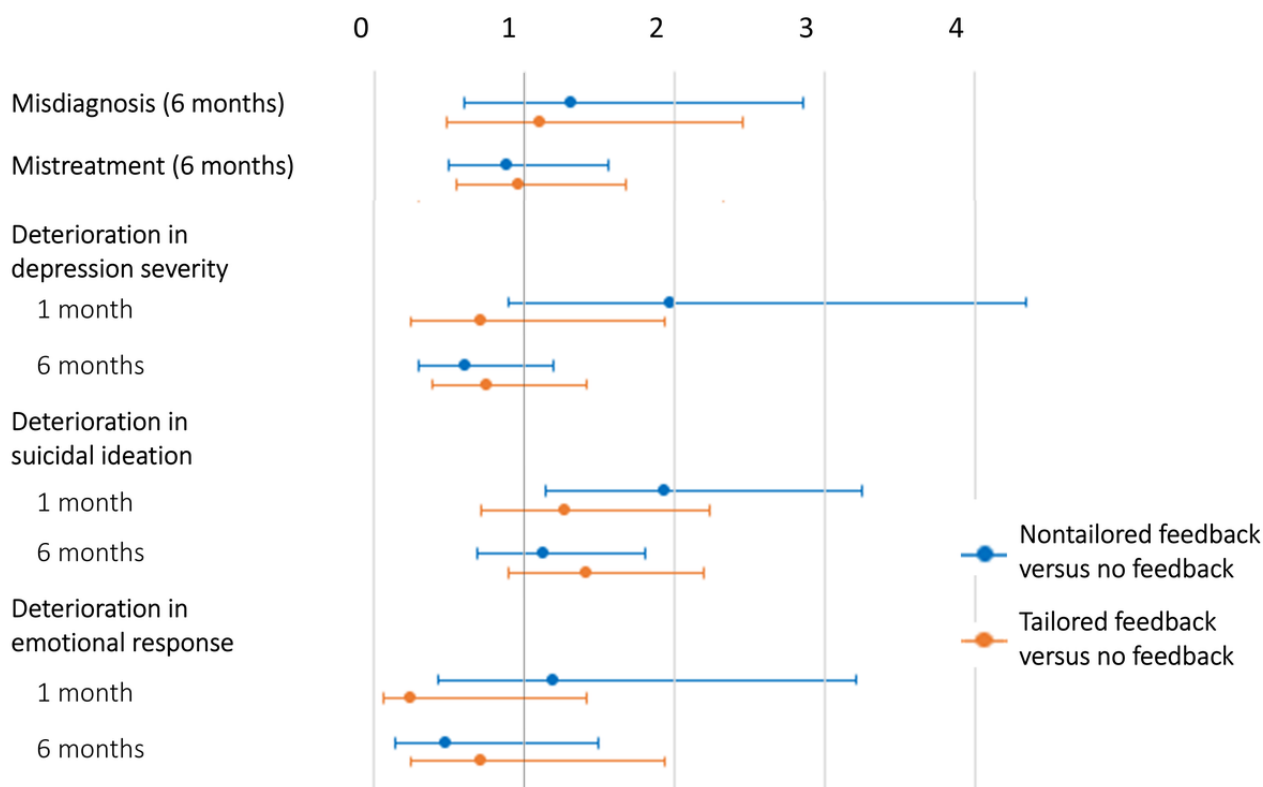
Discussion

To the best of our knowledge, this secondary analysis is the first study to systematically examine potential negative effects of feedback after web-based depression screening in a large sample of currently undiagnosed and untreated individuals with at least moderate depression severity.

