

# Task-sharing and telemedicine delivery of psychotherapy to treat perinatal depression: a pragmatic, noninferiority randomized trial

Received: 10 September 2024

Accepted: 18 December 2024

Published online: 3 March 2025

 Check for updates

Daisy R. Singla <sup>1,2,3,4</sup>✉, Richard K. Silver<sup>5,6,24</sup>, Simone N. Vigod <sup>4,7</sup>,  
Nour Schoueri-Mychasiw<sup>3</sup>, J. Jo Kim<sup>5,23</sup>, Laura M. La Porte<sup>5,23</sup>,  
Paula Ravitz <sup>2,3,4,23</sup>, Crystal E. Schiller<sup>8,23</sup>, Andrea S. Lawson<sup>1</sup>, Alex Kiss<sup>9</sup>,  
Steven D. Hollon<sup>10</sup>, Cindy-Lee Dennis <sup>3,4,11</sup>, Tara S. Berenbaum<sup>3</sup>,  
Holly A. Krohn<sup>8</sup>, Jamie E. Gibori<sup>5</sup>, Jaime Charlebois <sup>12</sup>, David M. Clark<sup>13</sup>,  
Ariel K. Dalfen<sup>2,3,4</sup>, Wendy Davis<sup>14</sup>, Bradley N. Gaynes<sup>8,15</sup>, Molyn Leszcz<sup>2,3,4</sup>,  
Sarah Rachel Katz<sup>3</sup>, Kellie E. Murphy<sup>16,17</sup>, John A. Naslund<sup>18</sup>,  
Mae Lynn Reyes-Rodríguez<sup>8</sup>, Alison M. Stuebe <sup>19,20</sup>, Claire Zlobin<sup>21</sup>,  
Benoit H. Mulsant <sup>1,4,24</sup>, Vikram Patel<sup>18,22,24</sup> & Samantha Meltzer-Brody<sup>8,24</sup>

Task-sharing and telemedicine can increase access to effective psychotherapies. Scaling Up Maternal Mental healthcare by Increasing access to Treatment (SUMMIT) is pragmatic, multisite, noninferiority, four-arm trial that tested the non-inferiority of provider (non-specialist vs. specialist providers) and modality (telemedicine vs. in-person) in delivering psychotherapy for perinatal depressive symptoms. Across three university-affiliated networks in the United States and Canada, pregnant and postpartum adult participants were randomized 1:1:1:1 to each arm (472 nonspecialist telemedicine, 145 nonspecialist in-person, 469 specialist telemedicine and 144 specialist in-person) and offered weekly behavioral activation treatment sessions. The primary outcome was depressive symptoms (Edinburgh Postnatal Depression Scale (EPDS)) and the secondary outcome was anxiety (Generalized Anxiety Disorder (GAD-7)) symptoms at 3 months post-randomization. Between 8 January 2020 and 4 October 2023, 1,230 participants were recruited. Noninferiority was met for the primary outcome comparing provider (EPDS: nonspecialist 9.27 (95% CI 8.85–9.70) versus specialist 8.91 (95% CI 8.49–9.33)) and modality (EPDS: telemedicine 9.15 (95% CI 8.79–9.50) versus in-person 8.92 (95% CI 8.39–9.45)) for both intention-to-treat and per protocol analyses. Noninferiority was also met for anxiety symptoms in both comparisons. There were no serious or adverse events related to the trial. This trial suggests compelling evidence for task-sharing and telemedicine to improve access to psychotherapies for perinatal depressive and anxiety symptoms. ClinicalTrials.gov NCT04153864

A full list of affiliations appears at the end of the paper. ✉ e-mail: [daisy.singla@utoronto.ca](mailto:daisy.singla@utoronto.ca)

One in five women experience depression or anxiety during the perinatal period (pregnancy up to the year following childbirth)<sup>1,2</sup>. Treatment is essential given the negative, long-term and intergenerational impact on maternal and child developmental outcomes<sup>3</sup>. Brief psychotherapies are first-line, evidence-based treatments<sup>4,5</sup>. They are preferred by perinatal populations over pharmacotherapy<sup>6</sup> and recommended by major clinical guidelines such as the UK National Institute for Health and Clinical Excellence<sup>7</sup>, the United States Preventive Task Force<sup>8</sup> and the Canadian Network for Mood and Anxiety Disorders Treatments<sup>9,10</sup>. However, access is limited, with barriers including costs, stigma and the inequitable distribution of mental health professionals<sup>11</sup>. As a result, only 10% of affected perinatal patients in high-income countries (HICs) receive psychotherapy<sup>12</sup>.

Both task-sharing and telemedicine are scalable, patient-centered solutions to improve access to psychotherapy. Scalability is defined as the potential of an intervention to reach large numbers of individuals<sup>13</sup>. In task-sharing, nonspecialist treatment providers—individuals without a specialized degree or prior experience delivering mental healthcare—are trained to deliver brief, manualized psychotherapies and have been shown to treat a range of mental health conditions worldwide<sup>14,15</sup>. A previous systematic review<sup>16</sup> yielded 45 randomized controlled trials of nonspecialist-delivered psychological treatments for perinatal populations with common mental health conditions. The results suggested that nonspecialists—namely nurses and midwives—could be trained to deliver psychological treatments for perinatal populations with depressive and anxiety symptoms in HICs. However, in both low- and middle-income and HICs, most trials use an inactive control group (for example, a waitlist) and no trials, to our knowledge, evaluated whether different provider types were able to deliver the same treatments comparably. Thus, while the efficacy of this approach has been shown to reduce perinatal depressive and anxiety symptoms worldwide<sup>14,16</sup>, the relative effectiveness of nonspecialists compared with specialist providers remains unknown.

Using telemedicine, psychotherapies can offer patients flexibility by reducing barriers such as scheduling, transportation and childcare<sup>11,17</sup>. Telemedicine-delivered psychotherapies are generally preferred by perinatal patients and are efficacious in reducing perinatal depressive and anxiety symptoms<sup>18,19</sup>. While the use of telemedicine rose dramatically during the COVID-19 pandemic, meaningful comparisons of telemedicine-delivered psychotherapy with in-person delivery have been impeded by inadequately powered trials to assess their effectiveness in common mental disorders. Thus, it is unclear whether telemedicine-delivered psychotherapy is a reasonable alternative to in-person psychotherapy, particularly for perinatal populations. Further, to our knowledge, no one trial has compared both task-sharing and telemedicine with specialists and in-person care in real-world healthcare settings. The combination of these two effective and patient-centered solutions has the potential to address the substantial and growing treatment gap for perinatal populations and beyond.

In the Scaling Up Maternal Mental healthcare by Increasing access to Treatment (SUMMIT) trial, we compared the effectiveness of provider (nonspecialist versus specialist) and modality (telemedicine versus in-person) in the delivery of one psychotherapy known as behavioral activation (BA) therapy. BA is a relatively brief<sup>20</sup>, evidence-based treatment that has been implemented effectively by a range of nonspecialists globally<sup>21–23</sup>. BA was selected because of its well-established efficacy for depressive and anxiety symptoms in both general and perinatal populations<sup>24,25</sup>, and its comparable efficacy to both antidepressant medication<sup>26,27</sup> and longer courses of psychotherapy<sup>27,28</sup>. The SUMMIT open-access treatment manual is available online at no cost ([www.thesummittrial.com](http://www.thesummittrial.com)). The current study focused on objectives that assessed outcomes at 3 months post-randomization.

