



Model of integrated mental health video consultations for people with depression or anxiety in primary care (PROVIDE-C): assessor masked, multicentre, randomised controlled trial

Markus W Haun,^{1 2} Justus Tönnies,¹ Mechthild Hartmann,^{1 2} Alina Wildenauer,¹ Michel Wensing,³ Joachim Szecsenyi,³ Manuel Feißt,⁴ Moritz Pohl,⁴ Markus Vomhof,^{5 6} Andrea Icks,^{5 6} Hans-Christoph Friederich^{1 2}

For numbered affiliations see end of the article

Correspondence to:

M W Haun markus.haun@med.uni-heidelberg.de (ORCID 0000-0003-1851-3747)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2024;386:e079921

<http://dx.doi.org/10.1136/bmj-2024-079921>

Accepted: 21 August 2024

ABSTRACT

OBJECTIVE

To evaluate whether an integrated mental health video consultation approach (PROVIDE model) can improve symptoms compared with usual care in adults with depression and anxiety disorders attending primary care.

DESIGN

Assessor masked, multicentre, randomised controlled trial (PROVIDE-C).

SETTING

In 29 primary care practices in Germany, working remotely online from one trial hub.

PARTICIPANTS

376 adults (18-81 years) who presented to their general practitioner (GP) with depression or anxiety, or both.

INTERVENTION

Participants were randomised (1:1) to receive the PROVIDE model (n=187) or usual care (n=189). Usual care was provided by GPs through interventions such as brief counselling and psychotropic medication prescriptions and may or may not have included referrals to mental health specialists. The PROVIDE model comprised transdiagnostic treatment provided through five real-time video sessions between the patient at the primary care practice and a mental health specialist at an offsite location.

MAIN OUTCOME MEASURES

The primary outcome was the absolute change in the mean severity of depressive and anxiety symptoms

measured using the patient health questionnaire anxiety and depression scale (PHQ-ADS) at six months, in the intention-to-treat population. Secondary outcomes, measured at six and 12 months, included PHQ-ADS subscores, psychological distress related to somatic symptoms, recovery, health related quality of life, quality and patient centredness of chronic illness care, and adverse events.

RESULTS

Between 24 March 2020 and 23 November 2021, 376 patients were randomised into treatment groups. Mean age was 45 years (standard deviation (SD) 14), 63% of the participants were female, and mean PHQ-ADS-score was 26 points (SD 7.6). Compared with usual care, the PROVIDE intervention led to improvements in severity of depressive and anxiety symptom (adjusted mean change difference in the PHQ-ADS score -2.4 points (95% confidence interval -4.5 to -0.4), P=0.02) at six months. The effects were sustained at 12 months (-2.9 (-5.0 to -0.7), P<0.01). No serious adverse events were reported in either group.

CONCLUSIONS

Through relatively low intensity treatment, the PROVIDE model led to a decrease in depressive and anxiety symptoms with small effects in the short and long term. Depression and anxiety disorders are prevalent and therefore the small effect might cumulatively impact on population health in this population.

TRIAL REGISTRATION

ClinicalTrials.gov NCT04316572.

Introduction

Globally, depression and anxiety disorders are among the top leading causes of years lived with disability in both sexes.¹ Primary care physicians can effectively treat many patients with depression and anxiety, but some individuals require specialised mental health care. Despite the availability of effective treatment options, most people with depression and anxiety disorders do not have access to specialised mental health care.^{2 3} This issue is primarily linked to the fact that services are not available, do not have capacity, or are unaffordable.⁴ Especially in rural and remote areas, people are hindered by transportation barriers and refrain from seeking help because of widespread stigma.⁵ Moreover, the ageing population entails an increasing number of individuals dealing with multiple health conditions, including mental disorders. These

WHAT IS ALREADY KNOWN ON THIS TOPIC

Previous studies, predominantly conducted by the United States Department of Veterans Affairs services, have shown the efficacy of telemedicine based collaborative care in primary care settings

Evidence is scarce for consultation models of integrated mental health videos that feature remote consultations between patients located at primary care settings and offsite mental health specialists

WHAT THIS STUDY ADDS

The PROVIDE model, an integrated mental health video consultation transdiagnostic treatment approach, was shown to be effective in reducing symptoms of depression and anxiety in the short and long term

Although the effect size is small, the improvement is meaningful given the prevalence of these disorders in community settings

The PROVIDE model shows promise as a scalable intervention that can collectively benefit population health in terms of depression and anxiety disorders

challenges require increased coordination of long term integrated care for complex multimorbidity.

According to the World Health Organization, primary care settings are considered the most suitable setting for treating patients with mental disorders.⁶ Most patients, even those with severe and chronic conditions, are treated solely by general practitioners (GPs) who provide effective care.⁷ However, a large number of individuals affected by depression and anxiety disorders are not identified, do not receive adequate treatment, or are simply in urgent need of specialised treatment.⁸ Most GPs prioritise providing holistic care for both physical symptoms and psychological distress. However, the so-called somatising effect, whereby mental or emotional factors can manifest in physical conditions, can lead healthcare professionals within primary care settings to focus more on evaluating and addressing physical symptoms.^{9,10}

In recent years, the integration of mental health care into primary care settings has been pursued.¹¹ One major approach features a referral model that involves the colocation of on-site mental health specialists in primary care settings. In this approach, GPs refer distressed patients to mental health specialists who assume the primary responsibility for the psychosocial management of the presenting problem.¹² In 2020, the results from a systematic review of 15 studies showed that co-located specialty care was associated with mental health benefits, and concluded that more rigorous randomised controlled trials are needed.¹³ Recently, virtual colocation (ie, live interactive videoconferencing) has been proposed to enrich integrated primary care.¹⁴ The limited number of published randomised controlled trials to date were conducted in highly regulated environments, such as the US Veterans Health Care Administration, or involved patients from inpatient facilities.^{15,16} The potential scalability of these models to primary care settings, particularly in countries where smaller, single handed, or rural and remote practices dominate, remains uncertain.

To explore this potential, we previously developed and successfully piloted a scalable integrated mental health video consultation model, designed to be implemented in primary care settings. This model, which targets patients with depression and anxiety, allows mental health specialists to be virtually embedded in primary care through co-located specialty care.¹⁷ The aim of this assessor masked randomised controlled trial was to investigate the effectiveness of this new mental health service model for treating people with depression or anxiety, or both, in primary care settings. We hypothesised that people randomly assigned to receive the intervention would lead to greater reductions in symptoms of depression and anxiety at six months compared with those assigned to receive treatment as usual.

Methods

Trial design and setting

This study was a multicentre, stratified, assessor masked, parallel group, randomised controlled trial

(the PROVIDE-C trial) with 1:1 allocation of patients and conducted in Germany. We recruited patients from 29 primary care practices in the federal states of Baden-Wuerttemberg and Rhineland-Palatinate (overall population 15.2 million; overall area size 55 605 km²). In Germany, GPs receive regionally negotiated fee-for-service payments from sickness funds up to a maximum number of services per quarter. Typically, no gatekeeping process exists, meaning that patients can directly access the services of a GP without prior registration (ie, a free access system). Although covered by all sickness funds, video consultations are not common, however, they increased during and after the covid-19 pandemic, that is, after commencement of the trial. The trial was approved by the Medical Faculty of the University of Heidelberg Ethics Committee (S-923/2019), and its protocol is available online.¹⁸ We reported the PROVIDE-C trial in accordance with the CONSORT 2010 statement.

Participants and recruitment

Patients were eligible if they had at least one of the following mental health conditions: (1) at least moderately severe depression, defined as a patient health questionnaire-9 (PHQ-9) score of 10 points or greater with item one or two being endorsed (5-9 mild, 10-14 moderate, 15-19 moderately severe, and 20-27 severe depressive symptoms); (2) at least moderately severe general anxiety, defined as a generalised anxiety disorder scale (GAD-7) score of 10 points or greater (5-9 mild, 10-14 moderate, 15-21 severe anxiety symptoms); or (3) a combined anxiety and depression score (patient health questionnaire anxiety and depression scale, PHQ-ADS) of 12 points or greater, had received no or insufficient treatment (psychotherapy, psychopharmacotherapy, or both); agreed to participate in the trial by providing written informed consent; were capable of providing consent, and were aged 18 years or older.

Patients were excluded if they (1) had substance misuse or dependence that was likely to compromise intervention adherence (identified during an unstructured assessment as part of the screening), (2) were acutely suicidal or put others at risk (PHQ-9 Item 9 endorsed and positive structured suicide screening result), (3) needed emergency medical treatment, such as, hospital admission (as assessed by the referring GP), (4) had acute psychotic symptoms, such as, persecutory delusions or thought insertion (identified during an unstructured assessment as part of the screening), (5) had severe cognitive impairment or dementia (as assessed by the referring GP), (6) had substantial hearing or visual impairment (as assessed by the referring GP), (7) were pregnant and in the second or third trimester (as assessed by the referring GP), (8) showed insufficient German language proficiency (identified during an unstructured assessment as part of the screening), (9) or had prior experience with video consultations through participation in the PROVIDE-B feasibility trial.¹⁷

GPs recruited patients during their regular clinic hours or by calling them at home. Based on their clinical

judgement, the GPs selected individuals suspected of having depression or anxiety, introduced the trial to them, obtained consent from them for screening, and referred them to the study team for screening. We also instructed GPs via weekly reminders to review their electronic health records to recruit potentially eligible patients. Moreover, we shifted to the research staff in the trial centre all time consuming tasks (eg, assessing eligibility, explaining the trial in detail, addressing patients' questions, obtaining consent, randomising patients, and collecting baseline data) that could have deterred GPs from recruiting patients with the routine consultations during busy times (eg, Monday mornings, holidays, and peak phases of the covid-19 pandemic).¹⁹ We obtained signed informed consent forms from all participants before performing the baseline assessment.

Randomisation and masking

We collected baseline measurements immediately before randomisation. Eligible participants were then randomly assigned (1:1) to the intervention or control group via a secure web based randomisation system (Randomiser V.2.0.2) operated by a data manager who was not involved in patient recruitment, centrally at the Institute of Medical Biometry, Heidelberg University. The treatment sequence was a computer generated sequence of random numbers and was stratified by centre (primary care practice) and symptom severity at baseline as measured by the PHQ-ADS (three levels of symptom severity with scores of 10, 20, and 30 points indicating mild, moderate, and severe depression and (10-19 mild, 20-29 moderate, 30-48 severe)) using randomly permuted block sizes of 2 and 4. While the patients, GPs, and mental health specialists were aware of the intervention assignment after allocation, the data analysts were masked to the allocation. While the patients reported outcomes in assessments after baseline through computer assisted telephone interviews, we masked the interviewers who questioned the patients and completed the questionnaire on their behalf. Specifically, we ensured that these interviewers were not present when discussing individual patients and avoided mentioning any patient names or assigned treatments. When scheduling the interviews with the patients, research assistants who were not involved in conducting the interviews instructed patients not to mention to the interviewers which group they belonged to.

