



Low-effort internet intervention to reduce students' stress delivered with Meta's Messenger chatbot (Stressbot): A randomized controlled trial

Jan Maciejewski^a, Ewelina Smoktunowicz^{b,*}

^a Department of Psychology, SWPS University, Warsaw, Poland

^b StresLab Research Centre, Institute of Psychology, SWPS University, Warsaw, Poland

ARTICLE INFO

Keywords:

Facebook
Messenger
Chatbot
Low-effort
Self-efficacy
Stress

ABSTRACT

In order to be more accessible and overcome the challenges of low adherence and high dropout, self-guided internet interventions need to seek new delivery formats. In this study, we tested whether a widely-adopted social media app – Meta's (Facebook) Messenger – would be a suitable conveyor of such an internet intervention. Specifically, we verified the efficacy of Stressbot: a Messenger chatbot-delivered intervention focused on enhancing coping self-efficacy to reduce stress and improve quality of life in university students. Participants ($N = 372$) were randomly assigned to two conditions: (1) an experimental group with access to the Stressbot intervention, and (2) a waitlist control group. Three outcomes, namely coping self-efficacy, stress, and quality of life, were assessed at three time points: a baseline, post-test, and one-month follow-up. Linear Mixed Effects Models were used to analyze the data. At post-test, we found improvements in the Stressbot condition compared to the control condition for stress ($d = -0.33$) and coping self-efficacy ($d = 0.50$), but not for quality of life. A sensitivity analysis revealed that the positive short-term intervention effects were robust. At the follow-up, there were no differences between groups, indicating that the intervention was effective only in the short term. In sum, the results suggest that the Messenger app is a viable means to deliver a self-guided internet intervention. However, modifications such as a more engaging design or boosters are required for the effects to persist.

1. Introduction

One of the most common distinctions between multiple formats of internet-delivered psychological interventions (Smoktunowicz et al., 2020) is whether they are guided or not. The guidance usually relates to the content of the intervention being provided by a trained professional, whereas self-guided programs are delivered with no or minimal human input (Karyotaki et al., 2017). While therapist-guided interventions produce effects similar to in-person therapy (Hedman-Lagerlöf et al., 2023), self-guided programs tend to lead to smaller effects (Karyotaki et al., 2021) and suffer from low uptake and adherence, and high attrition rates (e.g., Linardon and Fuller-Tyszkiewicz, 2020). Yet, they have several advantages. First, both app- and web-based self-guided interventions have been found to be effective in treating multiple mental disorders and enhancing psychological health, including reducing stress (e.g., Karyotaki et al., 2017; Linardon et al., 2019). Second, they are easier and cheaper to maintain, meaning they can be offered on a large scale, which is particularly important when access to more structured help is limited (Bockting et al., 2016). Finally, they provide a higher

level of anonymity and even more flexibility than guided interventions (Bücker et al., 2019). These benefits encourage attempts to identify new means of delivering self-guided interventions and overcoming their challenges. Mohr et al. (2017) propose that in order to achieve a true paradigm shift, internet interventions should aim beyond replacing psychotherapy and introduce new ways of providing psychological help. Moreover, rather than inviting users to interventions, researchers might want to meet participants in spaces where they already spend their time. In response, in this study, we tested the efficacy of a self-guided low-effort internet intervention delivered via an app that is already used by more than a billion users (Kemp, 2021), that is, Meta's Messenger.

Using Meta's Messenger as a delivery tool for a psychological program offers several benefits. First, it addresses the entry barriers usually associated with internet interventions. Interventions are often delivered via apps and web platforms that require participants to create a new account. Moreover, because these apps usually have no standard design, they come with a learning curve: Users need to learn how to navigate the program before they can start to benefit from its content. Employing a well-known messaging app to deliver an intervention responds to these

* Corresponding author at: SWPS University, Chodakowska 19/31, 03-815 Warsaw, Poland.

E-mail address: esmoktunowicz@swps.edu.pl (E. Smoktunowicz).

crucial challenges (Dederichs et al., 2021; Smith et al., 2023), as it is likely that the end users are already familiar with the app and even have it installed on their phones. Second, because an intervention delivered through Meta's Messenger means employing a chatbot design, it introduces a better user experience. Specifically, a chatbot interface could decrease the effort needed to participate in the intervention and enhance the rewarding aspect of completing exercises, as every interaction with a chatbot is met with an immediate response (Baumel and Muench, 2021). Moreover, we believe that a combination of a chatbot and a popular social media app responds to participants' needs for interventions that are user-friendly and easily integrated into their lives (Borghouts et al., 2021).

However, chatbots range from complex AI-based implementations to simple, rule-based bots (Adamopoulou and Moussiades, 2020). The former are still in early development, and although examples of their successful adaptation in psychological programs exist (e.g., Liu et al., 2022), their preliminary use presents critical risks, such as misunderstanding participants' intentions (Boucher et al., 2021). Moreover, some participants have reported concerns regarding the usability of chatbots (Boucher et al., 2021), which we suspect could result from disappointment with the current capabilities of human-mimicking AI applications in psychological programs. Simple, rule-based chatbots on the other hand, albeit less technologically sophisticated, are less likely to under-deliver when it comes to users' high expectations, especially if the chatbot's mechanism is transparently disclosed. With this assumption, we employed a rule-based chatbot as a user-friendly tool to deliver an internet intervention in a convenient conversational form.

Our overarching goal in this study was to test whether a self-guided internet intervention delivered via a chatbot in a messaging app—in this case, Meta's Messenger—would be effective in improving psychological outcomes. Specifically, we focused on an intervention that aimed to reduce stress and improve quality of life through the enhancement of coping self-efficacy (Chesney et al., 2006). Previous studies have found programs focused on improving self-efficacy to be effective in various contexts, including improving medical workers' well-being (Smoktunowicz et al., 2021) and reducing stress in professionals exposed to indirect trauma (Cieslak et al., 2016). Yet, these programs were web-based, required effort to learn to use and were time-consuming. These factors limited their uptake which manifested in, among others, a high dropout rate. In this study, we tested whether enhancing self-efficacy and improving well-being would also be achieved with a new low-effort delivery method, namely Meta's Messenger chatbot.

A group particularly familiar with the Messenger app is university students: In 2018, >80 % of people aged 18 to 24 in the US used it (Dixon, 2022). Testing a Messenger-based intervention in this demographic makes it possible to meet the interventions' recipients in the space they already use, and thus remove entry barriers while simultaneously increasing the availability of psychological help. It also allows for testing the intervention in a setting close to real life (Mohr et al., 2017). Moreover, a low-effort intervention seems to fit students' needs as they already face various demands ranging from academic to social and psychological: e.g., financial management, interpersonal relationships, or time management (Logan and Burns, 2021). Hence, a traditional in-person or internet-based intervention could itself become a demand. Yet, students do suffer from restricted access to psychological help. This has been illustrated both in the international context (Auerbach et al., 2018) and specifically in Poland (NZS, 2021), where the current study was carried out. Therefore, a low-effort format could be a way to bring psychological help to those who would not access it otherwise.