The primary objectives examined (1) whether BA delivered by nonspecialists was noninferior to BA delivered by specialists in treating perinatal depressive symptoms and (2) whether BA delivered via

telemedicine was noninferior to in-person BA in treating perinatal depressive symptoms. The secondary objectives assessed (1) whether BA delivered by nonspecialists was noninferior to BA delivered by specialists in treating perinatal anxiety symptoms and (2) the moderating effects of symptom severity (mild, moderate or severe) on the comparative effectiveness of providers and modality types on depressive and anxiety symptoms. The exploratory analyses examined (1) the noninferiority of key process variables and (2) whether participants recruited during different trial phases could be combined given recruitment occurred throughout the COVID-19 pandemic.

## Results

### Enrollment and participant characteristics

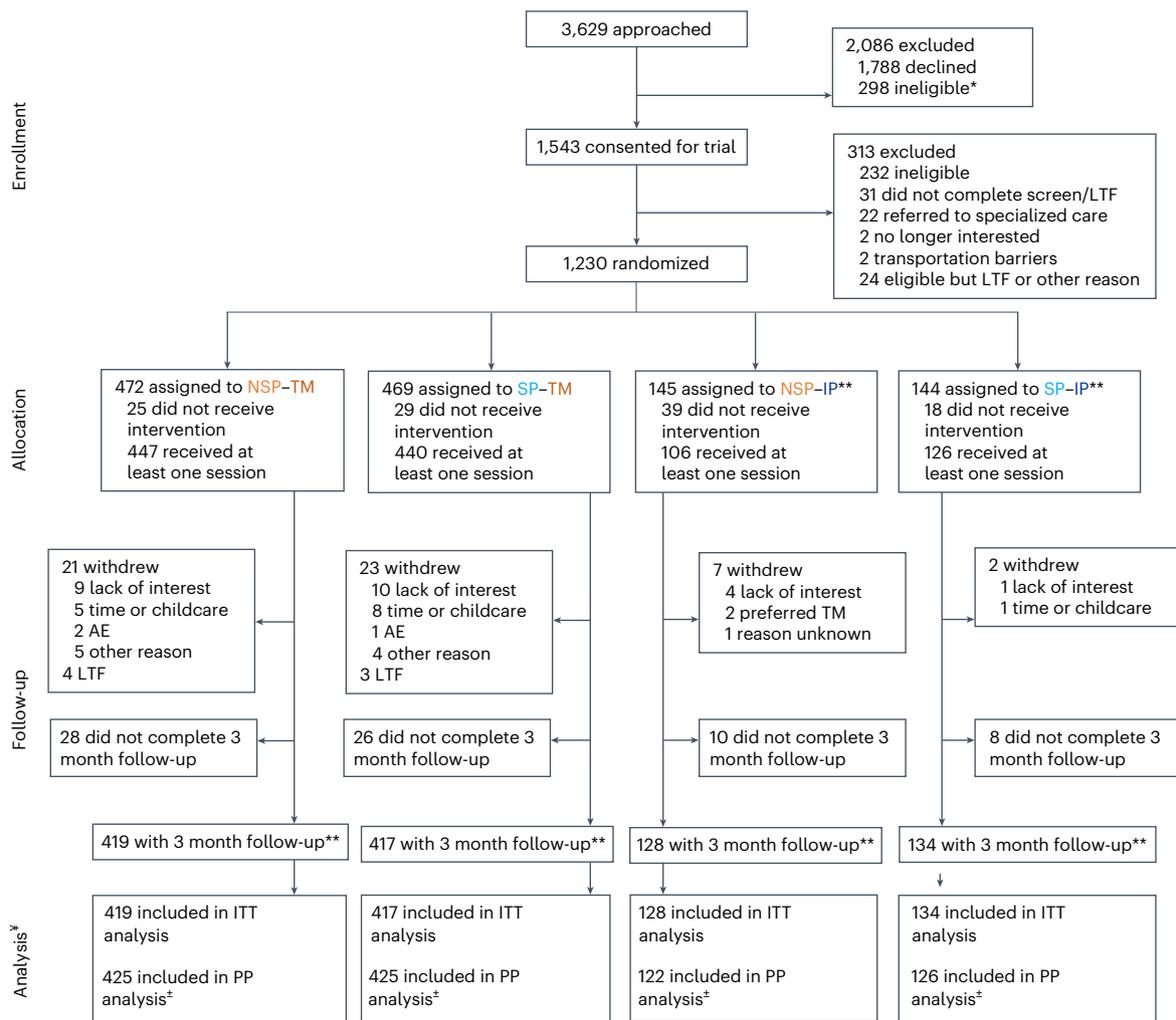
From 8 January 2020 to 4 October 2023, 3,629 individuals were approached from five clinical sites (Mount Sinai Hospital, Women's College Hospital, St. Michael's Hospital, University of North Carolina and Endeavor Health); 1,543 agreed to participate and 1,512 completed a second, more detailed screening for which consent was required.  $N = 1,230$  participants were enrolled and randomized into the trial (Fig. 1 and Extended Data Figs. 1 and 2): 472 were assigned to the nonspecialist-telemedicine arm, 469 to specialist-telemedicine, 145 to nonspecialist in-person and 144 to specialist in-person. The 3 month follow-up occurred from March 2020 to February 2024. All analyses in this paper involved outcomes assessed at 3 months post-randomization.

Baseline demographic and clinical characteristics are described in Table 1. Participants' mean age was 33.27 (95% confidence interval (CI) 33.00–33.55) years, who predominantly identified as cis-women (1,168/1,173, 99.57%); Black, Indigenous or Persons of Color (578/1,226, 47.15%); and nulliparous (668/1,226, 54.49%). Most participants (1,051/1,226, 85.73%) reported a history of depression or anxiety, and almost one-quarter (288/1,226, 23.49%) were taking psychotropic medications at enrollment. Baseline characteristics were balanced between the four arms; however, most participants expressed an initial preference for being treated by a specialist ( $n = 730$ , 60.68%) and via telemedicine ( $n = 747$ , 62.09%).

Participants attended a mean of 6.20 (95% CI 6.05–6.36) BA sessions, with the vast majority (982 of 1,230 or 80% across sites) reaching treatment completion status, defined as completing a minimum of four sessions over a 120 day period. While there were no statistically significant differences in treatment dosage (number of sessions attended) between provider conditions (nonspecialist 6.06 versus specialist 6.36), those randomized to telemedicine attended significantly more sessions than those randomized to in-person BA (6.55 versus 5.07,  $t_{(1,228)} = -8.15$ ,  $P < 0.001$ ). Further, BA fidelity scores were statistically higher for nonspecialists than specialists, with no differences between modalities (Supplementary Table 1). Neither treatment fidelity nor treatment modality preference were statistically correlated with patient depressive and or anxiety symptoms at 3 months post-randomization (Supplementary Table 2).

### Primary outcomes

**Providers: nonspecialists versus specialists.** The primary outcome was the Edinburgh Postnatal Depression Scale<sup>29</sup> (EPDS) at 3 months post-randomization. In the intention-to-treat (ITT) analyses comparing providers ((EPDS: nonspecialist 9.27 (95% CI 8.85–9.70) versus specialist 8.91 (95% CI 8.49–9.33), absolute difference in EPDS means (0.36)), the upper limit of the 95% CI for the difference in EPDS means (0.86) did not exceed the 10% noninferiority margin (EPDS 0.89). Thus, noninferiority of nonspecialist- to specialist-delivered psychotherapy was met. A per protocol (PP) analysis for the comparison of nonspecialist versus specialist was not carried out as the PP population remained the same as ITT due to the study participants not switching from one type of provider to another. Figure 2a illustrates the nonadjusted mean EPDS scores for both provider conditions over time.