Principal Findings

The results indicate that feedback, both nontailored and tailored, was not associated with increased rates of misdiagnosis, mistreatment, deterioration in depression severity, or deterioration in emotional response to symptoms as compared with no feedback. Deterioration of suicidal ideation, however, appeared to be more likely 1 month after receiving nontailored feedback, as compared with no feedback. Although almost 40% of the sample turned out to be screened false positive, irrespective of the study arm, rates of subsequent misdiagnosis and mistreatment were lower than 5% and 9%, respectively, with rates of pharmacotherapy ranging even lower than 4%. Across study arms, deterioration in emotional response to depressive symptoms was reported by at most 3% of participants, deterioration in depression severity by at most 7% of participants, and deterioration of suicidal ideation by at most 13% of participants.

Comparison With Prior Work

There are 3 main findings that contribute to the scientific debate on negative effects of web-based depression screening outlined in the Introduction section. First, the results regarding mistreatment and misdiagnosis emphasize that feedback after web-based depression screening is not associated with inadequate management and care for individuals who receive false-positive feedback—even when the rate of false positives is relatively high and when the feedback refers to a health system that covers depression care as it is true for Germany. These results extend on prior findings that feedback after web-based depression screening does not affect service uptake [12,43]. Furthermore, they refute the most prominent but opinion-based criticism against web-based depression screening [4,14].

Second, there is also no indication that feedback after web-based depression screening induces negative psychological effects such as deterioration in depression severity and emotional response to symptoms. Notably, the rates for deterioration in depression severity of at most 7% found in this study are comparable with those reported in care-as-usual conditions in psychotherapy trials [45]. The null findings regarding deterioration in emotional response to symptoms, however, appear to conflict with prior qualitative evidence suggesting that web-based depression screening does induce negative emotions and distress in some individuals [8,18]. An explanation for this discrepancy might be that negative emotional effects might be induced not only by the feedback but also by the screening questions alone, which has been reported in a

qualitative follow-up study of the DISCOVER trial [18]. Furthermore, it might be that the construct emotional response, defined by items assessing concern and emotional affectedness about the symptoms, relates more to a cognitive evaluation of symptoms rather than capturing an actual emotional state. Therefore, assessing outcomes such as distress or negative affectivity shortly after providing screening and comparing these between a screening only and a feedback condition appear worthwhile to further address these issues (see the studies by Gould et al [46] and Robinson et al [47] for exemplary study designs in suicide screening).

Third, the current results indicate that nontailored feedback, in contrast to tailored feedback, might lead to increased suicidal ideation after 1 month. This finding is contradictory to results from a randomized clinical trial on screening and feedback in the primary care setting [48] but in line with prior observational evidence regarding web-based screening [7]. Explanations for such an effect might be that receiving a diagnosis on the web might induce hopelessness, a known risk factor for suicidal ideation [49], or that the referral initiation process may be overwhelming, thereby triggering decompensation [7]. However, it remains an open question why nontailored but not tailored feedback should increase suicidal ideation: against our hypothesis, neither the usage of the feedback nor any other outcome differed between the 2 feedback arms [12].

Notably, when qualitatively asked in follow-up telephone interviews 6 months after screening, only 1% (9/909) of the participants retrospectively reported negative effects attributed to trial participation (see the study by Kohlmann et al [12]). Explanations for the discrepancy between this low rate and present negative effects rates of up to 13% might be that in qualitative interviews, negative effects might have been stigmatizing to report (eg, in the case of suicidal ideation), might not be remembered retrospectively, or might subjectively not be classified as an adverse event by participants.

Limitations

The interpretation of the present results should be considered in the context of the study's limitations. First, this secondary analysis of the DISCOVER trial was planned post hoc and therefore not powered to detect differences between study arms regarding the selected outcomes. Indeed, rates of negative effects turned out to be low (with a maximum of 39 participants per study arm), wherefore analyses may be underpowered to detect significant effects. On the other hand, multiple testing might have led to overestimation of significance in the case of deterioration in suicidal ideation. To robustly examine negative effects in the future, trials with higher sample sizes that prospectively address negative effects are needed.

Second, the underlying DISCOVER trial did not explicitly call for individuals seeking depression screening. As these may be more eager to follow the advice of the feedback, in the current sample, misdiagnosis and mistreatment might be underestimated as compared with individuals using public depression screening tools. To increase the generalizability of results, future studies should target the recruitment of participants who actively seek web-based depression screening.