In sum, the objective of this study was to determine whether a psychological self-guided intervention could be effective when delivered with a low-effort approach through a chatbot on Meta's Messenger. Specifically, we expected that completing exercises designed to enhance coping self-efficacy would result in students' higher coping self-efficacy (Hypothesis 1), lower stress (Hypothesis 2), and higher quality of life (Hypothesis 3) when compared to the waitlist control both at post-test

and at a one-month follow-up.

2. Methods

2.1. Study design

The study was a parallel randomized controlled trial (RCT) with two conditions compared on three time points: baseline (T1), post-intervention (T2), and a one-month follow-up (T3). The two conditions were: (1) an experimental condition with access to Stressbot – an intervention spanning seven days with the aim of enhancing coping self-efficacy delivered through a chatbot on Meta's Messenger, and (2) a waitlist control condition.

The study was approved by the Ethical Review Board at SWPS University (opinion 35/2022 issued on May 13, 2022) and preregistered on [ClinicalTrials.gov](https://www.clinicaltrials.gov) (NCT05500209).

2.2. Power analysis

To determine the required sample size, we conducted an a priori power calculation. Meta-analyses on internet interventions for stress reported effect sizes ranging from $d = 0.19$ to $d = 0.78$ (Amanvermez et al., 2022; Svärđman et al., 2022). A study comparing the effects of an experimental condition enhancing specific self-efficacy to an active control group reported an effect size of 0.49 (Cieslak et al., 2016). Since our intervention relied mainly on increasing self-efficacy but used a passive comparator, the effect size was expected to be similar or larger. A minimum effect size of $d = 0.50$ was therefore considered. Considering an alpha error level of 0.05, two conditions, and three measurement points, 240 participants were deemed to be necessary to achieve a power of 0.90. However, we recruited more participants as we expected a dropout rate of about 25 % (Linardon and Fuller-Tyszkiewicz, 2020). A post-hoc power analysis revealed that the final sample of 372 participants allowed us to preserve enough power to detect an effect size of $d = 0.50$ with up to 45 % attrition, or an effect as low as $d = 0.28$ when using multiple imputation. All power calculations were conducted using the *powerlmm* R package (Magnusson, 2019).

2.3. Procedure

Fig. 1 illustrates the study flow. The recruitment process was carried out between August and October 2022 via ads on Facebook. Specifically, the advertisements were targeted at Facebook groups for Polish university students. Participants were first directed to an online form and asked to accept the participation terms and sign the informed consent. Those who agreed to participate were screened for inclusion criteria. That is, they had to: (1) be at least 18 years old, (2) be a university student, and (3) be a Messenger user. Eight hundred and eleven interested individuals completed the screening. Seven hundred and ninety-four were eligible for inclusion. Subsequently, 563 of them completed the registration process by consenting to be redirected to the Messenger chatbot (called Stressbot in this study) to complete the registration process. Afterwards, all participants were asked to fill out the baseline (T1) assessment. Those who failed to do so despite two notifications ($n = 191$) were excluded as they did not fulfill the inclusion criteria. The final sample size of $N = 372$ was then randomly assigned to one of the two conditions. Participants in the experimental condition were given access to the Stressbot intervention that lasted for 7 days. Those in the control condition were asked to wait for one week. Finally, all participants were asked to fill out the posttest measurement (T2), and one month later, the follow-up assessment (T3). Participants in the control condition were given access to the Stressbot intervention immediately after the T3 measurement.

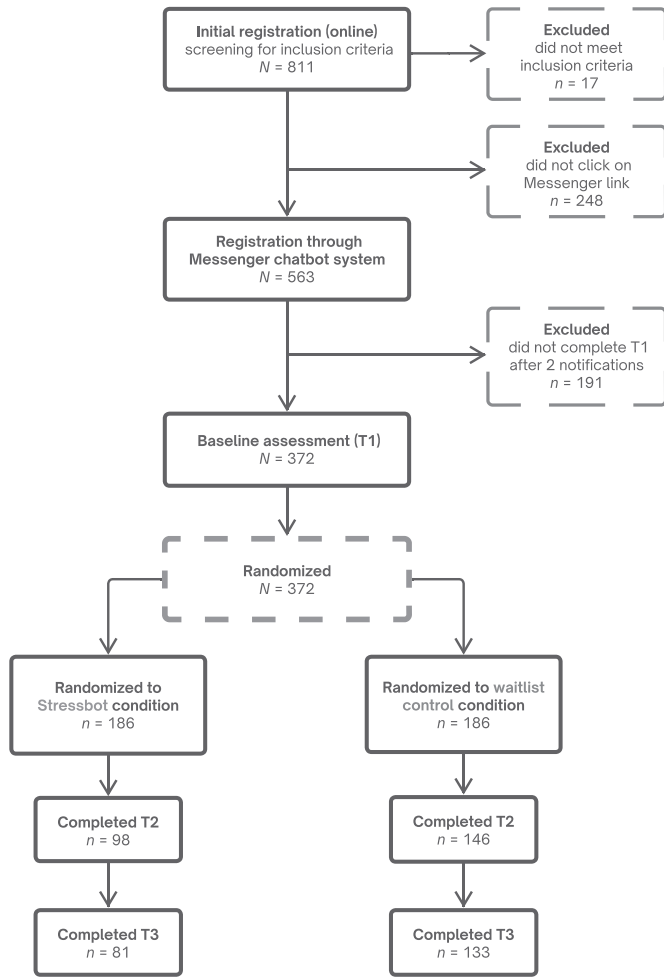


Fig. 1. Study flow.

2.4. Randomization procedure

All participants were recruited before the study started. Block randomization with a 1:1 ratio and a block size of 2 was used. The trial was not blinded. Participants were aware of the possibility of being allocated to one of two conditions and the general purpose of the Stressbot intervention before enrollment.

2.5. Stressbot intervention

Stressbot (illustrated in Fig. 2) is a self-guided internet intervention aiming to reinforce coping self-efficacy to reduce stress and improve quality of life in university students. Coping self-efficacy, defined as a “person’s confidence in his or her ability to cope effectively” (Chesney et al., 2006, p. 421), is an example of personal resources that, in line with the Conservation of Resources theory, is expected to improve stress and quality of life (Hobfoll, 1989). Coping self-efficacy-enhancing exercises are based on The Social Cognitive theory, in particular on increasing agency through mastery experiences (Bandura, 1997). The contents of each intervention day are described in Table 1.

