

Supplementary Information

Combining AI and human support in mental health: a digital intervention with comparable effectiveness to human-delivered care

Clare E Palmer¹, Emily Marshall¹, Edward Millgate¹, Graham Warren¹, Michael P. Ewbank¹, Elisa Cooper¹, Samantha Lawes¹, Malika Bouazzaoui¹, Alastair Smith¹, Chris Hutchins-Joss¹, Jessica Young¹, Morad Margoum², Sandra Healey², Louise Marshall¹, Shaun Mehew¹, Ronan Cummins¹, Valentin Tablan¹, Ana Catarino¹, Andrew E Welchman¹ and Andrew D Blackwell¹

¹ieso Digital Health, The Jeffrey's Building, Cowley Road, Cambridge, CB4 0DS, UK

²Dorset HealthCare University NHS Foundation, Sentinel House, Nuffield Industrial Estate, Nuffield Road, Poole, UK

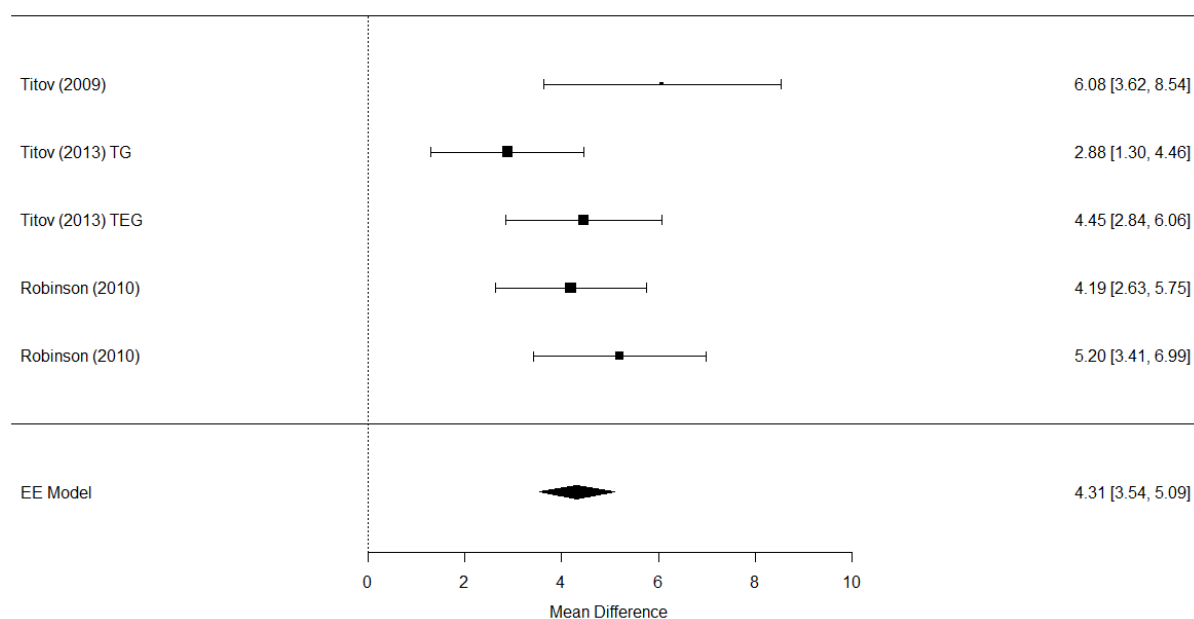
Supplementary Information Table of Contents

Supplementary Methods	
A priori definition of non-inferiority margin	2
Supplementary Tables	
Supplementary Table 1. Description of each module topic in the digital program.	3
Supplementary Table 2. Propensity-matching between groups.	4
Supplementary Table 3. Output from slope analysis comparing adherence rate over GAD-7 check-ins across groups.	5
Supplementary Table 4. Output of linear regression model measuring the association between participant characteristics and adherence for the ITT sample.	6
Supplementary Table 5. Output of logistic regression model measuring association between participant characteristics and non-adherence.	7
Supplementary Table 6. Rates of improvement, recovery, reliable recovery, responder and remission from baseline to final score across all groups.	8
Supplementary Table 7. Mean GAD-7 score across assessments stratified by baseline GAD-7 severity.	9
Supplementary Table 8. Change in GAD-7, PHQ-9 and WSAS scores from baseline to final score stratified by baseline GAD-7 severity for the intervention sample.	10
Supplementary Table 9. Output of linear regression model measuring association between participant characteristics and change in GAD-7 scores for the ITT intervention sample.	11
Supplementary Table 10. Mean PHQ-9 score across assessments stratified by baseline GAD-7 severity.	12