**Fig. 1 | Enrollment, randomization and follow-up of the participants by arm ( $N = 1,230$ ).** Participants ( $N = 1,230$ ) were randomly assigned to one of four arms, and all were offered BA. LTF, lost to follow-up; NSP, nonspecialist provider; SP, specialist provider; TM, telemedicine; IP, in-person. \*Includes one participant who was a screen fail. \*\*Due to the COVID-19 pandemic, recruitment to in-person (NSP-IP and SP-IP) occurred during the following time periods: January 2020 to March 2020, April 2022 to April 2023 and July 2021 to January 2022

(Extended Data Fig. 1). \*\*\*Reflects participants who completed the primary outcome (EPDS). <sup>z</sup>Per protocol analyses were only conducted for modality (TM versus IP) comparisons because protocol deviations were defined as instances where participants were switched from IP to TM due to the COVID-19 pandemic. <sup>y</sup>Imputed analyses including the full sample ( $N = 1,230$ ) were also conducted for all ITT and PP analyses (see Extended Data Fig. 2 for a flow chart by condition).

**Modality: telemedicine versus in-person.** In the ITT analyses comparing treatment modality (EPDS: telemedicine 9.15 (95% CI 8.79–9.50) versus in-person 8.92 (95% CI 8.38–9.45), absolute difference in EPDS means (0.23)), the upper limit of the 95% CI for the difference in EPDS means (0.77) did not exceed the 13% noninferiority margin (EPDS 1.16) at 3 months post-randomization. Due to institutional pandemic-related restrictions, 21 participants were switched from in-person to telemedicine and reconsidered for the PP analyses. Similar results were found for the PP analyses comparing treatment modality (EPDS: telemedicine 9.17 (95% CI 8.82–9.53) versus in-person 8.81 (95% CI 8.26–9.36), absolute difference in EPDS means (0.36)) the upper limit of the 95% CI for the difference in EPDS means (0.91) was below the noninferiority margin of EPDS (1.15; Table 2). Thus, noninferiority of telemedicine to in-person psychotherapy was met. Figure 2b illustrates the nonadjusted mean EPDS scores for both modalities over time.

In both comparisons, noninferiority was also met when analyses were conducted with imputed data, adjusted (regression) models (Table 2) and without outliers (Supplementary Table 6). Further, we tested for a modality by provider interaction and found it to be non-significant (EPDS:  $P = 0.93$ ; Fig. 2), suggesting that results comparing

provider and modality held when considering modality and provider type respectively. These results also held across sites and when considering other important covariates such as clinical baseline severity, race and ethnicity, site, treatment preference for provider or modality, and psychotropic medication use (see Supplementary Table 7 for the full list of covariates considered in the adjusted analyses).

### Secondary outcomes

**Perinatal anxiety symptoms.** Figure 3 presents nonadjusted mean Generalized Anxiety Disorder<sup>30</sup> (GAD-7) scores over time for providers and modalities. Noninferiority was met when comparing providers (ITT GAD-7: nonspecialist 6.44 (95% CI 6.01–6.86) versus specialist 6.36 (95% CI 5.95–6.78)) and modalities (ITT GAD-7: telemedicine 6.43 (95% CI 6.09–6.78) versus in-person 6.29 (95% CI 5.71–6.88)) on anxiety symptoms for all analyses at 3 months post-randomization (Table 2) except for the PP analyses comparing telemedicine versus in-person in which noninferiority was not met on anxiety symptoms unless outliers were removed (Supplementary Table 6). The small absolute differences between modalities for ITT (0.14) and PP analyses (0.30) are not clinically meaningful, as suggested by others (that is,

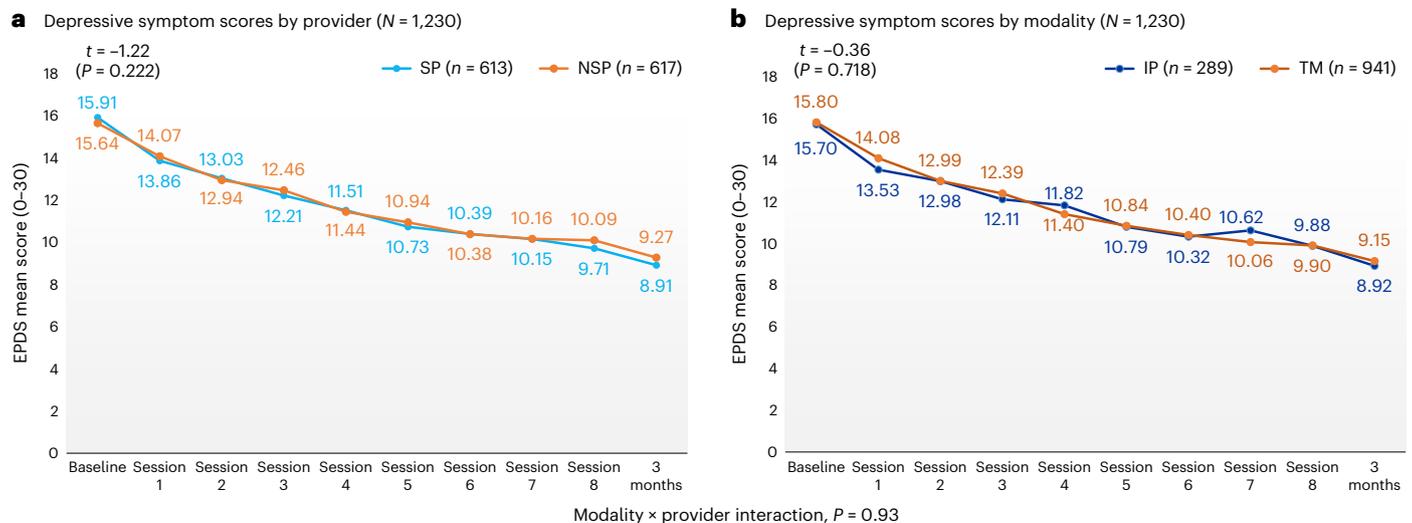
**Table 1 | Baseline characteristics of participants (N=1230), n (%), unless otherwise specified**