Procedures

Intervention group

The PROVIDE intervention aimed to reduce severity of depressive and anxiety symptoms by integrating specialised mental health care into primary care practices via video consultations.

We hypothesised that the PROVIDE intervention would increase in the (virtual) availability of mental health specialists, leverage patients' familiarity with the primary care practice for treatment engagement, and draw on a transdiagnostic treatment approach that

combined elements from problem solving treatment. Transdiagnostic treatment has been shown to yield moderate effects in alleviating depression and anxiety in primary care patients with a focus on building a strong working alliance. This approach has been promoted as a crucial element of manuals achieving high acceptability for both patients and clinicians.¹²

The five core components of the PROVIDE intervention were as follows: video consultations for primary care patients conducted by mental health specialists; five consultations over a period of approximately eight weeks; diagnostic clarification and case formulation plus stepped care, based on interim symptom monitoring using the PHQ-ADS after the third consultation; brief psychotherapy focused on therapeutic alliance and affect expression and regulation; and case supervision in a biweekly group format, led by a senior consultant specialising in psychiatry and psychosomatic medicine.^{22 23}

Following a stage model of psychotherapy manual development, we initially created a stage I intervention manual that outlined the treatment techniques, goals, and format for the PROVIDE-B feasibility trial.^{24 25} For the PROVIDE-C trial, we refined this manual into a full stage II intervention manual (supplementary material 1).

Materials used in the intervention

We provided the primary care practices with widescreen (12.3 inch) computer tablets and a handbook outlining the trial, its procedures, and feasible contingency plans in case of connectivity failures; mental health specialists were provided with the stage II intervention manual.

Main consultation-by-consultation

In the first consultation, the mental health specialist began by establishing a strong working alliance by inviting the patient to talk freely about the central problem and trying to understand the patient's concerns and symptoms. To derive a case formulation, the mental health specialist also gathered diagnostic information, such as through probing or applying established psychometric practises. The mental health specialist informed the patient about their condition (depression or anxiety) and the available treatment options. In the second consultation, the mental health specialist facilitated affect experience or expression by introducing the concept of emotional mindfulness and encouraging the patient to practise it between consultations.²⁶ In the third consultation, the mental health specialist maximised their effort in supporting the patient in experiencing and expressing their (avoided) affects. Then, the mental health specialist aimed to link the patient's narrative and the related affects to the central problem and, more importantly, to more adaptive responses. After the third consultation, the study team determined the patient's PHQ-ADS score to monitor their progress, and the results were sent to the mental health specialist. In the fourth consultation, the mental

health specialist and the patient reflected on the symptom severity as indicated by the PHQ-ADS score and the need for prolonged care, that is, the need for referral to more intensive specialised treatment. In the final session, the mental health specialist and the patient reviewed the treatment process and developed the next steps, which, if needed, may have included a recommendation for additional intensive specialised treatment. After the final session, the mental health specialist compiled a one page case summary that comprised recommendations for the GP on how to proceed with the patient's care.

Mode of delivery

The intervention was delivered by 22 mental health specialists (21 psychologists in advanced psychotherapy training and one physician with a licence for practising psychotherapy) trained in the PROVIDE model. To standardise the intervention, the physician was not allowed to prescribe medication directly. While all mental health specialists were allowed to discuss psychopharmacological treatment with the GPs, the initiation of psychopharmacological treatment was at the GPs' discretion. Before their first video consultation, all mental health specialists received a three hour introductory training session on the trial procedures, the intervention manual, and the videoconferencing platform.

The intervention was delivered through individual, synchronous one-to-one video consultations conducted via an encrypted, web based videoconferencing platform on a subscription basis (arztkonsultation ak GmbH, Schwerin, Germany, <https://arztkonsultation.de>). The patient was in a room of the primary care practice designated for the video consultations to ensure confidentiality, while the mental health specialist was at an offsite location.

The five consultations lasted 50 min each and were held at biweekly intervals over a period of approximately eight weeks. The consultations happened at fixed time slots, on which the primary care practice staff and the mental health specialists agreed.

Tailoring

The mental health specialists were aware of technical limitations such as non-muting microphones, poor visual definition, impaired audio, and speech delay. When the broadband connection was unstable, or the quality of the audio was poor, the mental health specialists switched to a phone call to avoid misinterpretations and frustration with the technology. For acute crises (eg, suicidal ideation, medical emergency, or violence), the mental health specialists had the contact information of the patients' significant others (those physically close to the patient) on hand and contacted the nearest safe and emergency care locations.

Modifications during the trial

We modified the design after the trial commenced: patients at an increased risk of covid-19 complications

in the intervention group received their video consultations during their lockdown period at home.

Intervention fidelity

We evaluated intervention fidelity, that is, the degree to which core components were delivered by providers and understood by participants as planned. Specifically, we computed a predefined intervention integrity score by applying the criteria specified in the statistical analysis plan (supplementary material S2) and the trial protocol.¹⁸ Moreover, we determined the proportion of video consultations that were conducted as planned.

Treatment as usual arm

For patients allocated to the control group, treatment as usual involved usual care provided by their GP. This may or may not have included interventions such as brief counselling, the prescription of psychotropic medication, and referrals to mental health specialists such as office based psychiatrists and psychotherapists or mental health clinics. Patients in the treatment as usual group needed to make appointments with their GP ad hoc. In both groups, participants were permitted to continue any treatment they had been receiving at the time of enrolment in the trial.

Outcomes

The primary outcome was the absolute change in the mean severity of depressive and anxiety symptoms measured using the PHQ-ADS from the baseline assessment to six months after the baseline assessment. For the sake of comparability, we selected six months because this timepoint is one of the most common for evaluating primary outcomes in integrated mental health care trials.^{15 16 27 28} The PHQ-ADS is a 16-item scale (all nine items of the PHQ-9 and all seven items of the GAD-7, scored from 0=not at all to 3=nearly every day). Higher scores indicated more severe symptoms. The PHQ-ADS is a psychometrically validated measure used in primary care settings.²⁹ It has shown effectiveness and sensitivity as an outcome measure in treatment trials, with a recommended minimal important difference of 3 to 5 points.³⁰ In the multicentre PROVIDE-C trial, outcomes were centrally assessed, including the PHQ-ADS scores.

Secondary outcomes included differences in the absolute changes in mean severity of depressive and anxiety symptoms on the PHQ-ADS at 12 months between the two groups. At six and 12 months, we also investigated: absolute changes in the mean severity of depressive (using PHQ-9) and anxiety (using GAD-7) symptoms; mean score for psychological distress related to somatic symptoms (somatic symptom disorder-B criteria scale, SSD-12); mean score on the personal confidence and hope, goal, and success orientation, willingness to ask for help, reliance on others, and no domination by symptoms subscales (five subscales of the recovery assessment scale, RAS-G); health related quality of life (12 item short-form health survey, SF-12); and quality and patient centredness

of chronic illness care score (patient assessment of chronic illness care-short form, PACIC-short form). At six and 12 months, we also evaluated differences in health service use between the groups by applying the questionnaire for the assessment of medical and non-medical resource use in mental disorders. However, findings on health service use and cost effectiveness, based on the scores on the EuroQol 5 dimension 5 level (EQ-5D-L), will be reported in a separate publication. We collected information about harmful outcomes from all randomly assigned participants. Patients, GPs, and mental health specialists in both groups had to immediately report harmful outcomes including severe adverse events, defined as life threatening and fatal events (eg, suicide attempts, death by suicide, and reported violence), to the trial team. For more detailed insights into possible adverse consequences of the intervention, we systematically assessed prespecified harmful outcomes in the intervention group by applying the inventory for the assessment of negative effects of psychotherapy during follow-up visits. Finally, we conducted a cost effectiveness analysis, which we will report in a separate publication that will include the trial registration number and findings on health service use.

Data collection and retention management

We collected participant data from the intervention and control arms at baseline immediately before randomisation. Follow-up measurements were conducted at six and 12 months after the baseline assessment. Specifically, we conducted computer assisted telephone interviews during which patients reported outcomes on validated questionnaires. We captured these outcomes by applying an online survey tool (the Enterprise Feedback Suite Survey, Questback GmbH). Retention management followed a standard operating procedure: we scheduled interviews for follow-up measurements contacting participants up to eight times. We then reminded participants about the scheduled interview via text message (SMS) 24 hours before the anticipated interview. If no interviews could be scheduled or conducted, we sent the questionnaires together with a paid return envelope to the people who had not responded and attempted to call them eight more times.

Sample size

To detect the minimal clinically important difference in the PHQ-ADS score of 3 points (SD 9 points) with a two sided 5% significance level and a power of 80%, a sample size of 160 patients per group was necessary. This size adjusted for the correlation between the baseline value and the change from the baseline value ($r=0.35$) and given an anticipated dropout rate of 20%. We expected an 18 month inclusion period to recruit these patients.

Statistical analysis

Initially, we compared sociodemographic and medical characteristics of the eligible patients enrolled in the

trial to the eligible patients who declined to participate or with whom we lost contact. Specifically, we did a logistic regression using trial participation status (yes/no) as the dependent variable.

The primary analysis used data from the intention-to-treat population, which included all patients in the group to which they were allocated by randomisation. We investigated the missing-at-random assumption using a description of the covariates grouped by missing data on at least one PHQ-ADS item at six months versus no missing data (ie, a second baseline table, a method to identify the potential for bias due to missing data is to compare participants with and without missing values). Missing data for the primary and secondary endpoints were replaced using multiple imputation at the item level (10 imputations with a maximum of 20 iterations per imputation and a fixed seed). Before analysis, assumptions for mixed linear modelling (normality of residuals, linearity, homogeneity, and extreme outliers) were evaluated graphically. We analysed the primary outcome with a mixed linear model, where a random intercept accounted for the primary care practice to which the patient belonged. The patient specific baseline variables, which are established predictors for symptom severity in primary care, were included as fixed effects in the model: age, gender, presence of a chronic physical disease, physical health (SF-12 physical component score), history of depression or anxiety, baseline PHQ-ADS score, trial group, and the number of days between the baseline assessment and randomisation.³¹ The treatment effect was quantified by the parameter estimate of the group together with the respective 95% Wald confidence interval. We supplemented the findings for the primary outcome with analyses of the intention-to-treat complete case dataset (missing data not imputed), the per protocol dataset (data from participants who followed the protocol, excluding their data if they do not adhere), the as treated dataset (considering the treatment actually received by the participant, without regard to adherence to their randomisation assignment). Additionally, we analysed a sensitivity dataset, which was a subset of the as treated dataset for which any form of psychological treatment in the control group did not lead to the exclusion of the patient from the analysis. During the review process for this paper, after the protocol was developed, we decided to report findings for both the primary and the secondary outcomes from the minimally adjusted model adjusting only for primary care practice and the baseline PHQ-ADS score. We decided to perform a responder analysis comparing the proportion of participants within each study arm who had a change at least as large as the minimal clinically important difference as part of the secondary analyses.³² Specifically, we calculated the multiple imputation pooled estimator for the difference between the two study arms in terms of the percentage of patients who met the minimal clinically important difference. The analyses of the secondary endpoints were exploratory and were conducted analogously to those of the primary endpoint. To assess the