Third, as this is a secondary analysis, outcome selection was limited. Although the existing data allowed for assessing a range of relevant negative effects, future research should consider including further outcomes such as distress, suicidal behavior, stigma, treatment side effects, or overdiagnosis (ie, the diagnosis of correctly diagnosed but mild cases that would not benefit from treatment [16]). Furthermore, the present outcomes are based on self-reports and would benefit from more objective data (eg, for misdiagnosis) from health care providers. Regarding misdiagnosis and mistreatment, it cannot be ruled out that participants (correctly) received a burnout diagnosis or antidepressant medication or psychotherapy for conditions other than depressive disorders, wherefore rates may be overestimated. Furthermore, the operationalizations of suicidal ideation and emotional response to depressive symptoms are based on a single item and a composite score, respectively, that are not well validated for this purpose (see the studies by Na et al [50] and Rossom et al [51] for research on the validity of the PHQ-9 suicide item). Future research on negative effects should use valid and reliable measures to assess these outcomes (see the study by Erford et al [52] for a review on suicide ideation assessment instruments).

Notably, the findings refer to the German health care system, where psychotherapy is available and covered by the social health insurance. Particularly, rates for misdiagnosis and mistreatment might differ in other countries with differing health policies.

Future Directions

Given that the current results should be interpreted with caution due to the study's limitations, more robust research is needed to further address negative effects in web-based depression screening. Particularly, prospective and well-powered trials that validly assess suicidal ideation, preferably directly after the provision of screening and feedback, are needed (see the studies by Gould et al [46] and Robinson et al [47] for exemplary study designs in suicide screening). If future studies corroborate an association of web-based screening and feedback with suicidal ideation, this finding needs to inform regulations of currently unmonitored web-based depression tests. Furthermore, the findings should also inform research regarding comparable depression screening in medical and primary care settings, which is currently recommended in many countries despite very uncertain evidence regarding potential harms [53].

Conclusions

The results of this secondary analysis indicate that feedback after web-based depression screening is neither associated with health care–related negative effects such as misdiagnosis and mistreatment nor with psychological negative effects such as deterioration in depression severity or emotional response to symptoms. However, it cannot be ruled out that nontailored feedback may be associated with increased suicidal ideation. Against the background of the study's secondary design, robust prospective research on negative effects and particularly suicidal ideation in web-based depression screening is needed to inform current practice of public web-based depression screening as well as research in the field of depression screening in general.

Data Availability

Individual participant data that underlie the results reported in this paper, after deidentification (text, tables, figures, and appendices), will be shared by the corresponding author upon reasonable request for academic and research purposes and subject to data-sharing agreements.

Authors' Contributions

FS, AD, and SK designed the study. SK and BL obtained the funding for the underlying DISCOVER trial. FS and SK collected the data of the DISCOVER trial. FS performed the data analyses and wrote the original draft of the manuscript. All authors contributed to the interpretation of the data for the paper and critically reviewed and edited the original draft. All authors had full access to all the data in the study, approved the final version of the manuscript, and take responsibility for its submission for publication.

Conflicts of Interest

FS declares that there is no conflict of interest. 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 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 has received remuneration for several scientific book articles from various book publishers 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 Conferin a Na ională de Psihosomatică, Cluj-Napoca, Romania, October 2023. He was a board member of the EAPM (unpaid) until 2022. AD 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, and German Cancer Aid.

Multimedia Appendix 1

Illustration of the digitized 9-item Patient Health Questionnaire as displayed to study participants.
[PPTX File , 672 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Illustrations of complete nontailored feedback.
[PPTX File , 1476 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Illustrations of complete tailored feedback.
[PPTX File , 1457 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Illustration of suicidal ideation feedback.
[PPTX File , 443 KB-Multimedia Appendix 4]

Multimedia Appendix 5

CONSORT-eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 767 KB-Multimedia Appendix 5]

Multimedia Appendix 6

Characteristics of intention-to-treat sample.