Variable (total respondents)	Overall (N=1,230)	Provider conditions		Modality conditions	
		Nonspecialist provider (n=617)	Specialist provider (n=613)	Telemedicine (n=941)	In-person (n=289)
<b>Age in years (1,226)<sup>a</sup></b> , mean (95% CI)	33.27 (33.00–33.55)	33.26 (32.86–33.66)	33.29 (32.91–33.68)	33.06 (32.74–33.38)	33.97 (33.40–34.53)
<b>Perinatal period at enrollment (1,230)<sup>a</sup></b>					
Pregnant	618 (50.24)	305 (49.43)	313 (51.06)	477 (50.69)	141 (48.79)
Postpartum	612 (49.76)	312 (50.57)	300 (48.94)	464 (49.31)	148 (51.21)
<b>Gestational age (1,230)</b>					
Weeks pregnant	23.72 (23.17–24.27)	23.79 (23.01–24.57)	23.65 (22.88–24.43)	23.95 (23.33–24.57)	22.94 (21.74–24.15)
Weeks postpartum	12.77 (12.20–13.34)	12.81 (12.01–13.61)	12.73 (11.91–13.55)	13.12 (12.46–13.78)	11.68 (10.54–12.82)
<b>Race/ethnicity (1,226)<sup>a</sup></b>					
American Indian/Alaska Native	5 (0.41)	1 (0.16)	4 (0.66)	4 (0.43)	1 (0.35)
Asian	209 (17.05)	114 (18.51)	95 (15.57)	160 (17.08)	49 (16.95)
Black/African American	124 (10.11)	59 (9.58)	65 (10.66)	86 (9.18)	38 (13.15)
Hawaiian/Pacific Islander	4 (0.33)	2 (0.33)	2 (0.33)	2 (0.21)	2 (0.69)
White	614 (50.08)	311 (50.49)	303 (49.67)	473 (50.48)	141 (48.79)
Multi-race	100 (8.16)	44 (7.14)	56 (9.18)	79 (8.43)	21 (7.27)
Hispanic (Latino/Latina)	101 (8.24)	48 (7.79)	53 (8.69)	77 (8.22)	24 (8.30)
Other	35 (2.85)	19 (3.08)	16 (2.62)	29 (3.09)	6 (2.08)
Prefer not to answer	34 (2.77)	18 (2.92)	16 (2.62)	27 (2.88)	7 (2.42)
<b>Born in country of current residence (1,226)<sup>a</sup></b>	860 (70.15)	433 (70.29)	427 (70.00)	653 (69.69)	207 (71.63)
<b>Private health benefits (948)<sup>b, d</sup></b>	654 (68.99)	327 (69.13)	327 (68.84)	466 (69.45)	188 (67.87)
<b>Marital status (1,226)<sup>a</sup></b>					
Married or stable relationship	1052 (85.81)	528 (85.71)	524 (85.90)	806 (86.02)	246 (85.12)
Single	94 (7.67)	47 (7.63)	47 (7.70)	71 (7.58)	23 (7.96)
Dating	41 (3.34)	19 (3.08)	22 (3.61)	31 (3.31)	10 (3.46)
Other	21 (1.71)	11 (1.79)	10 (1.64)	14 (1.49)	7 (2.42)
Prefer not to answer	18 (1.47)	11 (1.79)	7 (1.15)	15 (1.60)	3 (1.04)
<b>Gender identity (1,173)<sup>b</sup></b>					
Cis-woman	1,168 (99.57)	585 (99.32)	583 (99.83)	881 (99.44)	287 (100)
<b>Sexual orientation (1,173)<sup>b</sup></b>					
Straight/heterosexual	1,073 (91.47)	531 (90.15)	542 (92.81)	810 (91.42)	263 (91.64)
Bisexual	71 (6.05)	42 (7.13)	29 (4.97)	52 (5.87)	19 (6.62)
Other	7 (0.60)	3 (0.51)	4 (0.68)	6 (0.68)	1 (0.35)
Prefer not to answer	22 (1.88)	13 (2.21)	9 (1.54)	18 (2.03)	4 (1.39)
<b>Highest level of education completed (1,226)<sup>a</sup></b>					
University degree	859 (70.07)	432 (70.13)	427 (70.00)	655 (69.90)	204 (70.59)
College/trade school	216 (17.62)	104 (16.88)	112 (18.36)	161 (17.18)	55 (19.03)
High school or less	141 (11.50)	76 (12.34)	65 (10.66)	112 (11.95)	29 (10.03)
Prefer not to answer	10 (0.81)	4 (0.65)	6 (0.98)	9 (0.96)	1 (0.35)
<b>Employment (1,226)<sup>a</sup></b>					
Full time	484 (39.48)	238 (38.64)	246 (40.33)	366 (39.06)	118 (40.83)
Maternity leave	353 (28.79)	182 (29.55)	171 (28.03)	268 (28.60)	85 (29.41)
Part time	111 (9.05)	53 (8.60)	58 (9.51)	83 (8.86)	28 (9.69)
Not employed	260 (21.21)	133 (21.59)	127 (20.82)	207 (22.09)	53 (18.34)
Other	18 (1.47)	10 (1.62)	8 (1.31)	13 (1.39)	5 (1.73)
<b>Annual income (US\$) based on zip code (1,138)</b> , mean (95% CI)	43,165.59 (42,236.47–44,094.70)	43,318.76 (41,946.93–44,690.58)	43,016.14 (41,756.00–44,276.28)	43,493.40 (42,390.31–44,596.48)	42,141.77 (40,460.83–43,822.72)
<b>Nulliparous (1,226)<sup>a</sup></b>	668 (54.49)	333 (54.06)	335 (54.92)	524 (55.92)	144 (49.83)
<b>COVID-19 exposure (1,064)<sup>b</sup></b>	31 (2.91)	16 (2.98)	15 (2.85)	22 (2.80)	9 (3.25)

**Table 1 (continued) | Baseline characteristics of participants (N=1230), n (%), unless otherwise specified**

Variable (total respondents)	Overall (N=1,230)	Provider conditions		Modality conditions	
		Nonspecialist provider (n=617)	Specialist provider (n=613)	Telemedicine (n=941)	In-person (n=289)
<b>Self-reported history of depression/ anxiety (1,226)<sup>a</sup></b>	1,051 (85.73)	528 (85.71)	523 (85.74)	805 (85.91)	246 (85.12)
<b>Psychotropic medication use at baseline (1,226)<sup>c</sup></b>	288 (23.49)	138 (22.40)	150 (24.59)	227 (24.12)	61 (21.40)
<b>Pregnancy conditions during current pregnancy (≥1) (1,075)<sup>b</sup></b>	265 (24.65)	125 (23.06)	140 (26.27)	192 (24.06)	73 (26.35)
Preeclampsia	52 (4.84)	27 (4.98)	25 (4.69)	36 (4.51)	16 (5.78)
Gestational hypertension	127 (11.81)	66 (12.18)	61 (11.44)	82 (10.28)	45 (16.25)
Gestational diabetes	121 (11.26)	56 (10.33)	65 (12.20)	85 (10.65)	36 (13.00)
Preterm labor	69 (6.42)	30 (5.54)	39 (7.32)	51 (6.39)	18 (6.50)
<b>Symptom scores at baseline</b>					
Depressive symptoms (EPDS) Total (0–30), mean (95% CI)	15.77 (15.56–15.99)	15.64 (15.34–15.94)	15.91 (15.60–16.22)	15.80 (15.55–16.04)	15.70 (15.25–16.15)
Anxiety symptoms (GAD-7) Total (0–21) mean (95% CI)	11.85 (11.58–12.12)	11.94 (11.57–12.31)	11.75 (11.35–12.16)	11.82 (11.51–12.14)	11.93 (11.38–12.49)
<b>Modality preference at baseline (1,203)<sup>b</sup></b>					
Telemedicine	747 (62.09)	–	–	–	–
In-person	190 (15.79)	–	–	–	–
No preference	266 (22.11)	–	–	–	–
<b>Provider preference at baseline (1,203)<sup>b</sup></b>					
Specialist provider	730 (60.68)	–	–	–	–
Nonspecialist provider	22 (1.83)	–	–	–	–
No preference	451 (37.49)	–	–	–	–

<sup>a</sup>Four participants did not complete the baseline assessment. <sup>b</sup>This question was added after the trial commencement (in Spring 2020). <sup>c</sup>This question is a subquestion of another question. <sup>d</sup>All Canadian participants have government insurance and some also have supplemental private insurance. See Supplementary Tables 3 and 4 for baseline characteristics of participants by hub and arm, respectively, and Supplementary Table 5 (refs. 44–56) for the representativeness of trial participants.