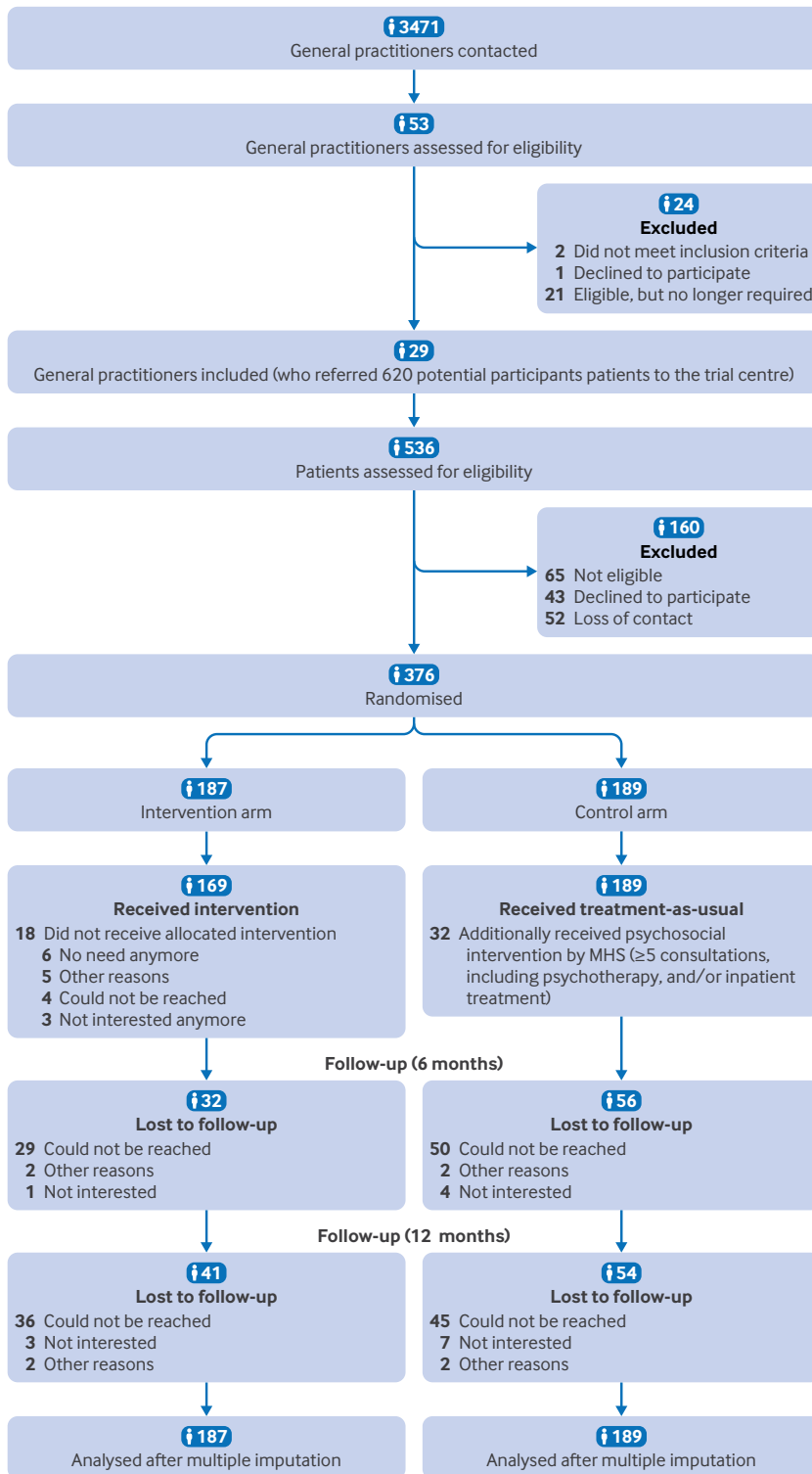


Fig 1 | Flow of participant inclusion in study. MHS=mental health specialists.

consistency of the observed effects across subgroups of patients defined by baseline characteristics, we performed nine prespecified subgroup analyses based on marital status, education level, employment status, psychiatric treatment or psychotherapy status at baseline, history of psychiatric treatment or psychotherapy, psychopharmacological treatment

status at baseline, history of psychopharmacological treatment, willingness to accept psychotherapy, and willingness to accept psychopharmacological treatment. For all subgroup analyses, we incorporated an interaction term between group and subgroup in the mixed linear regression models to analyse the primary endpoint. However, the trial was not powered to detect subgroup differences.

We prespecified the analyses in a statistical analysis plan before database lock (supplementary material S2). The analyses were performed using R 4.4.0 or higher. We registered the trial with ClinicalTrials.gov, NCT04316572, on 20 March 2020, before any participant was recruited.

Patient and public involvement

During the planning phase of the study, we involved two patient representatives (one female, one male) who had participated in video consultations in the PROVIDE-B feasibility trial. Specifically, the patient representatives participated in the conceptualisation of the trial procedures and materials. The selection of the outcome measures was informed by the representatives' priorities and experiences. They also revised the draft versions of this study protocol and all trial materials including information sheets, consent materials, and questionnaire sets with respect to clarity and understanding from the service user perspective. The patient representatives were not involved in the plans for participant recruitment. We continuously discussed the progress of the trial with these two patient representatives. Both patient representatives were compensated for their expenses. We will report the burden of the intervention on the patients' quality of life and health in a separate publication on the qualitative process evaluation of the trial.

Results

Sample description

Overall, we invited 3471 GPs to participate in the trial. Fifty three eligible GPs followed up our invitation, while the reasons for non-response among the others is unknown. We included 29 GPs in the trial, while the remaining 24 were not included given that in the meantime the target sample was reached. Between 24 March 2020, and 23 November 2021, GPs referred 620 potential participants to the trial and 536 were identified as potentially eligible. A total of 160 patients were excluded after the screening interview and baseline assessment (fig 1).

The eligibility fraction, which was the proportion of potential participants who underwent screening (including people who were referred but could not be screened) and were eligible for enrolment, was 76% (471/620). The enrolment fraction, which was the proportion of people who were eligible for participation and were enrolled, was 80% (376/471). The recruitment fraction, which was as the proportion of potential participants who enrolled (61% (376/620)), and the number of patients needed to be screened (1.6(620/376)). In total, 376 participants

were enrolled from 29 primary care practices and were randomly assigned to receive the integrated mental health video consultation model (n=187) or treatment as usual (n=189). Per general practitioner, a median of 10 patients were enrolled (range 1-58).

Among all 471 eligible patients, patients who were single were less likely to participate in the trial than people in a partnership (P=0.02). Whether eligible patients participated in the trial was not predicted by other sociodemographic and medical characteristics (age, gender, education level, employment status, degree of managing with the available income, chronic physical disease status, patient knowledge of local mental health services, symptom severity score (Patient Health Questionnaire Anxiety and Depression Scale), and the probability of post-traumatic stress disorder (primary care PTSD screen for the Diagnostic and Statistical Manual of Mental Disorders, fifth edition). Given that we were committed to including all eligible patients who were referred to the trial but still waiting to be enrolled when we reached our planned sample size, the final sample size slightly exceeded the planned sample size of 320. Of the 376 participants, 238 (63%) participants were female. The mean age was 45 years (SD 14; range 18-81). A total of 220 (59%) participants had at least one chronic physical disease. At baseline, 183 (49%) participants had never received any psychiatric treatment or psychotherapy. A total of 153 (41%) of the 376 participants had a symptom of moderate severity and 139 (37%) had a symptom of severe severity. The sociodemographic and medical characteristics were similar between the intervention group and the control group (table 1).

Intervention fidelity

Concerning intervention fidelity, 906 video consultations were completed in which 169 (90%) of 187 participants in the intervention group completed all five planned consultations. In 82 (9%) of these 906 video consultations, the patient was located at home due to being at an increased risk for covid-19 complications. Participants received a median of five consultations (range 0-5). In the intervention group, 172 (92%) participants were regarded as adherent based on the intervention integrity score. During the intervention period, 18 (10%) participants in the intervention group stopped treatment. Regarding potential contamination bias, we identified nine (5%) participants in the control group who had received some form of video consultation during the intervention period. Based on the description of the covariates grouped by missing data for at least one PHQ-ADS item at six months versus no missing data, we retained the missing at random assumption and proceeded with multiple imputation (supplementary material S3).

Primary outcome

At six months, 155 (83%) of 187 participants in the intervention group and 133 (70%) of 189 participants in the control group were followed up. Specifically,

we found a significant difference in the proportion of participants with missing data for at least one PHQ-ADS item at six months between the intervention group and the control group (P=0.01). Unintentional unmasking of outcome assessors occurred during 20% (59/288) of all interviews at six months. For the primary outcome, the mean change in the PHQ-ADS score was -9.2 points (95% CI -10.7 to -7.7) in the intervention group and -7.1 points (-8.8 to -5.4) in the control group (adjusted mean change difference -2.4 points (-4.5 to -0.4), P=0.02). The effect size (Cohen's d) was 0.21 (95% CI 0.03 to 0.39). No significant association between gender and the primary outcome was noted (P=0.95). Supplementary material S4 presents the parameter estimates calculated with the fully adjusted mixed effect model for the primary outcome based on the intention-to-treat dataset. Findings for the analyses based on the intention-to-treat complete case dataset (missing data not imputed), the per protocol dataset, the as treated dataset, and the sensitivity dataset did not indicate that the treatment effect was due to selection because they did not yield any significant differences between the trial groups. Table 2 presents findings for all analysis types applied to the primary outcome. The responder analysis compared the proportion of participants within each study arm who had a change at least as large as the minimal clinically important difference at six months. Our results showed that the proportion of participants in the intervention group who had an improvement at least as large as the minimal clinically important difference (median of 70.1%, range 69.0-72.7% for all 10 imputed datasets) was 6.8% (95% CI -4.5 to 18.1), P=0.23), which was greater than that in the control group (62.7%, range 60.8-67.2%). The intraclass correlation coefficient for the primary care practice as centre was 0.01.

Secondary outcomes

Follow-up results at 12 months were available for 146 (78%) of the 187 participants in the intervention group and 135 (71%) of the 189 participants in the control group. At 12 months, the mean change in the PHQ-ADS score (minimally adjusted model) was -11.0 points (95% CI -12.6 to -9.4) in the intervention group and -8.1 points (-9.8 to -6.4) in the control group (mean change difference -2.9 (-5.0 to -0.7), P=0.007). The effect size (Cohen's d) was 0.30 (95% CI 0.08 to 0.52). Table 3 presents the summary statistics and results from the mixed model analysis of the primary and secondary outcomes at all follow-ups.