[[DOCX File , 24 KB-Multimedia Appendix 6](#)]

Multimedia Appendix 7

Sensitivity analyses.
[[DOCX File , 33 KB-Multimedia Appendix 7](#)]

Multimedia Appendix 8

Post hoc analyses.
[[DOCX File , 23 KB-Multimedia Appendix 8](#)]

References

1. GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. Lancet Psychiatry. 2022;9(2):137-150. [doi: 10.1016/s2215-0366(21)00395-3]
2. Ghio L, Gotelli S, Marcenaro M, Amore M, Natta W. Duration of untreated illness and outcomes in unipolar depression: a systematic review and meta-analysis. J Affect Disord. 2014;152-154:45-51. [doi: 10.1016/j.jad.2013.10.002] [Medline: 24183486]
3. Barry MJ, Nicholson WK, Silverstein M, Chelmow D, Coker TR, Davidson KW, et al. Screening for depression and suicide risk in adults: US preventive services task force recommendation statement. JAMA. 2023;329(23):2057-2067. [doi: 10.1001/jama.2023.9297] [Medline: 37338872]
4. Duckworth K, Gilbody S. Should Google offer an online screening test for depression? BMJ. 2017;358:j4144. [doi: 10.1136/bmj.j4144] [Medline: 28903920]
5. Vaidyanathan U, Sun Y, Shekel T, Chou K, Galea S, Gabrilovich E, et al. An evaluation of internet searches as a marker of trends in population mental health in the US. Sci Rep. 2022;12(1):8946. [FREE Full text] [doi: 10.1038/s41598-022-12952-5] [Medline: 35624317]
6. Eichenberg C, Wolters C, Brähler E. The internet as a mental health advisor in Germany—results of a national survey. PLoS One. 2013;8(11):e79206. [FREE Full text] [doi: 10.1371/journal.pone.0079206] [Medline: 24278121]
7. Jacobson NC, Yom-Tov E, Lekkas D, Heinz M, Liu L, Barr PJ. Impact of online mental health screening tools on help-seeking, care receipt, and suicidal ideation and suicidal intent: evidence from internet search behavior in a large U.S. cohort. J Psychiatr Res. 2022;145:276-283. [FREE Full text] [doi: 10.1016/j.jpsychires.2020.11.010] [Medline: 33199054]
8. Kruzan KP, Meyerhoff J, Nguyen T, Mohr DC, Reddy M, Kornfield R. "I Wanted to See How Bad it Was": online self-screening as a critical transition point among young adults with common mental health conditions. Proc SIGCHI Conf Hum Factor Comput Syst. 2022;2022:328. [FREE Full text] [doi: 10.1145/3491102.3501976] [Medline: 35531062]
9. Junqueira DR, Zorzela L, Golder S, Loke Y, Gagnier JJ, Julious SA, et al. CONSORT Harms Group. CONSORT Harms 2022 statement, explanation, and elaboration: updated guideline for the reporting of harms in randomised trials. BMJ. 2023;381:e073725. [FREE Full text] [doi: 10.1136/bmj-2022-073725] [Medline: 37094878]
10. Rozental A, Andersson G, Boettcher J, Ebert DD, Cuijpers P, Knaevelsrud C, et al. Consensus statement on defining and measuring negative effects of internet interventions. Internet Interv. 2014;1(1):12-19. [doi: 10.1016/j.invent.2014.02.001]
11. Batterham PJ, Calear AL, Sunderland M, Carragher N, Brewer JL. Online screening and feedback to increase help-seeking for mental health problems: population-based randomised controlled trial. BJPsych Open. 2016;2(1):67-73. [FREE Full text] [doi: 10.1192/bjpo.bp.115.001552] [Medline: 27703756]
12. Kohlmann S, Sikorski F, König H, 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. 2024;6(7):e446-e457. [doi: 10.1016/s2589-7500(24)00070-0]
13. Thombs BD, Markham S, Rice DB, Ziegelstein RC. Screening for depression and anxiety in general practice. BMJ. 2023;382:1615. [doi: 10.1136/bmj.p1615] [Medline: 37460129]
14. Danczak A. Online screening test for depression is inappropriate. BMJ. 2017;359:j4736. [doi: 10.1136/bmj.j4736] [Medline: 29042347]
15. Nelson HD, Pappas M, Cantor A, Griffin J, Daeges M, Humphrey L. Harms of breast cancer screening: systematic review to update the 2009 U.S. preventive services task force recommendation. Ann Intern Med. 2016;164(4):256-267. [doi: 10.7326/m15-0970]
16. Thombs B, Turner KA, Shrier I. Defining and evaluating dverdiagnosis in mental health: a meta-research review. Psychother Psychosom. 2019;88(4):193-202. [doi: 10.1159/000501647] [Medline: 31340212]
17. Ryan A, Wilson S. Internet healthcare: do self-diagnosis sites do more harm than good? Expert Opin Drug Saf. 2008;7(3):227-229. [doi: 10.1517/14740338.7.3.227] [Medline: 18462181]