The seven-day program consisted of six daily exercises and a break of one day. Participants were informed that the exercises were designed to take a maximum of thirty minutes every day to complete. The exercises were delivered each morning through a notification from Meta’s Messenger app. Each exercise consisted of a flow of messages designed specifically for each day. Stressbot asked its users questions to which they could reply via text, voice, or by tapping buttons. Stressbot’s answers were based on predetermined rules, and this fact was disclosed to

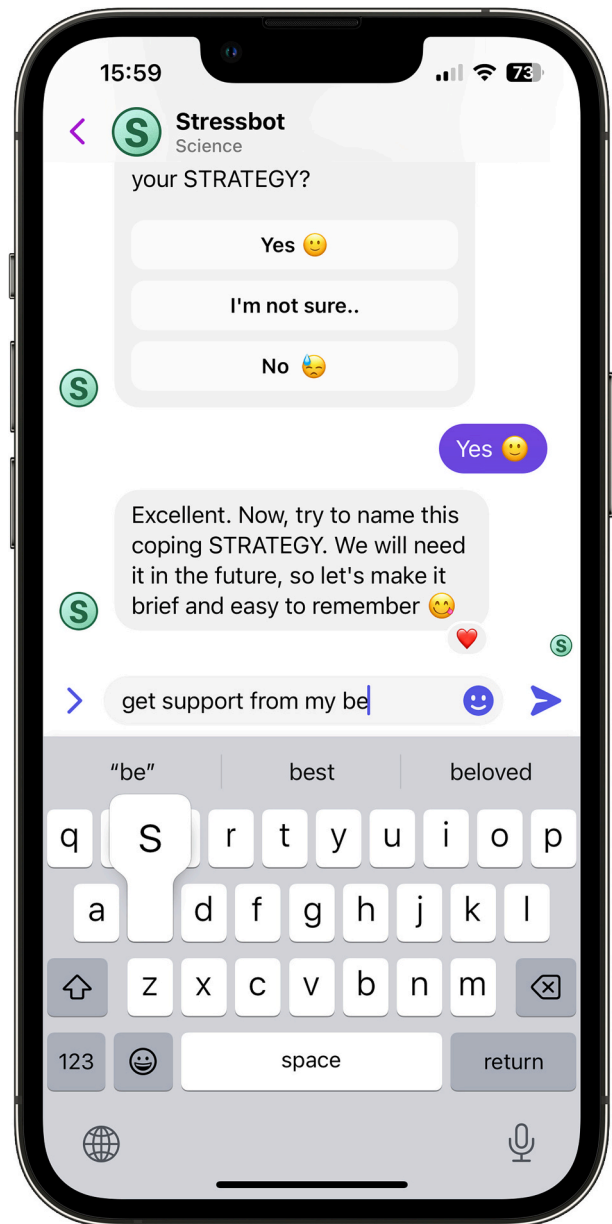


Fig. 2. A screenshot from the Stressbot intervention.

the participants, which is in line with recent recommendations for using chatbots in psychological programs (Kretschmar et al., 2019).

2.6. Measures

Three outcomes were assessed: coping self-efficacy, stress, and quality of life.

2.6.1. Coping self-efficacy

The Coping Self-Efficacy Scale (CSES; Chesney et al., 2006) consists of 26 items with responses ranging from 0 (cannot do at all) to 10 (certain can do). A higher total score represents higher coping self-efficacy. Cronbach’s alpha was 0.92 at T1, 0.95 at T2, and 0.95 at T3.

2.6.2. Stress

The Perceived Stress Scale (PSS-4; Cohen et al., 1983) consists of 4 items describing the frequency of stress symptoms with a response scale from 0 (never) to 4 (very often). A higher total score represents higher perceived stress. Cronbach’s alpha was 0.74 at T1, 0.80 at T2, and 0.81 at

Table 1
Contents of daily exercises in Stressbot intervention.

Day	Exercise	Description
1	Mastery experience	Recollecting past success with coping in a controllable situation
2	Mastery experience	Recollecting past success with coping in an uncontrollable situation
3	New situations	Anticipating future stressful situations in which recalled strategies can be helpful
4	Barriers	Anticipating barriers in using recalled strategies in new situations
5	Overcoming barriers	If-then approach: creating solutions to overcome barriers in new situations
6	Break	Summary of completed exercises
7	Vicarious experience	Recollecting others' successful coping experiences

T3.

2.6.3. Quality of life

The questionnaire used to assess Quality of life was the Brunnsvikken Brief Quality of Life Scale (BBQ; Lindner et al., 2016). The scale contains 12 items regarding the subjective quality of life with a response range from 0 (strongly disagree) to 4 (strongly agree). A higher total score represents higher quality of life. Cronbach's alpha was 0.75 at T1, 0.83 at T2, and 0.79 at T3.

2.7. Statistical analysis

First, we conducted a randomization check using chi-square and *t*-tests. Subsequently, we screened for missing data patterns with the Little's test and conducted a dropout analysis. To verify hypotheses regarding differences in outcome measures between the two conditions at T2 and T3, we built Linear Mixed Effects Models (LMEM) using *lme4* R package separately for each outcome with the intention to treat (ITT) principle. Models were fitted with the bottom-up principle and consisted of an interaction of time (measurement point) and group (Stressbot or control) as fixed effects and a random intercept. Using different analysis approaches (e.g., multiple imputation or complete case analysis) can lead to various biases when data are missing (Magnusson, 2019; Sullivan et al., 2018). To overcome this challenge and to verify the robustness of the effects, we conducted two parallel analyses: (1) Multiple imputation (MI) – pooled analyses performed on 30 datasets multiply imputed with *mice* and *mitml* R packages and (2) Completers (C) – analysis on data only from participants who completed all three measurements. The effect sizes were calculated as $M_1 - M_2 / SD_{pooled}$. Finally, to further test the robustness of the short-term effect (i.e., at T2) in the context of higher attrition in the intervention group, we conducted a sensitivity analysis comparing scenarios under which missing data represented worsening outcomes (Goldberg et al., 2021). All analyses were conducted with R version 4.1.2.

3. Results

3.1. Sample

The sample ($N = 372$) consisted of 26.8 % ($n = 96$) social sciences students, 18.4 % ($n = 66$) medical major students, 16.8 % ($n = 60$) economical major students, 14.5 % ($n = 52$) humanities students, 7 % ($n = 25$) STEM students, 6.1 % ($n = 22$) technical major students, 6.1 % ($n = 22$) biological major students, and 4.2 % ($n = 15$) art students. Participants' age ranged between 18 and 47 ($M = 20.98$, $SD = 3.05$). The sample consisted of 85 % ($n = 304$) women, 12.8 % ($n = 46$) men, and 2.2 % ($n = 8$) other genders.

3.2. Preliminary results

3.2.1. Descriptive statistics

Descriptive statistics are presented in Table 2.