For the secondary outcomes at six months (minimally adjusted model), we detected significantly greater improvements in severity of depressive symptoms (PHQ-9 score, Cohen's d 0.26 (95% CI 0.01 to 0.50)), psychological distress related to somatic symptoms (SSD-12 score, 0.24 (0.03 to 0.45)), health related quality of life (SF-12 mental component score, 0.18 (0.02 to 0.34)), and quality and patient centredness of chronic illness care (PACIC-short form, 0.39 (0.17 to 0.61)) in the intervention group compared with the control group. At 12 months in the minimally

Table 1 | Baseline characteristics of the intention-to-treat population

Characteristics	Control (n=189)	Intervention (n=187)	Total (n=376)
Age:			
Mean (SD)	47 (13)	44 (14)	45 (14)
Median (interquartile range)	50 (21)	44 (22)	47 (21)
Gender, n (%):			
Male	67 (35)	71 (38)	138 (37)
Female	122 (65)	116 (62)	238 (63)
Marital status, n (%):			
Single	58 (31)	49 (27)	107 (29)
In a partnership	129 (69)	134 (73)	263 (71)
Missing	2	4	6
Education level, n (%):			
No secondary general school leaving certificate	1	0	1 (1)
Secondary general school leaving certificate (9 years)	53	45	98 (26)
Certificate of 10 grade school of general education in the former German Democratic Republic/intermediate school leaving certificate (10 years)	65	78	143 (38)
Fachhochschule/University entrance qualification	66	64	130 (35)
Missing	4	0	4
Employment status, n (%):			
Vocational training	3 (2)	4 (2)	7 (2)
University studies	2 (1)	3 (2)	5 (1)
Employed/self-employed	95 (50)	102 (55)	197 (53)
Housework	5 (3)	5 (3)	10 (3)
Unemployed	17 (9)	11 (6)	28 (7)
On sick leave	52 (28)	44 (24)	96 (26)
On parental leave	2 (1)	4 (2)	6 (2)
Retired	13 (7)	13 (7)	26 (7)
Missing	0	1	1
Managing with available income, n (%):			
Easily	83 (45)	92 (49)	175 (47)
Not too bad	59 (32)	50 (27)	109 (29)
Difficult some of the time	28 (15)	29 (16)	57 (15)
Difficult all of the time	11 (6)	8 (4)	19 (5)
Impossible	5 (3)	7 (4)	12 (3)
Missing	3	1	4
Chronic physical disease(s):			
No	78 (41)	77 (41)	155 (41)
Yes	111 (59)	109 (59)	220 (59)
Missing	0	1	1
Current psychiatric treatment/psychotherapy, n (%):			
No	168 (89)	174 (93)	342 (91)
Yes	21 (11)	13 (7)	34 (9)
Past psychiatric treatment/psychotherapy, n (%):			
No	87 (52)	96 (55)	183 (54)
Yes	80 (49)	74 (43)	154 (45)
Declined to answer	1 (1)	4 (2)	5 (1)
Missing	21	13	34
Current psychopharmacological treatment, n (%):			
No	130 (69)	134 (72)	264 (71)
Yes	58 (31)	51 (28)	109 (29)
Missing	1	2	3
Past psychopharmacological treatment, n (%):			
No	89 (72)	94 (72)	183 (72)
Yes	35 (28)	36 (28)	71 (28)
Missing	65	57	122
Openness to psychotherapy, n (%):			
Strongly disagree	1 (1)	1 (1)	2 (1)
Disagree	10 (5)	5 (3)	15 (4)
Agree	29 (15)	30 (16)	59 (16)
Strongly agree	144 (76)	146 (78)	290 (77)
Declined to answer	5 (3)	5 (3)	10 (3)
Openness to psychopharmacological treatment, n (%):			
Strongly disagree	24 (13)	25 (13)	49 (13)
Disagree	58 (31)	56 (30)	114 (30)
Agree	38 (20)	43 (23)	81 (22)
Strongly agree	53 (28)	49 (26)	102 (27)
Declined to answer	16 (8)	14 (7)	30 (8)

(Continued)

Table 1 | Continued

Characteristics Control (n=189)		Intervention (n=187)	Total (n=376)
Severity of depressive and anxiety symptoms, mean (SD)*	26 (7.5)	27 (7.8)	26 (7.6)
Severity of depressive symptoms, mean (SD)†	14 (5.2)	15 (5.3)	14 (5.3)
Severity of anxiety symptoms, mean (SD)‡	12 (4.5)	12 (4.3)	12 (4.4)
Level of symptom severity, n (%):§			
Minimal	—	—	—
Mild	43 (23)	41 (22)	84 (22)
Moderate	79 (42)	74 (40)	153 (41)
Severe	67 (35)	72 (39)	139 (37)

*Patient health questionnaire anxiety and depression scale.
†Patient health questionnaire 9-item depression scale.
‡Generalised anxiety disorder scale.
§Patient health questionnaire anxiety and depression scale.

adjusted model, we found greater improvements in the intervention group than in the control group: PHQ-9 (0.20 (0.06 to 0.34)), GAD-7 (0.34 (CI 0.06 to 0.62)), and SSD-12 (0.39 (0.15 to 0.63)). Similar to the PHQ-ADS change scores, differences in the PHQ-9, GAD-7, and SSD-12 change scores, favouring the intervention group over the control group, were more pronounced at 12 months than at six months (table 3). At no time point did we identify any significant differences between the intervention group and the control group in recovery and health related quality of life (SF-12 physical component score). The prespecified subgroup analyses showed that depression and anxiety symptom severity, as the primary outcome, was equal across the different subgroups. No significant interactions were reported (supplementary material S5); however, the trial was not powered to detect subgroup differences.

Harms

No serious adverse events attributable to trial participation were reported in either group during the study. In the intervention group, self-reported data for negative effects, as measured using the inventory for the assessment of negative effects of psychotherapy at six months, were available for 81% (151/187) of

participants (supplementary material S6). In total, 28% (52/187) of participants in the intervention group reported at least one negative effect attributed to the intervention instead of their life circumstances (average number of adverse effects per patient 0.6 (SD 1.1)). Seventeen (9%) participants reported that they had experienced more “downs” than “ups” since the end of the intervention. Fifteen (8%) participants stated that they were anxious that their colleagues or friends would learn about their mental health treatment. No significant difference was noted in the frequency of reported negative effects between women (31/96, 32%) and men (21/55, 38%) (P=0.48).

Auxiliary treatment in the intervention group

For the intervention group, 67 (43%) of the 155 participants who were followed up at six months had received specialised mental health care in addition to the intervention. Specifically, 7% (11/155) of the participants had received inpatient mental health treatment and 37% (57/155) had received some form of outpatient care (psychiatric consultation, psychotherapy, or psychosocial counselling service) in addition to the PROVIDE intervention. Regular psychotropic medication was prescribed to 34%

Table 2 | Parameter estimates calculated with the mixed effect model for the primary outcome of the patient health questionnaire anxiety and depression scale

Analysis type	No. of people in control group	Control group, mean (SD)	No. of people in intervention group	Intervention group, mean (SD)	Minimally adjusted mixed model analysis (95% CI)*	P value	Effect size (minimally adjusted mixed model) (95% CI)	Adjusted mixed model analysis (95% CI)†	P value	Effect size (adjusted mixed model) (95% CI)
Baseline (six months)	189	26 (7.5)	187	27 (7.8)	—	—	—	—	—	—
Intention to treat	189	19 (9.8)	187	17 (10)	-2.43 (-4.48 to -0.38)	0.02	0.21 (0.03 to 0.39)	-2.26 (-4.28 to -0.23)	0.03	0.21 (0.02 to 0.39)
Intention to treat, complete case	112	18 (9.5)	125	17 (10)	-1.37 (-3.60 to 0.92)	0.23	0.08 (-0.05 to 0.21)	-1.45 (-3.62 to 0.79)	0.20	0.09 (-0.05 to 0.22)
Per protocol	85	17 (9.4)	97	17 (11)	-1.63 (-4.27 to 1.06)	0.23	0.09 (-0.06 to 0.24)	-2.00 (-4.54 to 0.56)	0.13	0.12 (-0.03 to 0.27)
As treated	93	17 (9.1)	144	18 (11)	-0.98 (-3.28 to 1.37)	0.41	0.05 (-0.08 to 0.18)	-1.25 (-3.50 to 1.05)	0.29	0.07 (-0.06 to 0.20)
Sensitivity dataset	124	18 (9.2)	78	17 (10)	-1.82 (-4.29 to 0.70)	0.15	0.10 (-0.04 to 0.24)	-2.03 (-4.45 to 0.42)	0.11	0.12 (-0.02 to 0.26)

Effect size measured with Cohen's d.
CI=confidence interval.

*Intervention v control condition; mixed effect linear regression model minimally adjusted for the respective baseline score and centre.

†Intervention v control condition; mixed effect linear regression models adjusted for the respective baseline score, age, gender, history of depression or anxiety, SF-12 Physical Component Score, chronic medical disease, and no. of days between baseline assessment and randomisation.

Table 3 | Summary statistics and results from the mixed model analysis of the secondary outcomes