18. 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. 2021;25:100435. [FREE Full text] [doi: 10.1016/j.invent.2021.100435] [Medline: 34401394]

19. ClinicalTrials.gov. The efficacy of automated feedback after internet-based depression screening (DISCOVER). Sep 21, 2023. URL: https://clinicaltrials.gov/study/NCT04633096 [accessed 2023-10-21]

20. Sikorski F, Kohlmann S. Does internet-based depression screening with feedback of results cause harm? A secondary analysis using data from the randomised controlled DISCOVER trial. Open Science Framework. May 02, 2023. URL: https://osf.io/tzyrd [accessed 2023-10-21]

21. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16(9):606-613. [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]

22. DISCOVER website. URL: https://discover-studie.de/rueckmeldung [accessed 2023-10-21]

23. Rohde F, Franke M, Sehili Z, Lablans M, Rahm E. Optimization of the Mainzelliste software for fast privacy-preserving record linkage. J Transl Med. 2021;19(1):33. [FREE Full text] [doi: 10.1186/s12967-020-02678-1] [Medline: 33451317]

24. Lowe B, Spitzer RL, Zipfel S. Gesundheitsfragebogen für Patienten (PHQ-D). Manual und Testunterlagen (Health questionnaire for patients (PHQ-D). Manual and test documents). Karlsruhe. Pfizer; 2002.

25. Negeri ZF, Levis B, Sun Y, He C, Krishnan A, Wu Y, et al. Depression Screening Data (DEPRESSD) PHQ Group. Accuracy of the Patient Health Questionnaire-9 for screening to detect major depression: updated systematic review and individual participant data meta-analysis. BMJ. 2021;375:n2183. [FREE Full text] [doi: 10.1136/bmj.n2183] [Medline: 34610915]

26. Du N, Yu K, Ye Y, Chen S. Validity study of patient health questionnaire-9 items for internet screening in depression among Chinese university students. Asia Pac Psychiatry. 2017;9(3). [FREE Full text] [doi: 10.1111/appy.12266] [Medline: 28856843]

27. Erbe D, Eichert H, Rietz C, Ebert D. Interformat reliability of the patient health questionnaire: validation of the computerized version of the PHQ-9. Internet Interv. 2016;5:1-4. [FREE Full text] [doi: 10.1016/j.invent.2016.06.006] [Medline: 30135800]

28. S3-Leitlinie Nationale VersorgungsLeitlinie Unipolare Depression. AWMF Leitlinienregister. 2022. URL: https://register.awmf.org/de/leitlinien/detail/nvl-005 [accessed 2024-04-02]