**Fig. 2 | Comparing depressive (EPDS) symptom scores by provider and modality. a, b.** The primary outcome was depressive symptom scores, as assessed by the EPDS at 3 months post-randomization, by provider (a) and modality (b). The *t* statistics and *P* values correspond to two-sample *t*-tests

conducted to evaluate the difference in depressive symptom scores at baseline that presented no significant differences between conditions. We tested for a modality by provider interaction (*P* = 0.93) for our primary (EPDS) outcome and we found it to be nonsignificant in all cases.

GAD-7 1.5–4)<sup>31</sup>, and the modality by provider interaction was nonsignificant (GAD-7: *P* = 0.71; Fig. 3).

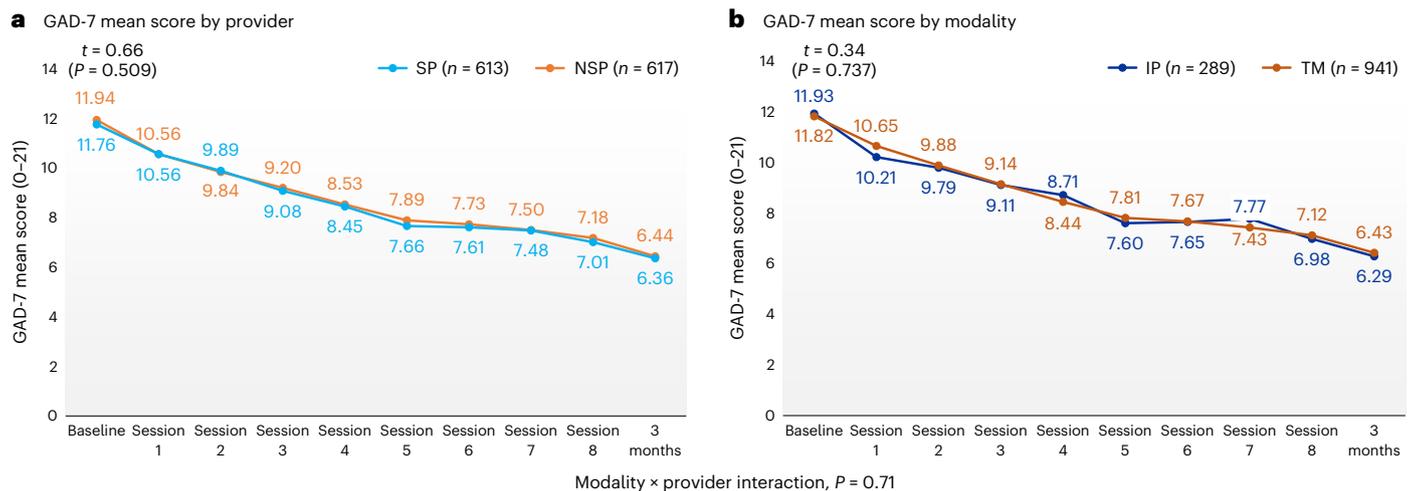
**Clinical severity.** We found no statistically significant differences for symptoms of depression (Fig. 4) or anxiety (Extended Data Fig. 3) at

3 months post-randomization when comparing providers (EPDS: mild: *F* = 0.09, *P* = 0.7619; moderate: *F* = 2.74, *P* = 0.0983; severe: *F* = 0.66, *P* = 0.4175) and modalities (EPDS: mild: *F* = 0.53, *P* = 0.4657; moderate: *F* = 0.26, *P* = 0.6076; severe: *F* = 2.59, *P* = 0.1090), accounting for baseline clinical severity. Scores decreased significantly between baseline

**Table 2 | Postintervention means (95% CI) and noninferiority results for primary, secondary and exploratory outcomes at 3 months post-randomization**