Supplementary Methods

A priori definition of non-inferiority margin

The Food and Drugs Administration (FDA) recommends that the threshold for non-inferiority is set based on estimates of the active comparator in previously conducted studies (70). This threshold can be defined, for example, as 50% or less of the lower confidence interval of the expected effect of the active comparator vs placebo. While it is difficult to develop an appropriate placebo in the context of psychotherapy research, previous studies exploring the efficacy of internet-delivered psychotherapy for GAD, against a waiting list control group, demonstrate significant clinical benefits of these interventions. In these studies, medium to large between-group effect sizes are reported, ranging from 0.38 to 1.25 (38–40). The outcomes from these studies are aggregated using a fixed effect meta-analysis, results shown below.



Here we define the non-inferiority margin as a GAD-7 score change of 1.8. This is equivalent to 50% of the lower limit of the 95% confidence interval of the fixed effect meta-analysis pooled result (i.e. 3.54/2).

Supplementary Tables

Supplementary Table 1. Description of each module topic in the digital program. The digital program consisted of six modules with three sessions each following a pattern of (1) learning; (2) activity; (3) practice. The final modules (5 + 6) focused on consolidation of learning.

Module	Topic	Description
1	Getting to Know You + Learning About Worry & Anxiety	Questions, reflections and information designed to help increase understanding of difficulties and prepare for making change.
2	Holding Thoughts Lightly	Focus on unhelpful patterns of thinking with activities and practices designed to promote more flexible ways to respond to thoughts.
3	Making Meaningful Moves Towards What Matters	Focus on understanding and targeting unhelpful avoidance behaviors, with activities and practices designed to gradually reduce these unhelpful behaviors.
4	Taking a Different Perspective	Focus on unhelpful beliefs about worry which may be maintaining anxiety symptoms, with activities and practices designed to reframe these beliefs.
5 + 6	Continuing On Your Journey and Bringing it All Together	Focus on consolidation of skills and techniques. Planning on maintaining progress once the program has been completed.

Supplementary Table 2. Propensity-matching between groups. Mean demographic data, baseline anxiety (GAD-7) and mood (PHQ-9) scores for the ITT sample (“ieso Digital Program”) and mean data from the propensity-matched external control groups for waiting controls, face-to-face CBT and typed CBT.

Variable	Per-protocol		Intention-to-treat		
	ieso Digital Program mean (sd) (N=169)	Waiting control mean (sd) (N=169)	ieso Digital Program mean (sd) (N=299)	Face-to-face CBT mean (sd) (N=299)	Typed CBT mean (sd) (N=299)
Age	41.7 (11.8)	41.7 (13.3)	39.8 (12.8)	40.1 (16.6)	39.8 (12.7)
Baseline GAD-7	12.4 (3.4)	12.5 (3.3)	12.5 (3.3)	12.9 (3.1)	12.6 (3.5)
Baseline PHQ-9	8.0 (3.8)	8.4 (3.4)	8.0 (3.7)	8.4 (3.6)	8.1 (3.6)
LTC ('yes')	70 (41.4)	78 (46.2)	114 (38.1)	119 (39.8)	108 (36.1)
LTC ('no')	91 (53.8)	90 (53.3)	167 (55.9)	180 (60.2)	173 (57.9)
LTC ('not known')	8 (4.7)	1 (0.6)	18 (6.0)	0	18 (6.0)

Supplementary Table 3. Output from slope analysis comparing adherence rate over GAD-7 check-ins across groups. Adherence rate was defined as proportion of participants completing the GAD-7 check-in (either within the ieso digital program or at the start of their treatment session for the control groups) across seven total instances of GAD-7 data collection. Overlapping confidence intervals indicate no significant difference between groups in adherence rate.