29. Seeralan T, Härter M, Koschnitzke C, Scholl M, Kohlmann S, Lehmann M, et al. Patient involvement in developing a patient-targeted feedback intervention after depression screening in primary care within the randomized controlled trial GET.FEEDBACK.GP. Health Expect. 2021;24 Suppl 1(Suppl 1):95-112. [FREE Full text] [doi: 10.1111/hex.13039] [Medline: 32286005]

30. Kohlmann S, Lehmann M, Eisele M, Braunschneider L, Marx G, Zapf A, et al. Depression screening using patient-targeted feedback in general practices: study protocol of the German multicentre GET.FEEDBACK.GP randomised controlled trial. BMJ Open. 2020;10(9):e035973. [FREE Full text] [doi: 10.1136/bmjopen-2019-035973] [Medline: 32958483]

31. Beesdo-Baum K, Zaudig M, Wittchen HU. First MB, Williams JBW, Karg RS, Spitzer RL, editors. SCID-5-CV Strukturiertes Klinisches Interview für DSM-5-Störungen Klinische Version: Deutsche Bearbeitung des Structured Clinical Interview for DSM-5 Disorders (SCID-5-CV Structured Clinical Interview for DSM-5 Disorders Clinical Version: German version of the Structured Clinical Interview for DSM-5 Disorders). Göttingen, Germany. Hogrefe; 2019.

32. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the Patient Health Questionnaire-9. Med Care. 2004;42(12):1194-1201. [doi: 10.1097/00005650-200412000-00006] [Medline: 15550799]

33. Broadbent E, Petrie KJ, Main J, Weinman J. The brief illness perception questionnaire. J Psychosom Res. 2006;60(6):631-637. [doi: 10.1016/j.jpsychores.2005.10.020] [Medline: 16731240]

34. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol. 1991;59(1):12-19. [doi: 10.1037//0022-006x.59.1.12] [Medline: 2002127]

35. McMillan D, Gilbody S, Richards D. Defining successful treatment outcome in depression using the PHQ-9: a comparison of methods. J Affect Disord. 2010;127(1-3):122-129. [doi: 10.1016/j.jad.2010.04.030] [Medline: 20569992]

36. Phillips R, Hazell L, Sauzet O, Cornelius V. Analysis and reporting of adverse events in randomised controlled trials: a review. BMJ Open. 2019;9(2):e024537. [FREE Full text] [doi: 10.1136/bmjopen-2018-024537] [Medline: 30826796]

37. Zou G. A modified poisson regression approach to prospective studies with binary data. Am J Epidemiol. 2004;159(7):702-706. [doi: 10.1093/aje/kwh090] [Medline: 15033648]

38. Gallis J, Turner EL. Relative measures of association for binary outcomes: challenges and recommendations for the global health researcher. Ann Glob Health. 2019;85(1):137. [FREE Full text] [doi: 10.5334/aogh.2581] [Medline: 31807416]

39. Knol MJ, Le Cessie S, Algra A, Vandenbroucke JP, Groenwold RH. Overestimation of risk ratios by odds ratios in trials and cohort studies: alternatives to logistic regression. CMAJ. 2012;184(8):895-899. [FREE Full text] [doi: 10.1503/cmaj.101715] [Medline: 22158397]

40. Boutron I, Altman DG, Moher D, Schulz KF, Ravaud P. CONSORT statement for randomized trials of nonpharmacologic treatments: a 2017 update and a CONSORT extension for nonpharmacologic trial abstracts. Ann Intern Med. 2017;167(1):40-47. [doi: 10.7326/m17-0046]

41. Eysenbach G. CONSORT-EHEALTH: implementation of a checklist for authors and editors to improve reporting of web-based and mobile randomized controlled trials. Stud Health Technol Inform. 2013;192:657-661. [Medline: 23920638]

42.   Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340(1):c869. [FREE Full text] [doi: 10.1136/bmj.c869] [Medline: 20332511]