3.2.2. Randomization check and dropout analysis

A randomization check confirmed no differences between

participants in the intervention and control conditions for any of the outcomes or demographic measures. Study dropout, defined as the loss between T1 and T2 measurements, was 34.4 %. Out of 372 randomized participants, 244 (146 in control and 98 in experimental condition) completed the T2 assessment. To test for discrepancies between those who did and did not complete T2, we first conducted the Little's test (Little and Rubin, 2002). Although the data turned out to be missing completely at random (MCAR) ($\chi^2_{73} = 88.60$, $p = .103$), we conducted a dropout analysis due to the risk of differential attrition (i.e., significantly higher dropout in the intervention condition (Goldberg et al., 2021)). The analysis revealed no differences between completers and non-completers. However, significant differential attrition was confirmed: Non-completers were more likely to have been randomly assigned to the Stressbot condition than the control condition ($\chi^2_1 = 27.44$, $p < .001$).

3.2.3. Intervention adherence

We calculated the adherence rate as the ratio of completed intervention days to the total number of days. The completion of each day was assessed automatically via the chatbot system. The average daily adherence rate was 47 %.

3.3. Hypotheses testing

We report our findings in the form of two parallel analyses in order to test the robustness of the effects (see 2.7. Statistical analysis section). The results of Linear Mixed Effects Models are presented in Table 3.

3.3.1. Coping self-efficacy

No effect of time or group was found for coping self-efficacy either in the MI, or the C model (Table 3). However, the interaction between group and time at T2 was significant for both models; MI: ($B = 0.55$, $SE = 0.15$, 95 % CI [0.25, 0.84]) and C: ($B = 0.49$, $SE = 0.17$, 95 % CI [0.17, 0.82]), indicating higher coping self-efficacy in the Stressbot condition immediately after the intervention. The effect size was moderate ($d = 0.50$ CI [0.29, 0.71]). At T3, the interaction was not significant in the MI model: ($B = 0.31$, $SE = 0.17$, 95 % CI [-0.03, 0.65]), but was significant in the C model: ($B = 0.40$, $SE = 0.17$, 95 % CI [0.08, 0.73]), indicating that the effect was not robust. Therefore, the first hypothesis was partially confirmed.

3.3.2. Stress

Similar to coping self-efficacy, we did not find an effect of time or group on stress in any of the models (Table 3). The interaction between condition and time was significant for both parallel analyses at T2; MI: ($B = -0.36$, $SE = 0.11$, 95 % CI [-0.57, -0.16]), and C: ($B = -0.32$, $SE = 0.12$, 95 % CI [-0.56, -0.09]), demonstrating lower stress in the Stressbot group after the intervention. The effect size was small ($d = -0.33$ CI [-0.13, -0.53]). However, at T3 we found no interaction of group and time in any of the models for stress (Table 3). The second hypothesis was thus only partially confirmed.

3.3.3. Quality of life

On quality of life, we found no effect of group or time at T2 (Table 3). At T3, we found an effect of time in both models; MI: ($B = 0.12$, $SE = 0.04$, 95 % CI [0.04, 0.19]), and C: ($B = 0.08$, $SE = 0.04$, 95 % CI [0.01,

Table 2
Means, standard deviations, and correlations for study variables.

Measure	<i>M</i>	<i>SD</i>	Range	1	2	3	4	5	6	7	8	9	10
1 Coping self-efficacy T1	4.24	1.60	0–8.58	–									
2 Coping self-efficacy T2	4.39	1.79	0–9.27	0.78***	–								
3 Coping self-efficacy T3	4.61	1.77	0–9.62	0.70***	0.81***	–							
4 Stress T1	2.20	0.84	0.25–4	–0.58***	–0.47***	–0.43***	–						
5 Stress T2	2.09	0.85	0.25–4	–0.46***	–0.59***	–0.55***	0.50***	–					
6 Stress T3	2.12	0.89	0.25–4	–0.37***	–0.42***	–0.60***	0.46***	0.59***	–				
7 Quality of life T1	2.78	0.55	0.75–4	0.61***	0.56***	0.53***	–0.47***	–0.35***	–0.28***	–			
8 Quality of life T2	2.77	0.62	0.50–4	0.50***	0.62***	0.52***	–0.40***	–0.46***	–0.32***	0.74***	–		
9 Quality of life T3	2.86	0.57	0.50–4	0.52***	0.55***	0.67***	–0.37***	–0.43***	–0.46***	0.70***	0.77***	–	
10 Age	20.98	3.05	18–47	0.05	0.17*	0.15*	–0.07	–0.09	–0.03	0.06	0.08	0.12	–

Note. T1 – baseline, T2 – post-intervention, T3 – one-month follow-up; *N* = 372.

* *p* < .05.
*** *p* < .001.

Table 3
Results of linear mixed effects models.

	Multiple imputation (MI)			Completers (C)		
	Fixed effects estimate (<i>SE</i>)	95 % CI	Random effect estimate	Fixed effects estimate (<i>SE</i>)	95 % CI	Random effect estimate
Coping self-efficacy						
Intercept	4.16 (0.13)	[3.91, 4.40]	2.28	4.20 (0.14)	[3.93, 4.48]	2.00
Time (T2)	0.01 (0.11)	[–0.21, 0.23]		–0.05 (0.10)	[–0.25, 0.15]	
Time (T3)	0.14 (0.10)	[–0.05, 0.33]		0.16 (0.10)	[–0.04, 0.36]	
Stressbot group	0.18 (0.18)	[–0.17, 0.53]		0.35 (0.23)	[–0.11, 0.80]	
T2 * Stressbot	0.55 (0.15)	[0.25, 0.84]		0.49 (0.17)	[0.17, 0.82]	
T3 * Stressbot	0.31 (0.17)	[–0.03, 0.65]		0.40 (0.17)	[0.08, 0.73]	
Stress						
Intercept	2.15 (0.06)	[2.02, 2.27]	0.39	2.13 (0.07)	[1.99, 2.28]	0.39
Time (T2)	0.03 (0.07)	[–0.11, 0.18]		0.02 (0.07)	[–0.12, 0.17]	
Time (T3)	0.06 (0.07)	[–0.07, 0.19]		0.01 (0.07)	[–0.13, 0.15]	
Stressbot group	0.12 (0.09)	[–0.06, 0.30]		0.06 (0.12)	[–0.17, 0.30]	
T2 * Stressbot	–0.36 (0.11)	[–0.57, –0.16]		–0.32 (0.12)	[–0.56, –0.09]	
T3 * Stressbot	–0.16 (0.11)	[–0.37, 0.06]		–0.13 (0.12)	[–0.36, 0.10]	
Quality of Life						
Intercept	2.73 (0.04)	[2.65, 2.82]	0.24	2.76 (0.05)	[2.66, 2.85]	0.24
Time (T2)	–0.02 (0.04)	[–0.10, 0.06]		–0.01 (0.04)	[–0.08, 0.07]	
Time (T3)	0.12 (0.04)	[0.04, 0.19]		0.08 (0.04)	[0.01, 0.15]	
Stressbot group	0.10 (0.06)	[–0.02, 0.21]		0.13 (0.08)	[–0.03, 0.29]	
T2 * Stressbot	–0.01 (0.06)	[–0.12, 0.10]		–0.01 (0.06)	[–0.13, 0.10]	
T3 * Stressbot	–0.03 (0.06)	[–0.15, 0.09]		–0.06 (0.06)	[–0.17, 0.06]	