Group	Slope	Std. Error	L95%CI	U95%CI
ieso Digital Program	-10.29	0.26	-10.80	-9.78
Face-to-face CBT	-9.55	0.31	-10.16	-8.95
Typed CBT	-10.99	0.58	-12.14	-9.85

Supplementary Table 4. Output of linear regression model measuring the association between participant characteristics and adherence for the ITT sample. Adherence was defined as the number of completed sessions in the digital program. All demographic data, GAD-7 score, and PHQ-9 score were collected at baseline. Enrolment path refers to whether a participant was a referred patient through NHS TT or was externally recruited. Reference values for each categorical variable were: chronic health condition = yes; gender = female; software version = 1; disability = yes; sexual orientation = heterosexual; employment = employed; ethnicity = white; qualification = degree; religion = Christian; enrolment path = NHS patient. * = $p < .05$; ** = $p < .005$.

Variable	Estimate	Std. Error	t value	p-value	significance
(Intercept)	10.083	3.360	3.001	0.0029	**
Baseline score: GAD-7	-0.082	0.155	-0.531	0.5959	
Baseline score :PHQ-9	0.060	0.133	0.448	0.6543	
Age (at screening)	0.112	0.042	2.690	0.0076	*
Chronic health condition: no	-0.156	1.091	-0.143	0.8863	
Chronic health condition: not known	-2.144	2.571	-0.834	0.4051	
Gender: male	-1.494	1.326	-1.126	0.2611	
Gender: other	-1.171	4.210	-0.278	0.7812	
Gender: not known	-0.718	5.159	-0.139	0.8894	
Software version: version 2	0.095	1.169	0.081	0.9355	
Disability: no perceived disability	0.620	1.292	0.480	0.6317	
Disability: prefer not to say	-1.549	3.899	-0.397	0.6914	
Sexual orientation: not known	2.065	2.789	0.740	0.4598	
Sexual orientation: other	1.741	1.396	1.247	0.2135	
Employment: not employed	-1.921	1.372	-1.400	0.1627	
Employment: not known	-6.595	3.399	-1.940	0.0534	
Ethnicity: not known	4.382	5.119	0.856	0.3928	
Ethnicity: other	0.740	1.859	0.398	0.6909	
Qualification: below degree	-1.371	1.214	-1.130	0.2595	
Qualification: not known	-2.040	6.078	-0.336	0.7374	
Qualification: other	-1.330	3.825	-0.348	0.7282	
Qualification: postgraduate	1.786	1.135	1.573	0.1169	
Religion: not known	-0.847	3.021	-0.280	0.7795	
Religion: other	-1.064	1.127	-0.944	0.3458	
Enrolment path: external recruit	0.887	1.114	0.796	0.4266	

Residual standard error: 7.835 on 274 degrees of freedom

Multiple R-squared: 0.1091, Adjusted R-squared: 0.0311

F-statistic: 1.399 on 24 and 274 DF, p-value: 0.1059

Supplementary Table 5. Output of logistic regression model measuring association between participant characteristics and non-adherence. Non-adherence was defined based on PP sample status, i.e. non-adherence (coded = 1) included all those *not* in the PP sample. All demographic data, GAD-7 score and PHQ-9 score were collected at baseline. Enrolment path refers to whether a participant was a referred patient through NHS TT or was externally recruited. Reference values for each categorical variable were: chronic health condition = yes; gender = female; software version = 1; disability = yes; sexual orientation = heterosexual; employment = employed; ethnicity = white; qualification = degree; religion = Christian; enrolment path = NHS patient. * = $p < .05$.

Variable	Estimate	Odds Ratio	Std. Error	z value	p-value	Significance
(Intercept)	0.846	2.33	0.908	0.932	0.3515	
Baseline score: GAD-7	0.000	1.00	0.042	0.003	0.9976	
Baseline score :PHQ-9	-0.027	0.97	0.036	-0.740	0.4590	
Age (at screening)	-0.030	0.97	0.011	-2.607	0.0091	*
Chronic health condition: no LTC	0.227	1.25	0.296	0.766	0.4439	
Chronic health condition: not known	0.252	1.29	0.681	0.371	0.7107	
Gender: male	0.204	1.23	0.358	0.570	0.5685	
Gender: other	0.209	1.23	1.126	0.186	0.8527	
Gender: not known	-0.353	0.70	1.665	-0.212	0.8319	
Product version: version 2	0.009	1.01	0.315	0.027	0.9783	
Disability: no perceived disability	-0.189	0.83	0.351	-0.538	0.5909	
Disability: prefer not to say	0.744	2.10	1.085	0.686	0.4926	
Sexual orientation: not known	-0.555	0.57	0.768	-0.723	0.4697	
Sexual orientation: other	-0.398	0.67	0.381	-1.044	0.2963	
Employment: not employed	0.694	2.00	0.372	1.863	0.0624	
Employment: not known	1.196	3.31	0.950	1.258	0.2082	
Ethnicity: not known	-0.644	0.53	1.672	-0.385	0.7001	
Ethnicity: other	0.159	1.17	0.503	0.317	0.7513	
Qualification: below degree	0.493	1.64	0.327	1.510	0.1311	
Qualification: not known	0.457	1.58	1.595	0.287	0.7745	
Qualification: other	1.096	2.99	1.124	0.974	0.3299	
Qualification: postgraduate	-0.074	0.93	0.309	-0.239	0.8110	
Religion: not known	-0.201	0.82	0.829	-0.243	0.8082	
Religion: other	0.413	1.51	0.309	1.337	0.1812	
Enrolment path: external recruit	-0.387	0.68	0.297	-1.305	0.1920	