Secondary outcomes	Control group (n=189)	Intervention group (n=187)	Minimally adjusted mixed model analysis (95% CI)*	P value	Effect size (minimally adjusted mixed model) (95% CI)	Adjusted mixed model analysis (95% CI)†	P value	Effect size (adjusted mixed model) (95% CI)	Patient health questionnaire anxiety and depression scale
Baseline	26 (7.5)	27 (7.8)	—	0.67	—	—	—	—	
12 months	18 (9.1)	15 (10)	-2.86 (-4.99 to -0.73)	0.01	0.30 (0.08 to 0.52)	-2.60 (-4.71 to -0.49)	0.02	0.29 (0.06 to 0.53)	
Patient health questionnaire 9-item depression scale									
Baseline	14 (5.2)	15 (5.3)	—	0.35	—	—	—	—	
Six months	10 (5.7)	9.1 (5.7)	-1.31 (-2.54 to -0.09)	0.04	0.26 (0.02 to 0.50)	-1.22 (-2.42 to -0.02)	0.05	0.24 (0.00 to 0.47)	
12 months	9.5 (5.3)	8.1 (5.8)	-1.31 (-2.26 to -0.36)	<0.01	0.20 (0.06 to 0.34)	-1.24 (-2.18 to -0.31)	0.01	0.19 (0.05 to 0.35)	
Seven item generalised anxiety disorder scale									
Baseline	12 (4.5)	12 (4.3)	—	0.97	—	—	—	—	
Six months	8.7 (4.9)	8 (5.1)	-0.91 (-1.95 to 0.13)	0.09	0.15 (-0.02 to 0.32)	-0.84 (-1.87 to 0.18)	0.11	0.14 (-0.03 to 0.32)	
12 months	8.2 (4.5)	7.1 (4.7)	-1.36 (-2.48 to -0.23)	0.02	0.34 (0.06 to 0.62)	-1.22 (-2.30 to -0.13)	0.03	0.30 (0.03 to 0.57)	
Somatic symptom disorder-B criteria scale									
Baseline	22 (11)	22 (12)	—	0.97	—	—	—	—	
Six months	19 (12)	18 (12)	-2.24 (-4.17 to -0.31)	0.02	0.24 (0.03 to 0.45)	-2.21 (-4.12 to -0.31)	0.02	0.24 (0.03 to 0.44)	
12 months	20 (11)	17 (11)	-3.56 (-5.76 to -1.36)	<0.01	0.39 (0.15 to 0.63)	-3.39 (-5.63 to -1.15)	<0.01	0.39 (0.13 to 0.65)	
12 item short form survey									
Physical component score§:									
Baseline	42 (10)	43 (11)	—	0.35	—	—	—	—	
Six months	43 (11)	45 (11)	1.26 (-0.67 to 3.20)	0.20	0.12 (-0.06 to 0.30)	1.13 (-0.83 to 3.09)‡	0.26	0.12 (-0.08 to 0.32)	
12 months	43 (11)	46 (11)	1.85 (-0.05 to 3.75)	0.06	0.13 (0.00 to 0.27)	1.51 (-0.37 to 3.40)‡	0.12	0.11 (-0.03 to 0.26)	
Mental component score:									
Baseline	30 (7.7)	29 (8.2)	—	0.36	—	—	—	—	
Six months	38 (11)	40 (11)	2.47 (0.23 to 4.71)	0.03	0.18 (0.02 to 0.34)	2.42 (0.19 to 4.65)	0.03	0.18 (0.02 to 0.35)	
12 months	39 (11)	41 (12)	2.28 (-0.20 to 4.76)	0.07	0.14 (-0.01 to 0.30)	2.29 (-0.15 to 4.73)	0.07	0.14 (-0.01 to 0.29)	
Recovery assessment scale									
Personal confidence and hope:									
Baseline	10 (3.4)	9.9 (3.7)	—	0.69	—	—	—	—	
Six months‡	12 (4.4)	13 (4.1)	0.54 (-0.22 to 1.29)	0.16	0.12 (-0.05 to 0.28)	0.54 (-0.23 to 1.31)	0.17	0.12 (-0.05 to 0.29)	
12 months‡	12 (3.9)	13 (4.2)	0.50 (-0.28 to 1.27)	0.21	0.13 (-0.07 to 0.32)	0.46 (-0.31 to 1.28)	0.24	0.12 (-0.08 to 0.31)	
Goal and success orientation:									
Baseline	5.3 (2)	5.4 (2)	—	0.78	—	—	—	—	
Six months‡	5.4 (2.2)	5.8 (2.2)	0.23 (-0.20 to 0.67)	0.28	0.14 (-0.12 to 0.39)	0.14 (-0.29 to 0.57)	0.51	0.09 (-0.18 to 0.35)	
12 months‡	5 (2.2)	5.4 (1.9)	0.16 (-0.22 to 0.53)	0.40	0.09 (-0.13 to 0.31)	0.11 (-0.26 to 0.49)	0.55	0.07 (-0.15 to 0.29)	
Willingness to ask others for help:									
Baseline	7.1 (2.8)	7 (2.9)	—	0.77	—	—	—	—	
Six months‡	8.4 (3.6)	8.8 (3.3)	0.47 (-0.16 to 1.09)	0.14	0.09 (-0.03 to 0.22)	0.53 (-0.10 to 1.15)	0.10	0.11 (-0.02 to 0.23)	
12 months‡	8.4 (3.1)	8.8 (3.1)	0.28 (-0.32 to 0.89)	0.36	0.07 (-0.09 to 0.23)	0.28 (-0.34 to 0.89)	0.37	0.07 (-0.09 to 0.24)	
Reliance on others:									
Baseline	5.9 (2)	6 (1.9)	—	0.79	—	—	—	—	
Six months‡	6.7 (1.8)	6.7 (1.8)	0.08 (-0.27 to 0.44)	0.64	0.05 (-0.15 to 0.25)	0.11 (-0.25 to 0.46)	0.56	0.06 (-0.14 to 0.26)	
12 months‡	6.5 (1.8)	6.5 (1.9)	0.01 (-0.32 to 0.34)	0.97	0.00 (-0.15 to 0.16)	0.01 (-0.32 to 0.33)	0.96	0.00 (-0.14 to 0.14)	
No domination by symptoms:									

(Continued)

Table 3 | Continued

Secondary outcomes	Control group (n=189)	Intervention group (n=187)	Minimally adjusted mixed model analysis (95% CI)*	P value	Effect size (minimally adjusted mixed model) (95% CI)	Adjusted mixed model analysis (95% CI)†	P value	Effect size (adjusted mixed model) (95% CI)	Patient health questionnaire anxiety and depression scale
Baseline	2.5 (1.8)	2.6 (1.8)	—	0.69	—	—	—	—	—
Six months‡	3.8 (2.5)	4.1 (2.5)	0.48 (-0.11 to 1.07)	0.11	0.19 (-0.05 to 0.42)	0.38 (-0.21 to 0.96)	0.20	0.15 (-0.08 to 0.39)	
12 months‡	4.1 (2.1)	4.5 (2.3)	0.51 (0.03 to 0.99)	0.04	0.16 (0.01 to 0.30)	0.46 (-0.02 to 0.93)	0.06	0.16 (-0.01 to 0.32)	
Patient assessment of chronic illness care-short form									
Baseline	2.6 (0.85)	2.4 (0.85)	—	0.03	—	—	—	—	—
Six months	2.4 (0.81)	2.8 (0.9)	0.34 (0.14 to 0.53)	<0.01	0.39 (0.17 to 0.61)	0.34 (0.15 to 0.53)	<0.01	0.38 (0.17 to 0.60)	
12 months	2.4 (0.85)	2.5 (0.94)	0.13 (-0.07 to 0.32)	0.19	0.13 (-0.07 to 0.32)	0.13 (-0.06 to 0.31)	0.18	0.12 (-0.05 to 0.29)	

Data are mean (SD), unless otherwise stated, effect size measured with Cohen's d.

*Mixed effect linear regression model, minimally adjusted for the respective baseline score and centre for intervention v control.

†Mixed effect linear regression models adjusted for the respective baseline score, age, gender, history of depression or anxiety, 12 item short form physical component score, chronic medical disease, and no. of days between baseline assessment and randomisation, for intervention v control.

‡The German version was used.

§Mixed effect linear regression model adjusted for the respective baseline score, age, gender, history of depression or anxiety, chronic medical disease, no. of days between baseline assessment and randomisation.

(53/155) of the participants in the intervention group at six months. Differences in the mean change from baseline in the PHQ-ADS scores between the subgroup of participants in the intervention group who received specialised mental health care in addition to the intervention and those who did not were small at six months (mean change from baseline -10.3 points (SD 10.03) v -9.2 (8.58), $t(153)=0.70$, $P=0.48$) and at 12 months (-12.9 (10.41) v -10.0 (9.47), $t(153)=1.80$, $P=0.07$).

Treatment in the treatment as usual group

In the control group, 86% (114/133) of the participants who were followed up six months had seen their GP a median of four times (range 1-35). The remaining 14% (19/133) of the participants had not seen their GP in that period at all. The proportion of participants in the control group who had received some form of specialised mental health care (psychiatric consultation, psychotherapy, psychosocial counselling service, or inpatient or day clinic treatment) was 47% (62/133) with 24% (32/133) having received at least five consultations, including psychotherapy or inpatient or day clinic treatment, or both. At six months, 27% (36/133) of participants in the control group had taken or were taking regular psychopharmacological medication.

Discussion

Principal findings

The aim of the PROVIDE-C trial was to investigate the effectiveness of a new integrated mental health video consultation model compared with treatment as usual for reducing symptom severity of depression and anxiety in patients receiving primary care. At a small effect size, the PROVIDE model was more effective in reducing not only severity of depressive and anxiety symptoms but also of psychological distress related to somatic symptoms, at both six months and 12

months after baseline assessment. Although the natural tendency for symptoms to improve over time without any specific treatment could account for some improvements in depression and anxiety severity scores, the effect cannot account for findings in favour of the PROVIDE model.

We did not detect any significant differences between the trial groups in the scores for some of the secondary outcomes. At six months, no significant differences were noted between the groups in the anxiety symptom severity score, the physical component score of health related quality of life, and the recovery subdomain scores. At 12 months, no significant differences were noted between the trial groups for the mental or physical component scores for health related quality of life, the recovery subdomains scores, or the patient centredness of chronic illness care score. Compared with treatment as usual, the PROVIDE model did not seem to offer any additional benefit for recovery. While the effect on patient centredness of chronic illness care seems to disappear over time, the effect on anxiety symptom severity may only occur after a longer period. These interpretations must be considered with caution because the subgroup analyses were not powered to detect significant differences between the trial groups.

Strengths and weaknesses

This study has several limitations. Selection bias is a pressing issue in practice based clinical research. Ensuring that the investigated population represents the targeted population is crucial, yet recruiting participants, particularly in primary care, is challenging.³³ We chose not to use random patient selection because of its impracticality in avoiding self-selection bias and the excessive prolongation of recruitment periods. Additionally, the use of a digital clinical trial recruitment support system was considered but ultimately rejected because these systems have not proven effective in mitigating selection bias in German primary care settings and faced substantial practical

implementation issues.^{34 35} Instead, we assumed that GPs would not be reluctant to offer trial participation to patients with high needs because even if a patient was randomised to the control group, the GP could have proceeded with usual care, including standard referrals to specialists. As such, we relied on GP consultations to leverage existing doctor and patient relationships.³⁶ Although we used broad but clearly defined eligibility criteria to enhance representativeness of the general population, including patients with conditions who are often excluded from psychotherapy trials (eg, medical comorbidity, history of psychopharmacological or electroconvulsive treatment, subacute schizophrenia, and history of substance misuse), this recruitment strategy may have introduced selection bias.³⁶⁻³⁸ To assess the representativeness of the PROVIDE-C sample, we first compared the group of eligible patients enrolled in the trial to the group of eligible patients who did not consent to enrolment. Most sociodemographic and medical variables (age, gender, education level, employment status, income management, chronic physical disease, knowledge of local mental health services, symptom severity (PHQ-ADS), and probability of having post-traumatic stress disorder) did not predict trial participation. However, patients who were single were less likely to participate ($P=0.02$). This may be explained by an absence of social support in single individuals, which is a significant factor in mental health care engagement, or by them experiencing higher levels of stigma related to mental health issues, leading to a preference for self-reliance over seeking professional help. On a broader scale, we then progressed with comparisons using external data. These comparisons showed that the trial participants had significantly lower physical health scores (SF-12) and higher psychological distress scores (SSD-12) than those in another German primary care sample, likely due to the inclusion of patients with depression and anxiety.³⁹ Enrolled patients' education levels corresponded to those of the German general population and their suicidal ideation frequencies were largely consistent with another sample from a primary care setting.⁴⁰ Our recruitment metrics—eligibility fraction (76%), enrolment fraction (80%), and recruitment fraction (61%)—were within the interquartile range for other primary care trials, indicating reasonable recruitment effectiveness.⁴¹ These findings highlight the inherent difficulties in meeting a fully representative and unbiased sample in practice based clinical research. Future studies should consider using more advanced digital recruitment systems and strategies to better mitigate selection bias and enhance representativeness. These steps could provide a more objective analysis and strengthen the validity of study findings.