Note. (1) Multiple imputation (MI) – pooled analyses performed on 30 datasets multiply imputed with *mice* and *mitml* R packages and (2) Completers (C) – analysis on data only from participants who completed all three measurements; Time T2/T3 – estimates representing the change in an outcome for the entire sample at the given measurement point; Stressbot group – estimates representing the difference resulting from allocation to the experimental condition, not accounting for the time point; T2/T3 * Stressbot – estimates representing the change in outcome value for participants in the Stressbot condition on the given measurement point.

0.15]), indicating an increase in quality of life with time. The effect size was small (*d* = 0.14 CI [0.00, 0.28]). However, we found no significant interaction of group and time either at T2 or T3 in either of the two models (Table 3). Therefore, the third hypothesis was not confirmed.

3.4. Sensitivity analyses

Although the short-term effect was robust to differences between the two analytical approaches, it can still be biased. This is because study dropout was higher in the Stressbot condition compared to the waitlist condition. As argued by Goldberg et al. (2021), a higher dropout in experimental conditions in internet interventions can result from participants not benefitting from or even deteriorating due to the intervention. If this is the case, the values in the experimental group that are unavailable due to attrition would represent worsening outcomes had they been recorded. This scenario violates the assumptions of MI, as it points to the missing-not-at-random missingness structure (Enders, 2022). Moreover, the realization that those who drop out could in fact deteriorate altogether challenges the validity of the effect in light of the ITT principle even when not using MI. To address this uncertainty, we followed the guidelines of Goldberg et al. (2021) and compared

outcomes of people who completed measures at T2 to scenarios in which the missing values represent outcomes that are 0.2, 0.5, and 0.8 *SD* worse than the mean residualized change score, as well as equal to the worst recorded outcome. Mann-Whitney *U* tests were conducted separately in each scenario to compare pre- to post-intervention ranked mean residualized change scores between Stressbot and waitlist conditions. The non-parametric tests were used to account for the influence of single-value imputation as suggested by Goldberg et al. (2021). Mean ranks alongside *p* values are available in Table 4. Graphs representing the comparisons are presented in Fig. 3.

In the case of coping self-efficacy, the analysis showed improvement in the intervention condition when compared to the control condition in the completers' sample (*W* = 4096, *p* < .001). The effect persisted when the missing data was assumed to be 0.20 *SD* below the mean residual (*W* = 14,482, *p* = .006), but not in the 0.50 *SD* scenario (*W* = 15,874, *p* = .161), or the 0.80 *SD* scenario (*W* = 17,138, *p* = .875). In the worst-case scenario, the effect reversed in favor of the control condition (*W* = 19,470, *p* = .032). For stress, the improvement demonstrated in the intervention condition compared to the control condition on the completers' sample (*W* = 9726, *p* < .001) persisted when missing values were assumed to be 0.20 *SD* above the mean residual (*W* = 20,358, *p* = .003).

Table 4
Sensitivity analyses.

	Coping self-efficacy				Stress				Quality of life			
	Sample	Mean rank (SD)	SE	p value	Sample	Mean rank (SD)	SE	p value	Sample	Mean rank (SD)	SE	p value
Stressbot												
Completers	98	145.44 (71.20)	7.19	<.001	102	104.15 (74.47)	7.37	<.001	98	124.02 (69.57)	7.03	.784
Worst	186	174.82 (116.39)	8.53	.032	186	197.89 (117.20)	8.59	.037	186	211.06 (104.98)	7.70	<.001
0.20 SD	186	201.64 (97.09)	7.12	.006	186	170.05 (98.42)	7.22	.003	186	180.85 (94.43)	6.92	.301
0.50 SD	186	194.16 (101.19)	7.42	.161	186	175.64 (100.95)	7.40	.047	186	177.06 (95.88)	7.03	.084
0.80 SD	186	187.36 (106.07)	7.78	.875	186	184.15 (106.21)	7.79	.668	186	174.96 (97.13)	7.12	.035
Waitlist												
Completers	146	107.10 (66.05)	5.47		148	140.22 (67.17)	5.52		146	121.48 (71.47)	5.91	
Worst	186	198.18 (91.65)	6.72		186	175.11 (91.39)	6.70		186	161.94 (99.93)	7.33	
0.20 SD	186	171.36 (111.16)	8.15		186	202.95 (110.17)	8.08		186	192.15 (115.16)	8.44	
0.50 SD	186	178.84 (109.02)	7.99		186	197.36 (109.28)	8.01		186	195.94 (113.45)	8.32	
0.80 SD	186	185.64 (104.84)	7.69		186	188.85 (105.25)	7.72		186	198.04 (111.99)	8.21	

Note. Worst – model assuming missing values to represent the worst recorded outcome; 0.20 SD, 0.50 SD, 0.80 SD - models assuming missing values to represent progressively worsening outcomes (e.g., 0.20 SD worse than the mean residual). Although the test is based on the sum of ranks, mean ranks for each group are presented for interpretation. Lower mean ranks represent better results for stress, while higher ranks represent better results for coping self-efficacy and quality of life when comparing the two groups under different assumptions; the mean ranks are non-standardized, and therefore should not be compared between different assumptions.

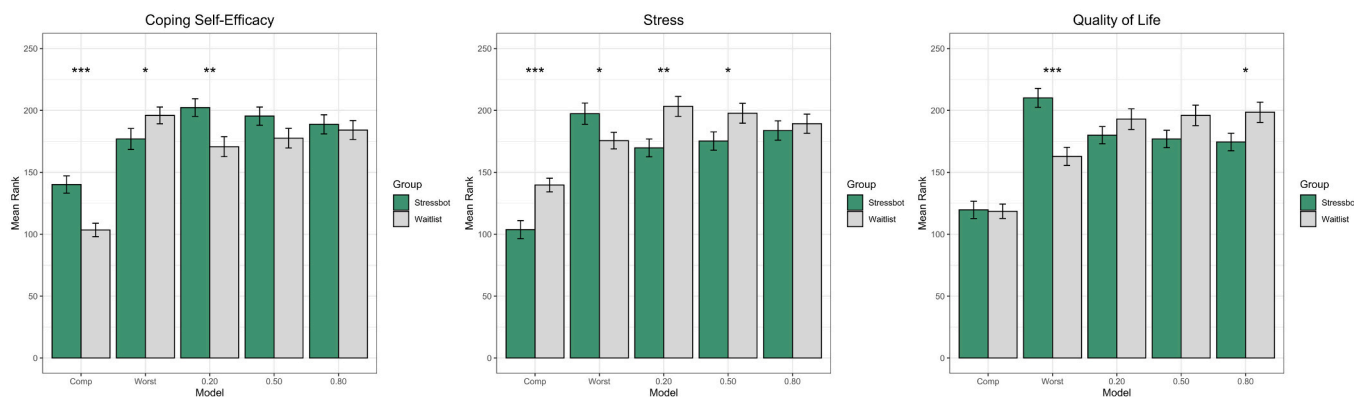


Fig. 3. Sensitivity analyses.