Outcome (measure; range)	Nonadjusted mean <sup>y</sup> (95% CI)		95% Clupper bound <sup>z</sup>	NIM <sup>*</sup>	Adjusted mean <sup>y</sup> (95% CI)		95% Clupper bound <sup>z</sup>	NIM <sup>*</sup>
<b>Primary and secondary outcomes</b>								
<b>ITT: nonspecialist versus specialist (10% NIM)</b>	<b>Nonspecialist provider</b>	<b>Specialist provider</b>			<b>Nonspecialist provider</b>	<b>Specialist provider</b>		
Depressive symptoms (EPDS, 0–30)	9.27 (8.85–9.70)	8.91 (8.49–9.33)	0.86	0.89*	8.66 (7.85–9.46)	8.33 (7.51–9.14)	0.77	0.89*
Anxiety symptoms (GAD-7, 0–21)	6.44 (6.01–6.86)	6.36 (5.95–6.78)	0.57	0.64*	6.04 (5.23–6.84)	6.17 (5.36–6.98)	0.31	0.64*
<b>ITT: telemedicine versus in-person (13% NIM)</b>	<b>Telemedicine</b>	<b>In-person</b>			<b>Telemedicine</b>	<b>In-person</b>		
Depressive symptoms (EPDS, 0–30)	9.15 (8.79–9.50)	8.92 (8.38–9.45)	0.77	1.16*	8.61 (7.83–9.39)	8.37 (7.50–9.23)	0.78	1.16*
Anxiety symptoms (GAD-7, 0–21)	6.43 (6.09–6.78)	6.29 (5.71–6.88)	0.73	0.82*	6.14 (5.34–6.93)	6.02 (5.15–6.90)	0.65	0.82*
<b>PP: telemedicine versus in-person (13% NIM)</b>	<b>Telemedicine</b>	<b>In-person</b>			<b>Telemedicine</b>	<b>In-person</b>		
Depressive symptoms (EPDS, 0–30)	9.17 (8.82–9.53)	8.81 (8.26–9.36)	0.91	1.15*	8.56 (7.77–9.34)	8.22 (7.34–9.10)	0.89	1.15*
Anxiety symptoms (GAD-7, 0–21)	6.47 (6.12–6.81)	6.17 (5.58–6.76)	0.90	0.80	6.20 (5.42–6.98)	5.86 (4.98–6.74)	0.89	0.80
<b>Exploratory outcomes</b>								
<b>ITT: nonspecialist versus specialist (10% NIM)</b>	<b>Nonspecialist provider</b>	<b>Specialist provider</b>			<b>Nonspecialist provider</b>	<b>Specialist provider</b>		
Quality of life (EQ-5D-5L domains, 1–5)								
Mobility	1.35 (1.29–1.41)	1.33 (1.27–1.38)	0.09	0.13*	1.34 (1.22–1.45)	1.31 (1.20–1.43)	0.08	0.13*
Self-care	1.24 (1.19–1.28)	1.25 (1.20–1.29)	0.04	0.13*	1.19 (1.10–1.28)	1.22 (1.12–1.31)	0.02	0.13*
Usual activities	1.74 (1.67–1.81)	1.71 (1.63–1.77)	0.12	0.17*	1.69 (1.55–1.84)	1.64 (1.50–1.78)	0.13	0.17*
Pain/discomfort	1.91 (1.84–1.99)	1.87 (1.80–1.94)	0.13	0.19*	1.87 (1.73–2.02)	1.83 (1.68–1.97)	0.12	0.19*
Anxiety/depression	2.27 (2.19–2.34)	2.26 (2.18–2.33)	0.09	0.23*	2.06 (1.92–2.21)	2.08 (1.94–2.23)	0.06	0.23*
Disability (WHODAS 2.0, 12–60)	20.93 (20.24–21.63)	20.79 (20.14–21.45)	0.94	2.08*	20.62 (19.47–21.77)	20.47 (19.31–21.63)	0.78	2.08*
<b>ITT: telemedicine versus in-person (13% NIM)</b>	<b>Telemedicine</b>	<b>In-person</b>			<b>Telemedicine</b>	<b>In-person</b>		
Quality of life (EQ-5D-5L domains, 1–5)								
Mobility	1.34 (1.29–1.39)	1.33 (1.26–1.41)	0.08	0.17*	1.31 (1.20–1.42)	1.34 (1.22–1.46)	0.04	0.17*
Self-care	1.24 (1.20–1.28)	1.24 (1.18–1.31)	0.06	0.16*	1.20 (1.11–1.29)	1.21 (1.11–1.31)	0.05	0.16*
Usual activities	1.71 (1.65–1.77)	1.76 (1.66–1.86)	0.05	0.23*	1.63 (1.49–1.76)	1.71 (1.55–1.86)	0.01	0.23*
Pain/discomfort	1.89 (1.83–1.95)	1.91 (1.81–2.01)	0.08	0.25*	1.83 (1.70–1.97)	1.87 (1.71–2.02)	0.06	0.25*
Anxiety/depression	2.27 (2.21–2.33)	2.23 (2.13–2.34)	0.15	0.29*	2.09 (1.95–2.23)	2.05 (1.90–2.21)	0.13	0.29*
Disability (WHODAS 2.0, 12–60)	20.77 (20.22–21.32)	21.16 (20.21–22.12)	–0.39	2.75*	19.94 (18.71–21.17)	20.41 (19.05–21.78)	0.36	2.75*
Outcome (measure; range)	Nonadjusted mean <sup>y</sup> (95% CI)		95% CI lower bound <sup>z</sup>	NIM <sup>*</sup>	Adjusted mean <sup>y</sup> (95% CI)		95% CI lower bound <sup>z</sup>	NIM <sup>*</sup>
<b>ITT: nonspecialist versus specialist (10% NIM)</b>	<b>Nonspecialist provider</b>	<b>Specialist provider</b>			<b>Nonspecialist provider</b>	<b>Specialist provider</b>		
Client satisfaction (CSQ-8, 1–4)	3.40 (3.34–3.46)	3.44 (3.39–3.49)	–0.10	–0.34*	3.43 (3.32–3.55)	3.51 (3.40–3.62)	–0.14	–0.34*
Therapeutic alliance (WAI-SR, 1–5)	4.20 (4.13–4.27)	4.12 (4.05–4.19)	–0.01	–0.41*	4.08 (3.93–4.23)	4.07 (3.92–4.22)	–0.07	–0.41*
Patient activation (PAAS, 0–4)	2.17 (2.09–2.24)	2.13 (2.06–2.21)	–0.06	–0.21*	2.11 (1.96–2.26)	2.07 (1.92–2.22)	–0.04	–0.21*
Perceived support (MSPSS, 1–7)	5.53 (5.43–5.62)	5.46 (5.35–5.57)	–0.05	–0.55*	5.50 (5.32–5.67)	5.43 (5.26–5.61)	–0.03	–0.55*
<b>ITT: telemedicine versus in-person (13% NIM)</b>	<b>Telemedicine</b>	<b>In-person</b>			<b>Telemedicine</b>	<b>In-person</b>		
Client satisfaction (CSQ-8, 1–4)	3.43 (3.39–3.47)	3.38 (3.29–3.46)	–0.13	–0.44*	3.53 (3.42–3.64)	3.41 (3.30–3.53)	0.04	–0.44*
Therapeutic alliance (WAI-SR, 1–5)	4.17 (4.11–4.22)	4.13 (4.02–4.24)	–0.14	–0.54*	4.13 (3.98–4.28)	4.02 (3.86–4.18)	0.01	–0.54*
Patient activation (PAAS, 0–4)	2.15 (2.09–2.21)	2.17 (2.06–2.27)	–0.08	–0.28*	2.07 (1.93–2.21)	2.11 (1.95–2.27)	–0.14	–0.28*
Perceived support (MSPSS, 1–7)	5.48 (5.40–5.56)	5.54 (5.39–5.69)	–0.08	–0.72*	5.45 (5.28–5.62)	5.48 (5.29–5.67)	–0.15	–0.72*

\*Noninferiority was met. <sup>y</sup>Per the SAP, nonadjusted means represent nonimputed analyses without covariates. Outliers for depressive and anxiety symptoms can be found in Supplementary Table 8. Imputed analyses can be found in Supplementary Table 9. Regression models include imputed data and covariates. Covariates: perinatal period, hub, age, baseline EPDS and GAD-7, race/ethnicity, employment, modality and provider preference, history of seeing a therapist, history of depression/anxiety, years lived in current home, nulliparous, time between randomization and 3 months, psychotropic medication before 3 months, started psychotropic medication after enrollment, pregnancy conditions, modality and provider (Supplementary Table 7). CSQ-8, Client Satisfaction Questionnaire<sup>60</sup>; EQ-5D-5L, EQ-5D 5-level<sup>61</sup>; MSPSS, Multidimensional Scale of Perceived Social Support<sup>62</sup>; NIM, noninferiority margin; NSP, nonspecialist provider; PAAS, Premium Abbreviated Activation Scale<sup>63</sup>; SP, specialist provider; WAI-SR, Working Alliance Inventory–Short Revision<sup>64</sup>; WHODAS, World Health Organization Disability Assessment Schedule 2.0 (ref. 65). <sup>z</sup>Per protocol analyses refer to the condition (telemedicine versus in-person) that the participant actually received and only apply to modality because no participants were switched between provider conditions. Of the 21 protocol deviations, 16 received telemedicine and 5 received in-person and were analyzed in the PP analyses accordingly. <sup>†</sup>The 95% upper bound is used when higher scores are worse and the 95% lower bound is used when lower scores are worse.



**Fig. 3 | Comparing anxiety (GAD-7) symptom scores by provider and modality.** **a,b**, The secondary outcome was anxiety symptom scores, as assessed by the GAD-7 at 3 months post-randomization by provider (**a**) and modality (**b**). The  $t$  statistics and  $P$  values correspond to two-sample  $t$ -tests conducted to

evaluate differences in anxiety symptom scores at baseline that presented no significant differences between conditions. We tested for a modality by provider interaction ( $P = 0.71$ ) for our secondary (GAD-7) outcome and we found it to be nonsignificant in all cases.

and posttreatment in all three symptom severity groups (Supplementary Tables 10 and 11).

### Safety

Eighteen serious adverse events (SAEs) (fetal or infant death, 6; hospitalization, 5; hospitalization and life-threatening event, 3; other serious important medical event, 2; and life-threatening event, 2) and two adverse events (AEs) were identified before or at 3 months post-randomization. All were reviewed by an independent Data Safety and Monitoring Board (DSMB) and none were deemed directly related or a result of the trial. We found no statistical differences in safety events between provider or modality conditions (Supplementary Table 12).

### Exploratory outcomes

Noninferiority between both provider and modality comparisons was met for all exploratory outcomes including client satisfaction (nonspecialist 3.40 (95% CI 3.34–3.46) versus specialist 3.44 (95% CI 3.39–3.49); telemedicine 3.43 (95% CI 3.39–3.47) versus in-person 3.38 (95% CI 3.29–3.46)), therapeutic alliance (nonspecialist 4.20 (95% CI 4.13–4.27) versus specialist 4.12 (95% CI 4.05–4.19); telemedicine 4.17 (95% CI 4.11–4.22) versus in-person 4.13 (95% CI 4.02–4.24)), perceived social support (nonspecialist 5.53 (95% CI 5.43–5.62) versus specialist 5.46 (95% CI 5.35–5.57); telemedicine 5.48 (95% CI 5.40–5.56) versus in-person 5.54 (95% CI 5.39–5.69)) and quality of life at 3 months post-randomization (Table 2). In addition, the interaction terms between providers (nonspecialist versus specialist) and study phase (Methods) in relation to EPDS and GAD-7 scores were not significant, supporting that participants recruited across trial phases could be combined (Supplementary Tables 13 and 14).