The proportions of missing outcome data at six months differed between the intervention group and the control group. This suggests that loss to follow-up might be related to participants' health status; for example, more patients with more severe symptoms in the control group were less likely to return for follow-up. However, while missing outcome data are inevitable

in all randomised controlled trials, we made every effort to prevent missing data (eg, prioritising short questionnaires, prespecifying analyses, monitoring missing outcome data, and sending frequent reminders) when designing the trial. To check whether missing outcomes led to bias, we compared the covariates (ie, the prognostic baseline characteristics) between the group of participants with missing data for at least one PHQ-ADS item at six months and the group with complete PHQ-ADS data at that timepoint and did not find any significant differences. Hence, the prognostic baseline characteristics were unlikely to be responsible for the differences in the proportion of missing data between the trial groups. A large difference was noted in the long term follow-up at 12 months, where the difference in the intervention group was substantially greater than that in the control group, which we aimed to account for by again applying multiple imputation for the missing information.

We decided to use patient reported outcomes as direct measures of benefits that were easy to record, easy to interpret, and relevant to the patient experience as indicated by the endorsement of the patient representatives involved in the trial. The assessor of outcomes has been argued to be the study participant, even if a masked interviewer questions the participant.⁴² In this regard, based on earlier evidence, the fact that unintentional unmasking occurred in 20% of the participants may have led to an exaggerated pooled effect size.⁴³ However, a large meta-epidemiological study found no evidence for an average difference in the estimated treatment effect between trials with and without masked outcome assessors.⁴⁴

In the secondary analysis, we did not detect a significant difference between the proportion of participants in the intervention group who had an improvement at least as large as the minimal clinically important difference and the respective proportion of participants in the control group. This inconsistency is frequently observed in clinical trials and is related to the fact that, due to the dichotomisation of continuous data, responder analysis ultimately entails a reduction in statistical power and should generally be avoided as the primary analysis approach.⁴⁵⁻⁴⁷ Use of minimal clinically important differences as dichotomisation thresholds does not ensure clinically relevant responder effects.^{46 48} Nevertheless, reporting findings on minimal clinically important differences usually facilitates clinicians' and policy makers' understanding of trial results.

The difference in the change of 2.3 points was lower than that of the minimal clinically important difference in the PHQ-ADS score (3-5 point change). How might this result be explained? According to a recent umbrella review of meta-analyses of randomised controlled trials of psychotherapies and pharmacotherapies for the most common mental disorders in adults, conditions for which very extensive evidence is available “almost always had such modest effect sizes when only studies with low risk of bias were considered or efforts were

made to correct for bias".⁴⁹ The observation that neither medication nor psychotherapy work as well as previously assumed fits well with the general trend towards low and decreasing effect sizes in several fields.⁵⁰ In a broader reading, small effects are more likely to have been correctly estimated and are more likely to reflect the true state of nature.⁵¹ Application in the real world to patients, many of whom have a much worse prognosis than those in trial settings, is more challenging and implementation is less effective; as such, the effects of both medication and psychotherapy become even weaker.⁴⁹ Therefore, small effect sizes are the norm when considering real-world data stemming from samples with increased participant diversity and captured under the typical conditions of daily practice.^{52 53} Trials that recruit patients in primary care usually result in effect sizes lower than 0.3, that is, the effect of psychotherapy in primary care is relatively small compared with its effect in truly specialised care.⁵⁴ Specifically, a high quality meta-analysis showed that cognitive behaviour therapy was more effective than the control treatment for depression in primary care settings, with a small effective size.⁵⁵

The PROVIDE-C trial may have featured a quite strong control condition in that primary care interventions are typically already patient centred, which may lessen the likelihood of detecting an effect.⁵⁶ Moreover, 47% of the patients in the control group were referred to specialists, including psychotherapists, which is covered by the German statutory health insurance. This rate is more than twice as high as that of 21% reported in a large German epidemiological study.⁵⁷ Indeed, GPs with a particular interest in mental health, with some of them even having an additional qualification in mental health care, were more likely to participate in the PROVIDE-C trial and may have managed their patients more effectively than the average GPs, decreasing the effect size observed. Furthermore, the fact that the GPs were sensitised to the distress of their respective patients through inclusion in the trial and were therefore possibly more likely to initiate specialised care may have also contributed to the positive outcomes for the control group.

To balance the trade-off between engaging high caseloads and offering an optimal treatment dose for improvement, the PROVIDE-C intervention comprised five consultations, which corresponds to the lower dose range for routinely delivered psychological treatments. Further research is needed to explore the observation that differences were small in PHQ-ADS scores between participants who received specialised mental health care in addition to the intervention and those who did not. Research needs to delineate underlying factors contributing to this result, such as the quality and consistency of the specialised care provided and individual variations in response to treatment. Additionally, investigations into whether some subgroups might benefit more from specialised mental health care would be beneficial to optimise personalised treatment approaches. Indeed, some evidence suggests that polysymptomatic primary care

populations with more severe symptoms may require up to 13 sessions to attain reliable and clinically significant improvements.⁵⁸

The trial has substantial strengths. The sample predominantly included female participants and middle class individuals with free access to health insurance. To ensure the generalisability of these results, testing the effects in samples that are more diverse in terms of economic and ethnic backgrounds will be useful. However, we included a wide range of patients with respect to age, employment status, comorbid chronic physical disease status, and severity of depression or anxiety. In summary, trials in any setting are rarely fully representative in terms of both patient and disease related characteristics. However, the population investigated in the PROVIDE-C trial is large plausibly represents the target population of patients with depression or anxiety disorders who are receiving primary care. The PROVIDE-C trial was based on sound reimplemention and pilot studies that leveraged feedback from key informants to optimise the intervention and mirror routine care.¹⁷ Although the complexity of patients, health professionals, clinical settings, cultures, and health care systems limits the applicability of trial results, the PROVIDE-C trial was a fit-for-purpose trial as indicated by a pragmatic explanatory continuum indicator summary-2 (PRECIS-2) score of 36/45 points.⁵⁹ For both clinicians and patients, the trial was likely clinically relevant: recruitment was completed on the original timescale, which is rare in studies conducted in primary care settings.^{36 53} Furthermore, four in five eligible patients were enrolled in the trial, indicating a high preference for the intervention. Given that the trial was tailored to the needs of patients whom the trial was intended to help and that a broad spectrum of patients with depression and anxiety was included, we assume that the intervention can be reliably delivered to the entire population of primary care patients. This assumption is supported by the very low intraclass correlation coefficient in our trial, which indicates that variation in the outcomes among individuals in one practice was less than that between practices. That is, in our study individual primary care practices did not have a disproportionate effect on the findings of the trial.⁶⁰

Comparison with other studies

The trial findings suggest that mental health specialists virtually co-located in primary care can effectively deliver patient centred psychological interventions at the most frequent point of care entry, primary care. Moreover, the pragmatic and relatively brief intervention, which featured five sessions of treatment, improved medium term outcomes across a broad spectrum of patients with depression, anxiety and psychological distress related to somatic symptoms. Our results align with the literature on the efficacy of both telemental health interventions integrated into primary care settings and the collocation of mental health specialists in primary care practices.^{13 15 16 27 61} Specifically, our findings strongly support evidence that telemental health

interventions conveying practical training for specific skills may be particularly suited for treating depression and anxiety disorders.^{62 63} Patients have been shown to prefer video consultations at their local primary care practice over both video consultations into their home and travelling for in-person appointments.⁶⁴ The PROVIDE model, in addition to offering fast access to a defined intervention, accounted for patients' health care delivery preferences. Leveraging the familiar primary care environment for mental health care may facilitate treatment engagement for patients who have difficulties developing trust and opening up to new people.⁶⁵ Moreover, relying on primary care practices as a place for mental health treatment may mitigate physical and psychosocial limitations related to patients' homes (eg, a scarcity of privacy or technical equipment, distraction, coercion, or violence). In Germany, as in many countries around the world, social distancing started with the beginning of the covid-19 pandemic in March 2020, which was the month when the PROVIDE-C trial started. In many countries, the relaxation of telemedicine regulations and telemental health use rapidly increased, with almost complete virtualisation of specialised mental health care occurring by the end of the first year of the pandemic.⁶⁶ Hence, the pandemic may have facilitated the adoption of the PROVIDE model into routine care. Leveraging major facilitators for positive remote care experiences from the service users' perspective, the model engaged patients in receiving telemental health services through the existing relationships with their GPs and offered low threshold access to the necessary technology, particularly for elderly individuals.⁶⁸

What the study adds

Although the small effect size of this study was small, we argue that it is still meaningful to patients and service providers for several reasons.

More patients might be able to receive treatment via video based mode of delivery.

A small effect size (Cohen's *d*) of 0.20, as measured in this trial, is equivalent to an 11% difference in outcomes between therapy conditions (ie, affecting 110 people per 1000 people treated).⁶⁹ Considering that thousands of GPs treat an even larger number of patients each year, such a small effect does matter because it is linked to a collective and cumulative positive impact on population health need in terms of depression and anxiety disorders. Small effects that accumulate over time and at scale are highly likely to be consequential from a public health perspective.⁵¹ This applies to a condition such as depression with important and consequential outcomes (eg, long term disability or suicide). Important outcomes are primarily, if not entirely, influenced by numerous factors, each exerting only a very small effect.⁷¹

The PHQ-ADS score was stable at 12 months, which is in accordance with evidence that psychotherapy has greater long term effects.⁷²

The PROVIDE-C intervention entailed relatively low intensity and short term treatment, yet its effect is

similar to what can be expected from psychological therapy by on-site mental health workers to patients in primary care settings and psychotherapy.⁷³ Given that the model has not been linked to serious adverse effects and is of minimal inconvenience for patients and providers, the extent of the observed treatment effect warrants further study.

Conclusions and policy implications

Meaning of the study

The trial assessed the effectiveness of the PROVIDE model, an integrated mental health video consultation model in reducing depression and anxiety symptoms among primary care patients. The PROVIDE model showed a small but significant effect in reducing these symptoms, along with psychological distress related to somatic symptoms, at six and 12 months. While symptoms tend to improve naturally over time, the PROVIDE model offers additional benefits beyond treatment as usual.

How this study could promote better decisions

Despite the small effect size below the average minimal clinically important difference, the consistent improvement in symptoms with the PROVIDE model is meaningful given the high prevalence of depression and anxiety in community settings. Implementing the PROVIDE model more widely could enhance access to mental health care, especially in primary care settings where it can mitigate geographical and temporal barriers.

Unanswered questions and future research

Several questions remain unanswered and warrant further research. The small effect size, although significant, indicates a need to explore the factors affecting the variability in individual responses to treatment with the PROVIDE model. Moreover, the under-representation of younger individuals in our sample suggests a need to better understand how different age groups might benefit from telehealth interventions. Additionally, the dose effect association of the PROVIDE intervention remains unclear. While our intervention entailed a relatively low intensity and short treatment duration, further research is necessary to determine the optimal number of sessions and intensity required for different patient subgroups to result in clinically significant improvements. Understanding these dynamics will help tailor interventions more effectively and maximise their public health impacts. Finally, future studies should also investigate the long term sustainability and cost-effectiveness of the PROVIDE model across diverse demographic and socioeconomic groups. Addressing these gaps will be essential for refining telehealth interventions and optimising their benefits for a broader patient population.