Note. Graphs illustrate sensitivity analyses representing situations in which the missing data would represent worsening outcomes for each of the outcome measures between T1 and T2; Comp – completers; Worst – worst-case-scenario assuming missing values to be equivalent to the worst observed outcome; 0.2, 0.5, 0.8 – conditions in which missing data represents progressively worsening outcomes; Lower mean ranks represent better results for stress, while higher ranks represent better results for coping self-efficacy and quality of life when comparing the two groups in different scenarios; The mean ranks are non-standardized, and therefore should not be compared between different assumptions; Error bars represent SE;

* $p < .05$, ** $p < .01$, *** $p < .001$.

and 0.50 SD above the mean residual ($W = 19,318, p = .047$). In the 0.80 SD scenario, the differences between groups disappeared ($W = 17,736, p = .668$), and eventually marginally switched in the worst-case scenario, demonstrating lower stress in the control group ($W = 15,179, p = .037$). For quality of life, there were no differences between groups on the complete-case sample (Table 4). This lack of effect remained until the 0.80 SD scenario when the control group eventually became favored ($W = 19,445, p = .035$). This was preserved in the worst-case scenario ($W = 12,729, p < .001$). The sensitivity analysis, therefore, showed that the short-term improvement in stress was robust to the MNAR missing data pattern resulting from higher attrition in the Stressbot group both when the unobserved data was assumed to deviate from the observed to small and moderate degrees. For coping self-efficacy, the robustness was shown only under the small deviation scenario. For quality of life, the lack of effect remained robust until a large deviation.

4. Discussion

This trial aimed to assess whether an internet intervention delivered with a chatbot on a widely adopted social media messaging platform, Meta's Messenger, would be effective. To test this, we conducted a

randomized controlled trial assessing the efficacy of Stressbot – a low-effort intervention focused on enhancing coping self-efficacy to reduce stress and improve quality of life in university students. We found that participants who were assigned to the intervention condition reported a moderate increase in coping self-efficacy compared to those assigned to the control group but only immediately after the intervention. At a one-month follow-up, a significant between-groups difference was identified for only one of the analytical approaches that we applied. This means that the follow-up effect was not robust. Thus, there was only a short-term improvement in coping self-efficacy. Similarly, a small between-groups effect for stress was found only immediately after the intervention, and not the follow-up. Lastly, we did not find significant differences between conditions on quality of life at any of the measurement points. Because the short-term effects in both coping self-efficacy and stress were consistent in both analyses (i.e., Multiple Imputation and Completers), they were robust to biases associated with each method. Furthermore, since these improvements were shown to be persistent even when the missing data was assumed to represent worsening outcomes, the overall short-term effect of the intervention can be considered robust. The increase in coping self-efficacy and reduction of stress might indicate that self-efficacy was the mechanism behind these

positive intervention effects. However, this is only speculation, since we did not test this process directly due to the methodological challenges associated with using mediation analyses for assessing causal relationships in randomized trials (Imai et al., 2013).

The immediate improvements in coping self-efficacy and stress were weaker than in previous self-efficacy-enhancing interventions (Cieslak et al., 2016; Rogala et al., 2016). Yet, these small effect sizes are in line with meta-analyses of internet interventions for stress reduction directed at university students (e.g., Amanvermez et al., 2022). Importantly and contrary to previous findings (e.g., Cieslak et al., 2016; Rogala et al., 2016), the effects of the Stressbot intervention did not persist until the follow-up. Moreover, although the intervention improved self-efficacy and stress, there was no improvement in quality of life. One explanation is that participants entered the study with already high baseline quality of life levels ($M = 2.78$, $SD = 0.55$, range = 0–4), as Stressbot was directed at a healthy population of students with no entry criteria regarding the levels of outcome variables.

The lack of effect at follow-up and lower-than-expected adherence rate (i.e., 47 %) might indicate that Messenger and similar apps are not viable means to deliver self-guided internet interventions. Yet, since the post-test effects were identified, we argue that the shortcomings are due to the current features of Stressbot itself. Namely, our implementation of a chatbot interface for a text-based intervention could have been insufficiently engaging. This could be because the interactions were limited to pressing buttons and text input was restricted to answering the bot's questions. Moreover, the daily activities could potentially still have been too time-consuming, which, combined with the text format, could have potentially led to low engagement and high dropout. Finally, the contents of the intervention focused mostly on enhancing mastery experiences to improve self-efficacy whereas it is likely important to also include other forms of increasing agency, as has been done previously in similar trials (e.g., Cieslak et al., 2016; Smoktunowicz et al., 2021).

4.1. Limitations and future directions

Some limitations of this study should be noted. First, it lacks a comparison with an active condition. We tested Stressbot in comparison to a waitlist control because we aimed to test whether using the Messenger app would be a viable delivery option to begin with, especially in light of conflicting results of previous chatbot-delivered interventions (Fitzpatrick et al., 2017; Greer et al., 2019). Yet, the necessary next step is to test whether it would be at least as effective as other successful interventions but at the same time more likely to be uptaken due to its user-friendliness. Furthermore, it is important to note that in our trial participants were not blind to their condition, which could have influenced results through a desirability bias. Specifically, those in the Stressbot condition could have potentially presented higher coping self-efficacy and lower stress at post-test due to the expectation of the intervention effect rather than a change in their self-efficacy beliefs. The way in which we measured adherence could also be refined. It was measured automatically upon completing each exercise and so we did not analyze the content of the responses. Moreover, we did not assess engagement by tracking metrics such as login frequency, duration of interactions, or points at which users disengage. Another limitation of this trial is that although we did use participants' insights from previous internet interventions that were created to enhance self-efficacy, we did not use a participatory approach when developing this particular program. Finally, the efficacy of Stressbot does not mean that this accessible format of delivering interventions is suitable for every case. While we showed the effects of this brief intervention intended for a general population of university students, the low-effort format could produce effects too weak for clinical use.