43.   Montgomery P, Grant S, Mayo-Wilson E, Macdonald G, Michie S, Hopewell S, et al. CONSORT-SPI Group. Reporting randomised trials of social and psychological interventions: the CONSORT-SPI 2018 extension. Trials. 2018;19(1):407. [FREE Full text] [doi: 10.1186/s13063-018-2733-1] [Medline: 30060754]

44.   Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Med. 2010;8(1):18. [FREE Full text] [doi: 10.1186/1741-7015-8-18] [Medline: 20334633]

45.   Cuijpers P, Karyotaki E, Ciharova M, Miguel C, Noma H, Furukawa TA. The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: a meta-analysis. Acta Psychiatr Scand. 2021;144(3):288-299. [FREE Full text] [doi: 10.1111/acps.13335] [Medline: 34107050]

46.   Gould MS, Marrocco FA, Kleinman M, Thomas JG, Mostkoff K, Cote J, et al. Evaluating iatrogenic risk of youth suicide screening programs: a randomized controlled trial. JAMA. 2005;293(13):1635-1643. [doi: 10.1001/jama.293.13.1635] [Medline: 15811983]

47.   Robinson J, Pan Yuen H, Martin C, Hughes A, Baksheev GN, Dodd S, et al. Does screening high school students for psychological distress, deliberate self-harm, or suicidal ideation cause distress—and is it acceptable? An Australian-based study. Crisis. 2011;32(5):254-263. [doi: 10.1027/0227-5910/a000087] [Medline: 21940259]

48.   Löwe B, Scherer M, Braunschneider L, Marx G, Eisele M, Mallon T, et al. Clinical effectiveness of patient-targeted feedback following depression screening in general practice (GET.FEEDBACK.GP): an investigator-initiated, prospective, multicentre, three-arm, observer-blinded, randomised controlled trial in Germany. Lancet Psychiatry. 2024;11(4):262-273. [doi: 10.1016/s2215-0366(24)00035-x]

49.   Kuo WH, Gallo JJ, Eaton WW. Hopelessness, depression, substance disorder, and suicidality—a 13-year community-based study. Soc Psychiatry Psychiatr Epidemiol. 2004;39(6):497-501. [doi: 10.1007/s00127-004-0775-z] [Medline: 15205735]

50.   Na PJ, Yaramala SR, Kim JA, Kim H, Goes FS, Zandi PP, et al. The PHQ-9 Item 9 based screening for suicide risk: a validation study of the Patient Health Questionnaire (PHQ)-9 Item 9 with the Columbia Suicide Severity Rating Scale (C-SSRS). J Affect Disord. 2018;232:34-40. [doi: 10.1016/j.jad.2018.02.045] [Medline: 29477096]

51.   Rossom RC, Coleman KJ, Ahmedani BK, Beck A, Johnson E, Oliver M, et al. Suicidal ideation reported on the PHQ9 and risk of suicidal behavior across age groups. J Affect Disord. 2017;215:77-84. [FREE Full text] [doi: 10.1016/j.jad.2017.03.037] [Medline: 28319695]

52.   Erford BT, Jackson J, Bardhoshi G, Duncan K, Atalay Z. Selecting suicide ideation assessment instruments: a meta-analytic review. Meas Eval Couns Dev. 2017;51(1):42-59. [doi: 10.1080/07481756.2017.1358062]

53.   Beck A, Hamel C, Thuku M, Esmaeilisaraji L, Bennett A, Shaver N, et al. Screening for depression among the general adult population and in women during pregnancy or the first-year postpartum: two systematic reviews to inform a guideline of the Canadian Task Force on Preventive Health Care. Syst Rev. 2022;11(1):176. [FREE Full text] [doi: 10.1186/s13643-022-02022-2] [Medline: 35996176]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
DSM-V: Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)
IPQ: Illness Perception Questionnaire
ITT: intention-to-treat
PHQ-9: 9-item Patient Health Questionnaire
RCI: reliable change index
RR: relative risk
SCID: Structured Clinical Interview for DSM-V Disorders

XSL•FO
RenderX