AUTHOR AFFILIATIONS

¹Department of General Internal Medicine and Psychosomatics, Heidelberg University, Im Neuenheimer Feld 410, Heidelberg, Germany

²German Center for Mental Health (DZPG), Partner Site Mannheim-Heidelberg-Ulm, Germany

³Department of General Practice and Health Services Research, Heidelberg University, Im Neuenheimer Feld 130.3, Heidelberg, Germany

⁴Institute of Medical Biometry (IMB), Heidelberg University, Im Neuenheimer Feld 130.3, Heidelberg, Germany

⁵Institute for Health Services Research and Health Economics, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany

⁶Institute for Health Services Research and Health Economics, Centre for Health and Society, Medical Faculty and University Hospital Düsseldorf, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany

We thank Stephanie Estel, who assisted with generating databases for the trial; the research teams at each of our research sites; Regina Brinster and Dorothea Weber for their support in drafting the trial protocol and Christina Klose for operating the randomisation system. We also thank Wolfgang Herzog for his valuable contribution to the design of the study, and Philip Winkler for assisting with the analysis; Michael Milz and the team at arztkonsultation ak GmbH for the sound technical support, and Franziska Zumbaum-Fischer for supervising the mental health specialists; and our patient representatives for their support in developing the intervention and for their valuable input throughout the trial.

Contributors: MWH, JT, and MP drafted the manuscript, assisted by AW, MH, and MW, and all authors reviewed and approved it. MWH, MH, MW, JS, MV, AI, and H-CF were responsible for the design of the trial. MWH, JT, and MH were responsible for intervention content, training, and supervision. MWH, JT, and AW were responsible for trial conduct. MWH, JT, and AW were responsible for database design and management. MP, MF, and MWH were responsible for the analyses. All authors contributed to the interpretation of the data. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication. MWH, JT, MF, and MP directly accessed and verified the underlying data reported in this manuscript. All authors read and approved the final manuscript. MWH is guarantor. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding: This trial was funded by a grant from the German Federal Ministry of Education and Research (BMBF) (grant no. 01GY16129). All researchers can confirm their independence from funders, and all authors, external and internal, had full access to all the data (including statistical reports and tables) in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/disclosure-of-interest/and declare: funding from German Federal Ministry of Education and Research (BMBF) (grant no. 01GY16129); no financial relationships with any funding from organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have affected the submitted work.

Ethical approval: The trial protocol was approved by the Medical Faculty of the University of Heidelberg Ethics Committee (S-923/2019) on 7 January 2020. All participants provided written informed consent to participate in this trial.

Data sharing: The trial investigators own and have complete control of the research data, which can be accessed at any time. For statistical analysis, the data will be downloaded and safely stored on a computing system maintained by Heidelberg University. For individual participant data meta-analysis, beginning nine months following article publication, deidentified participant data that underlie the results reported in this article (text, tables, figures, and appendices) and a data dictionary will be available to investigators whose proposed use of the data has been approved by an independent review committee identified for this purpose. Consent for data sharing was not obtained but the presented data are anonymised, and risk of identification is low. The study protocol has been published and the statistical analysis plan is provided in the supplementary material S1 (pages 1-17).

Transparency: The lead author affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Dissemination to participants and related patient and public communities: The results of this trial will be presented at relevant national and international conferences. To reach health care policy and practice audiences (eg, government bodies and health insurance providers) concerning the scale-up of the model, we will present the findings at conferences run by policy makers and service providers. We will also publish a press release and social media content on the main findings and implications of the trial. To directly inform health policy makers and practitioners, we will report the findings in plain language formats and compile an executive summary drawing together key findings of all aspects of the intervention with a series of journal articles included as appendices. The results of the trial will be disseminated by letter to all participants by the end of October 2024.

Provenance and peer review: Not commissioned; externally peer reviewed.

- James SL, Abate D, Abate KH, et al. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018;392:1789-858. doi:10.1016/S0140-6736(18)32279-7
- Alonso J, Liu Z, Evans-Lacko S, et al. WHO World Mental Health Survey Collaborators. Treatment gap for anxiety disorders is global: Results of the World Mental Health Surveys in 21 countries. *Depress Anxiety* 2018;35:195-208. doi:10.1002/da.22711
- Thornicroft G, Chatterji S, Evans-Lacko S, et al. Undertreatment of people with major depressive disorder in 21 countries. *Br J Psychiatry* 2017;210:119-24. doi:10.1192/bjp.bp.116.188078
- World Health Organization. World mental health report: transforming mental health for all. Geneva, Switzerland: World Health Organization; 2022 June. <https://archive.hshsl.umaryland.edu/handle/10713/20295>.
- Angermeyer MC, van der Auwera S, Carta MG, Schomerus G. Public attitudes towards psychiatry and psychiatric treatment at the beginning of the 21st century: a systematic review and meta-analysis of population surveys. *World Psychiatry* 2017;16:50-61. doi:10.1002/wps.20383
- World Health Organization (WHO). Integrating mental health into primary care: a global perspective. <https://www.who.int/publications/i/item/9789241563680>. 2008;206.
- Wang PS, Aguilar-Gaxiola S, Alonso J, et al. Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the WHO world mental health surveys. *Lancet* 2007;370:841-50. doi:10.1016/S0140-6736(07)61414-7
- Gask L, Kendrick T, Peveler R, Chew-Graham CA, eds. *Primary Care Mental Health*. 2nd ed. Cambridge University Press, 2018. <https://www.cambridge.org/core/books/primary-care-mental-health/3A19C2D8465A31708FB216D87EE4DDE4>, doi:10.1017/9781911623038.
- Ring A, Dowrick CF, Humphris GM, Davies J, Salmon P. The somatising effect of clinical consultation: What patients and doctors say and do not say when patients present medically unexplained physical symptoms. *Soc Sci Med* 2005;61:1505-15.
- Hoffmann M, Hartmann M, Wensing M, Friederich HC, Haun MW. Potential for integrating mental health specialist video consultations in office-based routine primary care: cross-sectional qualitative study among family physicians. *J Med Internet Res* 2019;21:e13382. doi:10.2196/13382
- Crowley RA, Kirschner N, Health and Public Policy Committee of the American College of Physicians. The integration of care for mental health, substance abuse, and other behavioral health conditions into primary care: executive summary of an American College of Physicians position paper. *Ann Intern Med* 2015;163:298-9. doi:10.7326/M15-0510
- Bower P, Gilbody S. Managing common mental health disorders in primary care: conceptual models and evidence base. *BMJ* 2005;330:839-42. doi:10.1136/bmj.330.7495.839
- Woods JB, Greenfield G, Majeed A, Hayhoe B. Clinical effectiveness and cost effectiveness of individual mental health workers colocated within primary care practices: a systematic literature review. *BMJ Open* 2020;10:e042052. doi:10.1136/bmjopen-2020-042052
- Hilty DM, Rabinowitz T, McCarron RM, et al. An update on telepsychiatry and how it can leverage collaborative, stepped, and integrated services to primary care. *Psychosomatics* 2018;59:227-50. doi:10.1016/j.psych.2017.12.005
- Fortney JC, Pyne JM, Edlund MJ, et al. A randomized trial of telemedicine-based collaborative care for depression. *J Gen Intern Med* 2007;22:1086-93. doi:10.1007/s11606-007-0201-9
- Fortney JC, Pyne JM, Kimbrell TA, et al. Telemedicine-based collaborative care for posttraumatic stress disorder: a randomized clinical trial. *JAMA Psychiatry* 2015;72:58-67. doi:10.1001/jamapsychiatry.2014.1575