### Sensitivity analyses

A priori planned sensitivity analyses accounted for differences in data collection time points and outcome variables (for example, participants who started or had a change in psychotropic medication between baseline and subsequent treatment sessions). In all analyses, no significant differences were found between provider or modality condition.

### Post hoc analyses

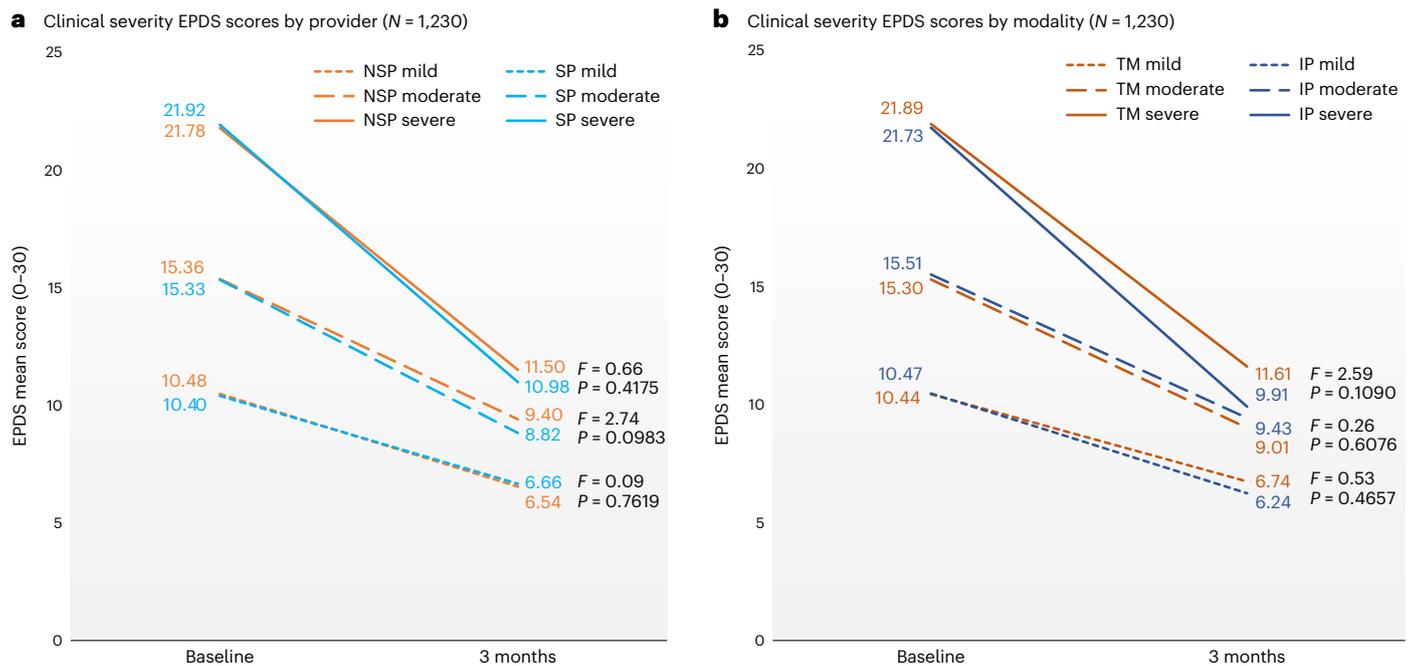
Post hoc sensitivity analysis included a comparison of participants who received no sessions to those who received one or more session, in terms of their depressive or anxiety symptom scores at 3 months

post-randomization (Extended Data Figs. 4 and 5). Those who received one or more sessions had significantly lower EPDS scores than those who received no sessions (EPDS 8.98 versus 11.03,  $P < 0.01$ ). In addition, we examined the intraclass correlation coefficient (ICC) among providers and found the values to be small in magnitude (for primary outcome EPDS, ICC 0.02 (95% CI 0.00–0.08) and for secondary outcome GAD-7, ICC 0.00 (95% CI 0.00–0.00)). This small ICC may be expected given the large number of providers ( $n = 67$ ) delivering the treatment across sites, and because the treatment was delivered individually and therefore participants were unlikely to interact with each other. Finally, we also conducted a sensitivity analysis to compare modality (telemedicine versus in-person) using the original noninferiority margin of 10% for our primary outcome (EPDS). The upper bound of the 95% CI was EPDS 0.77 and fell below the noninferiority margin of EPDS 0.89; thus, we found that noninferiority was met using the original noninferiority margin on the primary outcome of EPDS.

## Discussion

In this large, multisite trial, we found that nonspecialists (individuals without a specialized mental health background), were noninferior to specialists in delivering BA, whether BA was provided in-person or via telemedicine. Noninferiority was met across all primary, secondary and exploratory outcomes, with minor exceptions. Given the high prevalence of depression and anxiety in perinatal populations and their negative, intergenerational impact when left untreated<sup>3</sup>, these findings have important public health implications.

Congruent with growing, robust evidence supporting task-sharing worldwide<sup>13,14</sup>, our results in the SUMMIT Trial suggest that training nonspecialists can increase the mental health workforce and improve access to relatively brief<sup>20</sup>, evidence-based psychotherapies such as BA. Globally, task-sharing has been examined with the promise to scale access to a large number of services—including mental health care—because of the decreased reliance on specialist providers who are overburdened, scarce, inequitably distributed and often work in the private sector, beyond the reach of the majority of populations in most countries<sup>32</sup>. This study extends the task-sharing literature by showing the comparable effectiveness of nonspecialists to specialists delivering the same psychotherapy to a diverse perinatal population in North America, thus suggesting that nonspecialist-delivered psychotherapy is not second rate to specialist-delivered care and can enhance access to quality-assured psychotherapy. Despite the majority



**Fig. 4 | Comparing baseline clinical severity depressive (EPDS) scores by provider and modality. a,b**, Clinical severity EPDS scores were assessed at baseline and the change over time within each severity group between conditions by provider (a) and modality (b) were not significant. Severity groups based on

baseline EPDS score: mild (10–11), moderate (12–19) and severe (20–30). The *F* statistics and *P* values correspond to the time by condition interaction term from a linear mixed model.

of participants indicating a preference for treatment by a specialist provider, the SUMMIT trial demonstrated that nonspecialist-delivered care is noninferior to specialist-delivered care with high, and noninferior, patient satisfaction between provider conditions. These results suggest encouraging and reassuring evidence for patients to receive care from nonspecialists.

In addition, we found that BA was delivered by nonspecialists with statistically higher fidelity scores than by specialists. This may be due to the structured form of supervision nonspecialists received compared with the ad hoc consultation that was available to specialists following the training period. The higher fidelity scores among nonspecialists may also be because this was their first treatment delivery experience, whereas specialists were required to have a minimum of 5 years of experience delivering psychotherapies to be eligible for the trial and therefore less rigorous to follow the treatment protocol.