Null deviance: 409.40 on 298 degrees of freedom

Residual deviance: 380.51 on 274 degrees of freedom; p-value = 2.123265e-05

AIC: 430.51

Supplementary Table 6. Rates of improvement, recovery, reliable recovery, responder and remission from baseline to final score across all groups. Improvement was defined as a reduction on the PHQ-9 or GAD-7 scales greater than the reliable change index (≥ 4 for GAD-7; ≥ 6 for PHQ-9) and no reliable increase on either measure. Recovery was defined as reduction on both scales to below the clinical cutoff (PHQ-9 score < 10 and GAD-7 score < 8). Reliable recovery was defined as having both improved and recovered. Responder rate was defined as an improvement of ≥ 4 on the GAD-7 or ≥ 6 on the PHQ-9. Remission rate was defined as having a final GAD-7 < 8 or PHQ-9 < 10 , i.e. below the clinical cut-off. Unlike improvement, recovery, and reliable recovery, responder and remission rates were calculated for each questionnaire independently. For PHQ-9, remission and responder rates were only calculated for those above clinical cut-off at baseline (i.e. PHQ-9 ≥ 10 at baseline).

Sample	Arm	N Total	Improvement		Recovery		Reliable Recovery		GAD-7 Remission		GAD-7 Responder		PHQ > cut-off	PHQ-9 Remission		PHQ-9 Responder	
			N	%	N	%	N	%	N	%	N	%	N	N	%	N	%
Per-protocol	Waiting control	169	61	36.1	46	27.2	34	20.1	48	28.4	56	33.1	66	36	54.5	16	9.5
	ieso Digital Program	169	139	82.2	130	76.9	122	72.2	136	80.5	138	81.7	69	57	82.6	46	27.2
	Face-to-face CBT	253	189	74.7	164	64.8	157	62.1	170	67.2	185	73.1	103	84	81.6	78	30.8
	Typed CBT	229	195	85.2	196	85.6	180	78.6	198	86.5	192	83.8	76	66	86.8	81	35.4
Intention-to-treat	ieso Digital Program	299	198	66.2	174	58.2	161	53.8	156	76.8	160	78.8	80	63	78.8	49	24.1
	Face-to-face CBT	299	211	70.6	189	63.2	178	59.5	196	65.6	207	69.2	116	89	76.7	86	28.8
	Typed CBT	299	230	76.9	226	75.6	204	68.2	229	76.6	228	76.3	104	82	78.8	88	29.4

Supplementary Table 7. Mean GAD-7 score across assessments stratified by baseline GAD-7 severity. Mean GAD-7 scores for both PP and ITT samples. Stratification by severity based on baseline GAD-7 score. Check-in scores were collected prior to each module within the digital program software. Screening, completion and follow-up scores were collected outside of the digital program. Data reported are for each specific data collection point.