- 17 Tönnes J, Hartmann M, Wensing M, et al. Mental health specialist video consultations versus treatment-as-usual for patients with depression or anxiety disorders in primary care: randomized controlled feasibility trial. *JMIR Ment Health* 2021;8:e22569. doi:10.2196/22569
- 18 Haun MW, Tönnes J, Krisam R, et al. Mental health specialist video consultations versus treatment as usual in patients with depression or anxiety disorders in primary care: study protocol for an individually randomised superiority trial (the PROVIDE-C trial). *Trials* 2021;22:327. doi:10.1186/s13063-021-05289-3
- 19 Wilson S, Delaney BC, Roalfe A, et al. Randomised controlled trials in primary care: case study. *BMJ* 2000;321:24-7. doi:10.1136/bmj.321.7252.24
- 20 Zhang A, Franklin C, Jing S, et al. The effectiveness of four empirically supported psychotherapies for primary care depression and anxiety: a systematic review and meta-analysis. *J Affect Disord* 2019;245:1168-86. doi:10.1016/j.jad.2018.12.008
- 21 Fonagy P. Achieving evidence-based psychotherapy practice: a psychodynamic perspective on the general acceptance of treatment manuals. *Clin Psychol Sci Pract* 1999;6:442-4.
- 22 Diener MJ, Hilsenroth MJ, Weinberger J. Therapist affect focus and patient outcomes in psychodynamic psychotherapy: a meta-analysis. *Am J Psychiatry* 2007;164:936-41. doi:10.1176/ajp.2007.164.6.936
- 23 Bower P, Gilbody S, Richards D, Fletcher J, Sutton A. Collaborative care for depression in primary care. Making sense of a complex intervention: systematic review and meta-regression. *Br J Psychiatry* 2006;189:484-93. doi:10.1192/bjp.bp.106.023655
- 24 Carroll KM, Nuro KF. One size cannot fit all: a stage model for psychotherapy manual development. *Clin Psychol Sci Pract*. 2002;9:396-406 doi:10.1093/clipsy.9.4.396.
- 25 Tönnes J, Hartmann M, Wensing M, et al. Mental health specialist video consultations for patients with depression or anxiety disorders in primary care: protocol for a randomised controlled feasibility trial. *BMJ Open* 2019;9:e030003. doi:10.1136/bmjopen-2019-030003
- 26 Johansson R, Björklund M, Hornborg C, et al. Affect-focused psychodynamic psychotherapy for depression and anxiety through the Internet: a randomized controlled trial. *PeerJ* 2013;1:e102. doi:10.7717/peerj.102
- 27 Fortney JC, Pyne JM, Mouden SB, et al. Practice-based versus telemedicine-based collaborative care for depression in rural federally qualified health centers: a pragmatic randomized comparative effectiveness trial. *Am J Psychiatry* 2013;170:414-25. doi:10.1176/appi.ajp.2012.12050696
- 28 Yellowlees PM, Parish MB, Gonzalez AD, et al. Clinical outcomes of asynchronous versus synchronous telepsychiatry in primary care: randomized controlled trial. *J Med Internet Res* 2021;23:e24047. doi:10.2196/24047
- 29 Pflanz CP, Gallacher J, Bauermeister S. A psychometric evaluation of the 16-item PHQ-ADS concomitant anxiety and depression scale in the UK biobank using item response theory. *J Affect Disord* 2024;347:335-44. doi:10.1016/j.jad.2023.11.067
- 30 Kroenke K, Baye F, Lourens SG. Comparative validity and responsiveness of PHQ-ADS and other composite anxiety-depression measures. *J Affect Disord* 2019;246:437-43. doi:10.1016/j.jad.2018.12.098
- 31 Chondros P, Davidson S, Wolfe R, et al. Development of a prognostic model for predicting depression severity in adult primary patients with depressive symptoms using the diamond longitudinal study. *J Affect Disord* 2018;227:854-60. doi:10.1016/j.jad.2017.11.042
- 32 Butcher NJ, Monsour A, Mew EJ, et al. Guidelines for reporting outcomes in trial reports: The CONSORT-Outcomes 2022 Extension. *JAMA* 2022;328:2252-64. doi:10.1001/jama.2022.21022
- 33 Goodyear-Smith F, York D, Petousis-Harris H, et al. Recruitment of practices in primary care research: the long and the short of it. *Fam Pract* 2009;26:128-36. doi:10.1093/fampra/cmp015
- 34 Köpcke F, Prokosch HU. Employing computers for the recruitment into clinical trials: a comprehensive systematic review. *J Med Internet Res* 2014;16:e161. doi:10.2196/jmir.3446
- 35 Heinemann S, Thüring S, Wedeken S, et al. A clinical trial alert tool to recruit large patient samples and assess selection bias in general practice research. *BMC Med Res Methodol* 2011;11:16. doi:10.1186/1471-2288-11-16
- 36 Moffat KR, Shi W, Cannon P, Sullivan F. Factors associated with recruitment to randomised controlled trials in general practice: a systematic mixed studies review. *Trials* 2023;24:90. doi:10.1186/s13063-022-06865-x
- 37 van der Lem R, de Wever WW, van der Wee NJ, van Veen T, Cuijpers P, Zitman FG. The generalizability of psychotherapy efficacy trials in major depressive disorder: an analysis of the influence of patient selection in efficacy trials on symptom outcome in daily practice. *BMC Psychiatry* 2012;12:192. doi:10.1186/1471-244X-12-192
- 38 Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials* 2015;16:495. doi:10.1186/s13063-015-1023-4
- 39 Toussaint A, Riedl B, Kehler S, Schneider A, Löwe B, Linde K. Validity of the Somatic Symptom Disorder-B Criteria Scale (SSD-12) in primary care. *Fam Pract* 2018;35:342-7. doi:10.1093/fampra/cmz116
- 40 Horton M, Perry AE. Screening for depression in primary care: a Rasch analysis of the PHQ-9. *BJPsych Bull* 2016;40:237-43. doi:10.1192/pb.bp.114.050294
- 41 Jones R, Jones RO, McCowan C, Montgomery AA, Fahey T. The external validity of published randomized controlled trials in primary care. *BMC Fam Pract* 2009;10:5. doi:10.1186/1471-2296-10-5
- 42 Higgins JP, Savović J, Page MJ, Elbers RG, Sterne JA. Assessing risk of bias in a randomized trial. In: *Cochrane Handbook for Systematic Reviews of Interventions*. John Wiley & Sons, Ltd, 2019: 205-28, <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119536604.ch8>, doi:10.1002/9781119536604.ch8.
- 43 Hróbjartsson A, Thomsen ASS, Emanuelsson F, et al. Observer bias in randomized clinical trials with measurement scale outcomes: a systematic review of trials with both blinded and nonblinded assessors. *CMAJ* 2013;185:E201-11. doi:10.1503/cmaj.120744
- 44 Moustgaard H, Clayton GL, Jones HE, et al. Impact of blinding on estimated treatment effects in randomised clinical trials: meta-epidemiological study. *BMJ* 2020;368:l6802. doi:10.1136/bmj.l6802
- 45 Lane M, Miao T, Turgeon RD. Clinician's approach to advanced statistical methods: win ratios, restricted mean survival time, responder analyses, and standardized mean differences. *J Gen Intern Med* 2024;39:1196-203. doi:10.1007/s11606-023-08582-w
- 46 Collister D, Bangdiwala S, Walsh M, et al. Patient reported outcome measures in clinical trials should be initially analyzed as continuous outcomes for statistical significance and responder analyses should be reserved as secondary analyses. *J Clin Epidemiol* 2021;134:95-102. doi:10.1016/j.jclinepi.2021.01.026
- 47 Snapinn SM, Jiang Q. Responder analyses and the assessment of a clinically relevant treatment effect. *Trials* 2007;8:31. doi:10.1186/1745-6215-8-31
- 48 Abugor R, Clark J, Higginbotham L, et al. Should responder analyses be conducted on continuous outcomes? *Pharm Stat* 2023;22:312-27. doi:10.1002/pst.2273
- 49 Leichsenring F, Steinert C, Rabung S, Ioannidis JPA. The efficacy of psychotherapies and pharmacotherapies for mental disorders in adults: an umbrella review and meta-analytic evaluation of recent meta-analyses. *World Psychiatry* 2022;21:133-45. doi:10.1002/wps.20941
- 50 Götz FM, Gosling SD, Rentfrow PJ. Effect sizes and what to make of them. *Nat Hum Behav* 2024;8:798-800. doi:10.1038/s41562-024-01858-z
- 51 Funder DC, Ozer DJ. Evaluating effect size in psychological research: sense and nonsense. *Adv Methods Pract Psychol Sci* 2019;2:156-68 doi:10.1177/2515245919847202.
- 52 Ormel J, Hollon SD, Kessler RC, Cuijpers P, Monroe SM. More treatment but no less depression: The treatment-prevalence paradox. *Clin Psychol Rev* 2022;91:102111. doi:10.1016/j.cpr.2021.102111
- 53 Heinmüller S, Schneider A, Linde K, DFG Network Clinical Trials in General Practice. Quantity, topics, methods and findings of randomised controlled trials published by German university departments of general practice - systematic review. *Trials* 2016;17:211. doi:10.1186/s13063-016-1328-y
- 54 Cuijpers P, van Straten A, van Schaik A, Andersson G. Psychological treatment of depression in primary care: a meta-analysis. *Br J Gen Pract* 2009;59:e51-60. doi:10.3399/bjgp09X395139
- 55 Santoft F, Axelsson E, Öst LG, Hedman-Lagerlöf M, Fust J, Hedman-Lagerlöf E. Cognitive behaviour therapy for depression in primary care: systematic review and meta-analysis. *Psychol Med* 2019;49:1266-74. doi:10.1017/S0033291718004208
- 56 Fortin M, Stewart M, Almirall J, Beaupré P. Challenges in multimorbidity research: lessons learned from the most recent randomized controlled trials in primary care. *Front Med (Lausanne)* 2022;9:815783. <https://www.frontiersin.org/articles/10.3389/fmed.2022.815783>. doi:10.3389/fmed.2022.815783
- 57 Trautmann S, Beesdo-Baum K. The treatment of depression in primary care. *Dtsch Arztebl Int* 2017;114:721-8. doi:10.3238/arztebl.2017.0721
- 58 Robinson L, Delgadoillo J, Kellett S. The dose-response effect in routinely delivered psychological therapies: A systematic review. *Psychother Res* 2020;30:79-96. doi:10.1080/10503307.2019.1566676
- 59 Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ* 2015;350:h2147-h2147.

- 60 Adams G, Gulliford MC, Ukoumunne OC, Eldridge S, Chinn S, Campbell MJ. Patterns of intra-cluster correlation from primary care research to inform study design and analysis. *J Clin Epidemiol* 2004;57:785-94. doi:10.1016/j.jclinepi.2003.12.013
- 61 Yeung A, Martinson MA, Baer L, et al. The effectiveness of telepsychiatry-based culturally sensitive collaborative treatment for depressed Chinese American immigrants: a randomized controlled trial. *J Clin Psychiatry* 2016;77:e996-1002. doi:10.4088/JCP.15m09952
- 62 Hagi K, Kurokawa S, Takamiya A, et al. Telepsychiatry versus face-to-face treatment: systematic review and meta-analysis of randomised controlled trials. *Br J Psychiatry* 2023;223:407-14. doi:10.1192/bjp.2023.86
- 63 Fernandez E, Woldgabreal Y, Day A, Pham T, Gleich B, Aboujaoude E. Live psychotherapy by video versus in-person: A meta-analysis of efficacy and its relationship to types and targets of treatment. *Clin Psychol Psychother* 2021;28:1535-49. doi:10.1002/cpp.2594
- 64 Snoswell CL, Smith AC, Page M, Caffery LJ. Patient preferences for specialist outpatient video consultations: A discrete choice experiment. *J Telemed Telecare* 2021;1357633X211022898.
- 65 Hall JD, Danna MN, Hoelt TJ, et al. Patient and clinician perspectives on two telemedicine approaches for treating patients with mental health disorders in underserved areas. *J Am Board Fam Med* 2022;35:465-74. doi:10.3122/jabfm.2022.03.210377
- 66 Zangani C, Ostinelli EG, Smith KA, et al. Impact of the covid-19 pandemic on the global delivery of mental health services and telemental health: systematic review. *JMIR Ment Health* 2022;9:e38600. doi:10.2196/38600
- 67 Kinoshita S, Cortright K, Crawford A, et al. Changes in telepsychiatry regulations during the covid-19 pandemic: 17 countries and regions' approaches to an evolving healthcare landscape. *Psychol Med* 2022;52:2606-13. doi:10.1017/S0033291720004584
- 68 Vera San Juan N, Shah P, Schlieff M, et al. Service user experiences and views regarding telemental health during the covid-19 pandemic: A co-produced framework analysis. *PLoS One* 2021;16:e0257270. doi:10.1371/journal.pone.0257270
- 69 Barkham M. Smaller effects matter in the psychological therapies: 25 years on from Wampold et al. *Psychother Res* 2022;530-532. <https://www.tandfonline.com/doi/abs/10.1080/10503307.2022.2141589>.
- 70 Berk M, Marx W, Nierenberg AA. The aggregation of marginal gains: a pragmatic philosophy of clinical care in psychiatry. *World Psychiatry* 2024;23:291-2. doi:10.1002/wps.21174
- 71 Götz FM, Gosling SD, Rentfrow PJ. Small effects: the indispensable foundation for a cumulative psychological science. *Perspect Psychol Sci* 2022;17:205-15. doi:10.1177/1745691620984483
- 72 Furukawa TA, Shinohara K, Sahker E, et al. Initial treatment choices to achieve sustained response in major depression: a systematic review and network meta-analysis. *World Psychiatry* 2021;20:387-96. doi:10.1002/wps.20906
- 73 Harkness EF, Bower PJ. On-site mental health workers delivering psychological therapy and psychosocial interventions to patients in primary care: effects on the professional practice of primary care providers. *Cochrane Database Syst Rev* 2009;CD000532. <https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD000532.pub2/abstract>.

Web appendix: Extra material supplied by authors