Future studies should further evaluate the potential for low-effort chatbot-delivered interventions by focusing on investigating engagement. Factors such as the number and duration of interactions, or the dropout point could be of interest. Furthermore, to develop the optimum

low-effort intervention, a complex user-centric approach should be used. This could take the form of including participants in codesigning and evaluating the intervention. Studies should also employ strategies to make chat-based interactions more engaging. The use of voice- or picture-based input are some obvious enhancements. Additionally, in light of the recent advancements in AI, such as the introduction of Open AI's Chat GPT, it is clear that in the future, psychological internet interventions will eventually implement AI, but should be done with caution (Boucher et al., 2021; Carlbring et al., 2023). More exercises or boosters should also be considered to enhance the efficacy of similar interventions. Yet, this could challenge the intended low-intensity format. Perhaps two divergent directions for future low-effort interventions exist: either focusing on longer, more intensive programs but with an easy-to-use format, or conversely, on short but more engaging interventions. The former would focus on building long-term effects, while the latter could be used as short-term, in-the-moment help.

5. Conclusions

In sum, the results of this study confirm that an internet intervention can be successfully delivered with a chatbot on a popular social media app – Meta's Messenger. Although Stressbot produced a smaller effect than similar interventions, it offered a promising delivery design that required less effort and was less time-consuming for participants.

Funding

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

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Adamopoulou, E., Moussiades, L., 2020. Chatbots: history, technology, and applications. *Mach. Learn. Applic.* 2, e100006 <https://doi.org/10.1016/j.mlwa.2020.100006>.
- Amanvermez, Y., Zhao, R., Cuijpers, P., de Wit, L.M., Ebert, D.D., Kessler, R.C., Bruffaerts, R., Karyotaki, E., 2022. Effects of self-guided stress management interventions in college students: a systematic review and meta-analysis. *Internet Interv.* 28, e100503 <https://doi.org/10.1016/j.invent.2022.100503>.
- Auerbach, R.P., Mortier, P., Bruffaerts, R., Alonso, J., Benjet, C., Cuijpers, P., Demyttenaere, K., Ebert, D.D., Green, J.G., Hasking, P., Murray, E., Nock, M.K., Pinder-Amaker, S., Sampson, N.A., Stein, D.J., Vilagut, G., Zaslavsky, A.M., Kessler, R.C., WHO WMH-ICS. Collaborators, 2018. WHO world mental health surveys international college student project: prevalence and distribution of mental disorders. *J. Abnorm. Psychol.* 127 (7), 623–638. <https://doi.org/10.1037/abn0000362>.
- Bandura, A., 1997. *Self-efficacy: The Exercise of Control*, 1st edition. Worth Publishers.
- Baumel, A., Muench, F.J., 2021. Effort-optimized intervention model: framework for building and analyzing digital interventions that require minimal effort for health-related gains. *J. Med. Internet Res.* 23 (3), e24905 <https://doi.org/10.2196/24905>.
- Bockting, C.L.H., Williams, A.D., Carswell, K., Grech, A.E., 2016. The potential of low-intensity and online interventions for depression in low- and middle-income countries. *Glob. Ment. Health* 3, e25. <https://doi.org/10.1017/gmh.2016.21>.
- Borghouts, J., Eikev, E., Mark, G., Leon, C.D., Schueller, S.M., Schneider, M., Stadnick, N., Zheng, K., Mukamel, D., Sorkin, D.H., 2021. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. *J. Med. Internet Res.* 23 (3), e24387 <https://doi.org/10.2196/24387>.
- Boucher, E.M., Harake, N.R., Ward, H.E., Stoeckl, S.E., Vargas, J., Minkel, J., Parks, A.C., Zilca, R., 2021. Artificially intelligent chatbots in digital mental health interventions: a review. *Expert Rev. Med. Devices* 18, 37–49. <https://doi.org/10.1080/17434440.2021.2013200>.
- Bücker, L., Westermann, S., Kühn, S., Moritz, S., 2019. A self-guided Internet-based intervention for individuals with gambling problems: study protocol for a randomized controlled trial. *Trials* 20, 74. <https://doi.org/10.1186/s13063-019-3176-z>.
- Carlbring, P., Hadjistavropoulos, H., Kleiboer, A., Andersson, G., 2023. A new era in internet interventions: the advent of Chat-GPT and AI-assisted therapist guidance. *Internet Interv.* e100621 <https://doi.org/10.1016/j.invent.2023.100621>.