Study timepoint	Sample	Per-protocol				Intention-to-treat			
	Severity	Mild	Moderate	Severe	Overall	Mild	Moderate	Severe	Overall
Screening	N	39	82	48	169	62	150	87	299
	Mean (95%CI)	8.6 (8.4, 8.7)	11.7 (11.3, 12.0)	16.9 (16.4, 17.4)	12.4 (11.9, 13.0)	8.5 (8.4, 8.6)	11.7 (11.5, 11.9)	16.9 (16.5, 17.2)	12.5 (12.2, 12.9)
Check-in 1	N	39	82	48	169	59	144	81	284
	Mean (95%CI)	9.8 (8.5, 11.1)	11.2 (10.4, 12.0)	14.0 (12.9, 15.2)	11.7 (11.1, 12.3)	9.9 (8.9, 10.9)	10.9 (10.3, 11.5)	14.0 (13.2, 14.9)	11.6 (11.1, 12.1)
Check-in 2	N	39	82	48	169	54	120	66	240
	Mean (95%CI)	6.0 (5.1, 6.9)	8.1 (7.3, 8.9)	9.2 (8.1, 10.4)	7.9 (7.4, 8.5)	6.4 (5.5, 7.3)	8.1 (7.4, 8.7)	9.6 (8.5, 10.7)	8.1 (7.6, 8.6)
Check-in 3	N	39	82	48	169	46	106	57	209
	Mean (95%CI)	5.8 (4.7, 6.8)	7.0 (6.2, 7.9)	8.2 (6.9, 9.5)	7.1 (6.5, 7.7)	5.6 (4.7, 6.6)	7.2 (6.5, 8.0)	8.5 (7.3, 9.7)	7.2 (6.7, 7.8)
Check-in 4	N	39	82	48	169	40	90	50	180
	Mean (95%CI)	4.5 (3.7, 5.3)	5.9 (5.1, 6.6)	7.0 (5.8, 8.2)	5.8 (5.3, 6.4)	4.8 (3.8, 5.7)	5.9 (5.2, 6.6)	7.1 (5.9, 8.3)	6.0 (5.5, 6.5)
Check-in 5	N	31	68	36	135	31	70	37	138
	Mean (95%CI)	4.7 (3.6, 5.8)	5.4 (4.6, 6.1)	6.9 (5.2, 8.6)	5.6 (5.0, 6.2)	4.7 (3.6, 5.8)	5.3 (4.6, 6.0)	7.3 (5.5, 9.1)	5.7 (5.0, 6.3)
Check-in 6	N	23	61	28	112	23	62	28	113
	Mean (95%CI)	3.9 (2.8, 5.1)	5.2 (4.4, 6.1)	6.5 (4.6, 8.4)	5.3 (4.6, 6.0)	3.9 (2.8, 5.1)	5.2 (4.3, 6.0)	6.5 (4.6, 8.4)	5.2 (4.5, 5.9)
Completion	N	39	82	48	169	46	101	56	203
	Mean (95%CI)	4.2 (3.3, 5.1)	4.8 (4.2, 5.5)	6.2 (4.8, 7.6)	5.1 (4.5, 5.6)	4.3 (3.4, 5.1)	5.4 (4.7, 6.0)	6.7 (5.3, 8.0)	5.5 (4.9, 6.0)
Follow-up	N	39	80	47	166	48	106	56	210
	Mean (95%CI)	4.5 (3.5, 5.4)	4.7 (4.0, 5.5)	6.3 (4.8, 7.6)	5.1 (4.5, 5.7)	4.6 (3.8 ; 5.5)	5.5 (4.7, 6.2)	6.7 (5.3, 8.1)	5.6 (5.0, 6.2)

Supplementary Table 8. Change in GAD-7, PHQ-9 and WSAS scores from baseline to final score stratified by baseline GAD-7 severity for the intervention sample. Change (i.e. mean difference) in GAD-7, PHQ-9 and WSAS scores was calculated using the difference between baseline and final scores for the digital intervention group. A negative mean difference denotes a reduction in scores. Within-subject effect sizes (Cohen's *d*) calculated for the total sample and each severity subgroup based on baseline GAD-7 severity stratification.

Sample	GAD-7 severity	N	Baseline score		Change in score				
			Mean	SD	Mean	SD	Lower 95% CI	Upper 95% CI	Within-subjects effect size (<i>d</i>)
<i>GAD-7</i>									
Per-protocol	Mild	39	8.6	0.5	-4.4	3.0	-3.4	-5.3	1.4
	Moderate	82	11.7	1.5	-6.8	3.4	-6.1	-7.6	2.0
	Severe	48	16.9	1.7	-10.7	5.3	-9.2	-12.3	2.0
	Overall	169	12.4	3.4	-7.4	4.6	-6.7	-8.1	1.6
Intention-to-treat	Mild	62	8.5	0.5	-2.9	4.1	-1.9	-4.0	0.7
	Moderate	150	11.7	1.4	-5.0	4.2	-4.3	-5.7	1.2
	Severe	87	16.9	1.6	-7.9	6.0	-6.6	-9.2	1.3
	Overall	299	12.5	3.3	-5.4	5.1	-4.8	-6.0	1.1
<i>PHQ-9</i>									
Per-protocol	Mild	39	6.9	3.5	-2.3	3.0	-1.3	-3.3	0.8
	Moderate	82	7.5	3.7	-2.7	4.4	-1.8	-3.7	0.6
	Severe	48	9.7	3.8	-4.5	5.5	-2.9	-6.1	0.8
	Overall	169	8.0	3.8	-3.1	4.5	-2.4	-3.8	0.7
Intention-to-treat	Mild	62	6.7	3.1	-1.3	4.0	-0.3	-2.3	0.3
	Moderate	150	7.5	3.5	-1.3	4.6	-0.6	-2.0	0.3
	Severe	87	9.7	3.9	-2.3	5.5	-1.2	-3.5	0.4
	Overall	299	8.0	3.7	-1.6	4.8	-1.1	-2.1	0.3
<i>WSAS</i>									
Per-protocol	Mild	39	12.9	5.9	-3.7	6.0	-1.7	-5.6	0.6
	Moderate	82	15.1	6.7	-5.7	6.0	-4.3	-7.0	0.9
	Severe	48	17.4	5.5	-6.0	6.4	-4.2	-7.9	0.9
	Overall	169	15.3	6.4	-5.3	6.2	-4.4	-6.2	0.9
Intention-to-treat	Mild	58	12.2	5.6	-3.4	6.0	-1.6	-5.2	0.6
	Moderate	150	14.9	6.9	-4.8	6.4	-3.5	-6.1	0.8
	Severe	87	16.8	6.1	-5.5	6.9	-3.7	-7.4	0.8
	Overall	295	14.9	6.6	-4.7	6.5	-3.8	-5.6	0.7

Supplementary Table 9. Output of linear regression model measuring association between participant characteristics and change in GAD-7 scores for the ITT intervention sample. Dependent variable was change in GAD-7 score from baseline to final score. All demographic data, GAD-7 score and PHQ-9 score were collected at baseline. Enrolment path refers to whether a participant was a referred patient from NHS TT or was externally recruited. Reference values for each categorical variable were: chronic health condition = yes; gender = female; software version = 1; disability = yes; sexual orientation = heterosexual; employment = employed; ethnicity = white; qualification = degree; religion = Christian; enrolment path = NHS patient. * = $p < .05$; ** = $p < .005$; *** = $p < .001$

Variable	Estimate	Std. Error	t value	p-value	significance
(Intercept)	-4.126	1.993	-2.071	0.0393	*
Baseline score: GAD-7	0.696	0.092	7.575	0.0000	***
Baseline score :PHQ-9	-0.125	0.079	-1.584	0.1143	
Age (at screening)	0.074	0.025	2.999	0.0030	**
Chronic health condition: no LTC	-0.516	0.647	-0.798	0.4256	
Chronic health condition: not known Known	-1.749	1.525	-1.147	0.2523	
Gender: male	0.177	0.787	0.225	0.8219	
Gender: other	-2.606	2.497	-1.044	0.2975	
Gender: not known	0.253	3.059	0.083	0.9342	
Product version: version 2	0.022	0.693	0.031	0.9751	
Disability: no perceived disability	-0.222	0.766	-0.290	0.7722	
Disability: prefer not to say	-0.156	2.312	-0.067	0.9464	
Sexual orientation: not known	-1.156	1.654	-0.699	0.4853	
Sexual orientation: other	0.812	0.828	0.981	0.3274	
Employment: not employed	-1.589	0.814	-1.953	0.0519	.
Employment: not known	-1.081	2.016	-0.536	0.5921	
Ethnicity: not known	1.052	3.035	0.347	0.7292	
Ethnicity: other	1.240	1.103	1.124	0.2619	
Qualification: below degree	-0.372	0.720	-0.516	0.6061	
Qualification: not known	-1.272	3.604	-0.353	0.7245	
Qualification: other	-0.333	2.268	-0.147	0.8833	
Qualification: postgraduate	-0.756	0.673	-1.123	0.2626	
Religion: not known	0.010	1.792	0.006	0.9954	
Religion: other	0.104	0.668	0.156	0.8759	
Enrolment path: external recruit	-0.250	0.661	-0.378	0.7056	