- Chesney, M.A., Neilands, T.B., Chambers, D.B., Taylor, J.M., Folkman, S., 2006. A validity and reliability study of the coping self-efficacy scale. *Br. J. Health Psychol.* 11 (3), 421–437. <https://doi.org/10.1348/135910705X53155>.
- Cieslak, R., Benight, C.C., Rogala, A., Smoktunowicz, E., Kowalska, M., Zukowska, K., Yeager, C., Luszczynska, A., 2016. Effects of internet-based self-efficacy intervention on secondary traumatic stress and secondary posttraumatic growth among health and human services professionals exposed to indirect trauma. *Front. Psychol.* 7 <https://doi.org/10.3389/fpsyg.2016.01009>.
- Cohen, S., Kamarck, T., Mermelstein, R., 1983. A global measure of perceived stress. *J. Health Soc. Behav.* 24 (4), 385. <https://doi.org/10.2307/2136404>.
- Dederichs, M., Weber, J., Pischke, C.R., Angerer, P., Apolinário-Hagen, J., 2021. Exploring medical students' views on digital mental health interventions: a qualitative study. *Internet Interv.* 25, 100398 <https://doi.org/10.1016/j.invent.2021.100398>.
- Dixon, S., 2022. Percentage of U.S. Internet Users Who Use Facebook Messenger as of January 2018, by Age Group. Statista. <https://www.statista.com/statistics/814100/share-of-us-internet-users-who-use-facebook-messenger-by-age/>.
- Enders, C.K., 2022. *Applied Missing Data Analysis (Second Edition)*. The Guilford Press.
- Fitzpatrick, K.K., Darcy, A., Vierhile, M., 2017. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR Ment. Health* 4 (2), e19. <https://doi.org/10.2196/mental.7785>.
- Goldberg, S.B., Bolt, D.M., Davidson, R.J., 2021. Data missing not at random in mobile health research: assessment of the problem and a case for sensitivity analyses. *J. Med. Internet Res.* 23 (6), e26749 <https://doi.org/10.2196/26749>.
- Greer, S., Ramo, D., Chang, Y.-J., Fu, M., Moskowitz, J., Haritatos, J., 2019. Use of the Chatbot “Vivibot” to deliver positive psychology skills and promote well-being among young people after cancer treatment: randomized controlled feasibility trial. *JMIR MHealth UHealth* 7 (10), e15018. <https://doi.org/10.2196/15018>.
- Hedman-Lagerlöf, E., Carlbring, P., Svärdman, F., Ripper, H., Cuijpers, P., Andersson, G., 2023. Therapist-supported internet-based cognitive behaviour therapy yields similar effects as face-to-face therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. *World Psychiatry* 22 (2), 305–314. <https://doi.org/10.1002/wps.21088>.
- Hobfoll, S.E., 1989. Conservation of resources: a new attempt at conceptualizing stress. *Am. Psychol.* 44 (3), 513–524. <https://doi.org/10.1037/0003-066X.44.3.513>.
- Imai, K., Tingley, D., Yamamoto, T., 2013. Experimental designs for identifying causal mechanisms: *experimental designs for identifying causal mechanisms*. *J. R. Stat. Soc. Ser. A Stat. Soc.* 176 (1), 5–51. <https://doi.org/10.1111/j.1467-985X.2012.01032.x>.
- Independent Association of Students [Niezależne Zrzeszenie Studentów] (NZS), 2021. Mental Health of Students. Report by the Analytical Center of NZS and PSSiAP. [Zdrowie Psychiczne Studentów. Raport Centrum Analiz NZS i PSSiAP.]. https://nzs.org.pl/wp-content/uploads/2021/01/Zdrowie-Psychiczne-Studentow_Raport-CA-NZS.pdf.
- Karyotaki, E., Ripper, H., Twisk, J., Hoogendoorn, A., Kleiboer, A., Mira, A., Mackinnon, A., Meyer, B., Botella, C., Littlewood, E., Andersson, G., Christensen, H., Klein, J.P., Schröder, J., Bretón-López, J., Scheider, J., Griffiths, K., Farrer, L., Huibers, M.J.H., Cuijpers, P., 2017. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. *JAMA Psychiatry* 74 (4), 351–359. <https://doi.org/10.1001/jamapsychiatry.2017.0044>.
- Karyotaki, E., Efthimiou, O., Miguel, C., Bermpohl, F.M. genannt, Furukawa, T.A., Cuijpers, P., Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration, 2021. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network Meta-analysis. *JAMA Psychiatry* 78 (4), 361–371. <https://doi.org/10.1001/jamapsychiatry.2020.4364>.
- Kemp, Simon, 2021. Digital 2021: Global Overview Report. <https://datareportal.com/reports/digital-2021-global-overview-report>.
- Kretschmar, K., Tyroll, H., Pavarini, G., Manzini, A., Singh, I., NeurOx Young People's Advisory Group, 2019. Can your phone be your therapist? Young people's ethical perspectives on the use of fully automated conversational agents (Chatbots) in mental health support. *Biomed. Inform. Insights* 11, e117822261982908. <https://doi.org/10.1177/1178222619829083>.
- Linardon, J., Fuller-Tyszkiewicz, M., 2020. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. *J. Consult. Clin. Psychol.* 88 (1), 1–13. <https://doi.org/10.1037/ccp0000459>.
- Linardon, J., Cuijpers, P., Carlbring, P., Messer, M., Fuller-Tyszkiewicz, M., 2019. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. *World Psychiatry* 18 (3), 325–336. <https://doi.org/10.1002/wps.20673>.
- Lindner, P., Frykheden, O., Forsström, D., Andersson, E., Ljótsson, B., Hedman, E., Andersson, G., Carlbring, P., 2016. The Brunsviken brief quality of life scale (BBQ): development and psychometric evaluation. *Cogn. Behav. Ther.* 45 (3), 182–195. <https://doi.org/10.1080/16506073.2016.1143526>.
- Little, R.J.A., Rubin, D.B., 2002. *Statistical Analysis With Missing Data*: Little/statistical Analysis With Missing Data. John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119013563>.
- Liu, H., Peng, H., Song, X., Xu, C., Zhang, M., 2022. Using AI chatbots to provide self-help depression interventions for university students: a randomized trial of effectiveness. *Internet Interv.* 27, e100495 <https://doi.org/10.1016/j.invent.2022.100495>.
- Logan, B., Burns, S., 2021. Stressors among young Australian university students: a qualitative study. *J. Am. Coll. Heal.* 1–8 <https://doi.org/10.1080/07448481.2021.1947303>.
- Magnusson, K., 2019. *Methodological Issues in Psychological Treatment Research: Applications to Gambling Research and Therapist Effects*. Karolinska Institutet.
- Mohr, D.C., Weingardt, K.R., Reddy, M., Schueller, S.M., 2017. Three problems with current digital mental health research. . . and three things we can do about them. *Psychiatr. Serv.* 68 (5), 427–429. <https://doi.org/10.1176/appi.ps.201600541>.
- Rogala, A., Smoktunowicz, E., Żukowska, K., Kowalska, M., Cieślak, R., 2016. The helpers' stress: effectiveness of a web-based intervention for professionals working with trauma survivors in reducing job burnout and improving work engagement. *Med. Pr.* 67 (2), 223–237. <https://doi.org/10.13075/mp.5893.00220>.
- Smith, K.A., Blease, C., Faurholt-Jepsen, M., Firth, J., Van Daele, T., Moreno, C., Carlbring, P., Ebner-Priemer, U.W., Koutsouleris, N., Ripper, H., Mouchabac, S., Torous, J., Cipriani, A., 2023. Digital mental health: challenges and next steps. *BMJ Ment. Health* 26 (1), e300670. <https://doi.org/10.1136/bmjment-2023-300670>.
- Smoktunowicz, E., Barak, A., Andersson, G., Banos, R.M., Berger, T., Botella, C., Dear, B. F., Donker, T., Ebert, D.D., Hadjistavropoulos, H., Hodgins, D.C., Kaldov, V., Mohr, D. C., Nordgreen, T., Powers, M.B., Ripper, H., Ritterband, L.M., Rozenal, A., Schueller, S.M., Carlbring, P., 2020. Consensus statement on the problem of terminology in psychological interventions using the internet or digital components. *Internet Interv.* 21, e100331 <https://doi.org/10.1016/j.invent.2020.100331>.
- Smoktunowicz, E., Lesnierowska, M., Carlbring, P., Andersson, G., Cieślak, R., 2021. Resource-based internet intervention (med-stress) to improve well-being among medical professionals: randomized controlled trial. *J. Med. Internet Res.* 23 (1), e21445 <https://doi.org/10.2196/21445>.
- Sullivan, T.R., White, I.R., Salter, A.B., Ryan, P., Lee, K.J., 2018. Should multiple imputation be the method of choice for handling missing data in randomized trials? *Stat. Methods Med. Res.* 27 (9), 2610–2626. <https://doi.org/10.1177/0962280216683570>.
- Svärdman, F., Sjöwall, D., Lindsäter, E., 2022. Internet-delivered cognitive behavioral interventions to reduce elevated stress: a systematic review and meta-analysis. *Internet Interv.* 29, e100553 <https://doi.org/10.1016/j.invent.2022.100553>.