Residual standard error: 4.656 on 274 degrees of freedom
 Multiple R-squared: 0.232, Adjusted R-squared: 0.1647 F-statistic: 3.449 on 24 and 274 DF, p-value: 3.546e-07

Supplementary Table 10. Mean PHQ-9 score across assessments stratified by baseline GAD-7 severity. Mean scores for both PP and ITT samples. Stratification by severity based on baseline GAD-7 scores. Check-in scores were collected prior to each module within the digital program software. Screening, completion and follow-up scores were collected outside of the digital program. Data reported are for each specific data collection point.

Study timepoint	Sample	Per-protocol				Intention-to-treat			
	Severity	Mild	Moderate	Severe	Overall	Mild	Moderate	Severe	Overall
Screening	N	39	82	48	169	62	150	87	299
	Mean (95%CI)	6.9 (5.7, 8.0)	7.5 (6.7; 8.3)	9.7 (8.6; 10.8)	8.0 (7.4; 8.6)	6.7 (5.9, 7.5)	7.5 (6.9, 8.1)	9.7 (8.9, 10.5)	8.0 (7.5, 8.4)
Check-in 1	N	39	82	48	169	59	144	81	284
	Mean (95%CI)	9.7 (8.1, 11.3)	10.6 (9.7, 11.5)	12.5 (11.3, 13.8)	10.9 (10.2, 11.6)	9.2 (8.0, 10.4)	10.3 (9.6, 10.9)	12.3 (11.3, 13.3)	10.6 (10.2, 11.2)
Check-in 2	N	39	82	48	169	54	120	66	240
	Mean (95%CI)	6.7 (5.4, 7.9)	7.5 (6.5, 8.4)	7.7 (6.5, 8.9)	7.3 (6.7, 8.0)	6.5 (5.6, 7.6)	7.3 (6.6, 8.0)	8.0 (6.9, 9.1)	7.3 (6.8, 7.9)
Check-in 3	N	39	82	48	169	46	106	57	209
	Mean (95%CI)	5.7 (4.7, 6.7)	6.5 (5.5, 7.4)	6.9 (5.7, 8.2)	6.4 (5.8, 7.0)	5.6 (4.6, 6.5)	6.4 (5.6, 7.2)	6.9 (5.8, 8.1)	6.4 (5.8, 6.9)
Check-in 4	N	39	82	48	169	40	90	50	180
	Mean (95%CI)	5.1 (4.0, 6.2)	5.9 (5.1, 6.7)	5.7 (4.6, 6.8)	5.7 (5.1, 6.2)	5.4 (4.1, 6.7)	6.0 (5.2, 6.7)	5.9 (4.8, 7.0)	5.8 (5.3, 6.4)
Check-in 5	N	31	68	36	135	31	70	37	138
	Mean (95%CI)	5.8 (4.3, 7.2)	5.5 (4.6, 6.4)	6.1 (4.8, 7.5)	5.7 (5.1, 6.4)	5.8 (4.3, 7.2)	5.4 (4.5, 6.3)	6.5 (5.0, 8.1)	5.8 (5.1, 6.5)
Check-in 6	N	23	61	28	112	23	62	28	113
	Mean (95%CI)	4.2 (2.9, 5.5)	4.9 (3.9, 5.9)	5.4 (3.7, 7.1)	4.9 (4.2, 5.6)	4.2 (2.9, 5.5)	4.9 (4.0, 5.9)	5.4 (3.7, 7.1)	4.9 (4.2, 5.6)
Completion	N	39	82	48	169	46	101	56	203
	Mean (95%CI)	4.6 (3.5, 5.7)	4.8 (4.0, 5.6)	5.2 (4.0, 6.4)	4.9 (4.3, 5.4)	4.5 (3.5, 5.5)	5.1 (4.3, 5.9)	5.3 (4.2, 6.4)	5.0 (4.5, 5.6)
Follow-up	N	39	80	47	166	48	106	56	210
	Mean (95%CI)	5.2 (3.9, 6.6)	5.3 (4.2, 6.3)	5.7 (4.1, 7.2)	5.4 (4.7, 6.1)	5.1 (4.0, 6.3)	5.6 (4.7, 6.5)	5.9 (4.5, 7.3)	5.6 (4.9, 6.2)