The SUMMIT Trial was a pragmatic trial that was embedded in the clinical care pathways of five clinical sites, and included referrals from obstetrical and from psychiatry services for patients for whom psychotherapy was indicated. Given the rapid recruitment rate, adherence to the protocols and clear integration into the existing clinical care pathways across geographically and socio-demographically diverse cities in both the United States and Canada, we believe our results demonstrate an ability for scale and spread. Across general healthcare settings, frontline healthcare workers such as nurses, midwives and doulas can be trained to deliver and increase access to psychotherapies for perinatal depressive and anxiety symptoms. The eight-session BA treatment manual was adapted from the Healthy Activity Program in India<sup>22</sup> where mental health literacy is low and reconfirms that BA is an easy and user-friendly treatment that may be suitable for different populations<sup>24</sup> and can be delivered by a range of treatment providers<sup>21,22</sup>. We found similar treatment effects to other evaluations of BA treatments<sup>21,22,25</sup> using inactive controls, which reduced moderate and severe depressive and anxiety symptoms to mild or minimal severity levels<sup>33</sup>. Finally, these data may also be reassuring for low- and middle-income countries, where task-sharing of nonspecialist care is routinely applied due to the severe

shortage of specialist mental health providers and limited financial and infrastructure resources for mental healthcare<sup>34</sup>. Thus, the SUMMIT trial builds on the growing evidence that training nonspecialists has the potential to address the global mental health treatment gap by empowering communities to build local capacity, optimize available resources and still deliver evidence-based care.

Our findings also showed that telemedicine-delivered BA is noninferior to in-person BA when treating perinatal depression and anxiety symptoms. Telemedicine reduces barriers to care, especially for underserved populations, and was preferred over in-person psychotherapy in our sample (60.28%). This preference for telemedicine aligns with other previous studies highlighting a preference for virtual care among perinatal populations<sup>19,35</sup> and women in general<sup>36,37</sup> because they assist to overcome barriers related to coordinating childcare and transportation. In addition, our findings of noninferiority in key process variables challenge concerns that the patient–provider relationship is compromised using telemedicine<sup>38</sup>. The use of telemedicine in healthcare has grown substantially since the start of the COVID-19 pandemic and can increase access to care; however, it is now at risk of being removed from covered services in many jurisdictions<sup>39</sup>. This is important because nationally representative samples have demonstrated that the most prominent reason for not using telemedicine was because providers fail to offer this option<sup>36</sup>.

Noninferiority between modalities support the continued reimbursement of telemedicine-delivered psychotherapies. We also note that telemedicine was preferred by perinatal participants and this preference was consistent throughout the study regardless of when participants were enrolled. These results indicate that telemedicine-delivered psychotherapy is noninferior to the traditional in-person model, thus offering a more patient-centered and scalable approach.

Given the pragmatic nature of our trial conducted in real-world settings, our results may inform a stepped-care approach to treating depression and anxiety beyond perinatal populations. Nonspecialist-delivered psychotherapy via telemedicine may be the ideal initial step in a collaborative stepped-care model. This is

important for three reasons. First, this approach can make brief and effective first-line psychotherapies accessible. Second, both the comparable effectiveness between providers and modalities and the marked symptom reduction in those with severe baseline symptoms highlight the benefit of this approach either on its own or while waiting for more specialized services. Third, this aligns with other task-sharing and stepped-care models (for example, ref. 40), whereby specialists ensure quality care through training and supervision and provide clinical care to patients who require more than first-step psychotherapies. To scale this task-sharing approach within stepped care, payers are required to reimburse nonspecialists delivering care and specialists providing quality assurance for the nonspecialists. Further research is required to examine whether self-guided or single-session interventions may be a suitable first-step for specific populations (for example, those with mild symptoms<sup>41</sup>).

The SUMMIT trial has many strengths. This includes a large population, a relatively high retention rate<sup>42,43</sup> despite the challenges of a global pandemic, a racially and geographically diverse sample improving generalizability (Supplementary Table 5 (refs. 44–56)) and high patient satisfaction rates. Less than 1% of participants within our large sample required Internet access, suggesting that telemedicine is unlikely to raise a new barrier to care. Furthermore, the greater adherence to the number of sessions favoring telemedicine supports its feasibility. Finally, the conservative and clinically meaningful noninferiority margins strengthen the interpretation of our findings.

The study also has limitations. Substantial modifications were required due to COVID-19, including a change in noninferiority margin from 10% to 13% in the telemedicine versus in-person comparison, and deviation from the assigned modality due to institutional mandates during the pandemic. Most of the participants (70.07%) had a university degree, thus potentially limiting the generalizability of our sample. While this may be typical of psychotherapy research participants, our findings need to be extended to other groups such as adolescents or those without higher education. The trial did not include an inactive control group. This is common and even recommended in noninferiority trials<sup>57,58</sup>. Owing to the established efficacy of BA<sup>24,25</sup>, the comparison to a nontreated group in real-world healthcare settings would be inappropriate and arguably unethical<sup>25,57</sup>. In both comparisons of modality and provider, the magnitude of symptom improvement was significantly higher among the participants who received BA than in the small number who did not (Extended Data Figs. 4 and 5) and higher than what has been reported with inactive control conditions in other BA trials<sup>25</sup>. Further, the difference between their final EPDS (2.05) falls within the range of a minimally clinically important difference of EPDS 1.4–6.4 (ref. 59). Future applications of personalized medicine algorithms could determine ‘what works for whom’ and whether specific subgroups may benefit more from one type of provider or modality compared with another.

The noninferiority results of the SUMMIT trial suggests the comparable effectiveness of task-sharing and telemedicine-delivered psychotherapies to treat perinatal depression. These population-based approaches are both effective and patient centered and have the potential to transform access to treatment for perinatal mental health globally.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41591-024-03482-w>.

## References

- O'hara, M. W. & Swain, A. M. Rates and risk of postpartum depression—a meta-analysis. *Int. Rev. Psychiatry* **8**, 37–54 (1996).
- Dennis, C.-L., Falah-Hassani, K. & Shiri, R. Prevalence of antenatal and postnatal anxiety: systematic review and meta-analysis. *Br. J. Psychiatry* **210**, 315–323 (2017).
- Howard, L. M. et al. Non-psychotic mental disorders in the perinatal period. *Lancet* **384**, 1775–1788 (2014).
- Sokol, L. E., Epperson, C. N. & Barber, J. P. A meta-analysis of treatments for perinatal depression. *Clin. Psychol. Rev.* **31**, 839–849 (2011).
- Curry, S. J. et al. Interventions to prevent perinatal depression: US preventive services task force recommendation statement. *JAMA* **321**, 580–587 (2019).
- O'Mahen, H. A. & Flynn, H. A. Preferences and perceived barriers to treatment for depression during the perinatal period. *J. Womens Health* **17**, 1301–1309 (2008).
- Depression in Adults: Treatment and Management. NICE Guidelines Vol. 2024* (National Institute of Clinical Excellence, 2022).
- O'Connor, E., Rossom, R. C., Henninger, M., Groom, H. C. & Burda, B. U. Primary care screening for and treatment of depression in pregnant and postpartum women: evidence report and systematic review for the US preventive services task force. *JAMA* **315**, 388–406 (2016).
- Parikh, S. V. et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) 2016 clinical guidelines for the management of adults with major depressive disorder: section 2. Psychological treatments. *Can. J. Psychiatry* **61**, 524–539 (2016).
- Vigod, S. N. et al. Canadian network for mood and anxiety treatments (CANMAT) 2024 clinical practice guideline for the treatment of perinatal mood, anxiety and related disorders. *Can. J. Psychiatry* (in the press).
- Webb, R. et al. Barriers and facilitators to implementing perinatal mental health care in health and social care settings: a systematic review. *Lancet Psychiatry* **8**, 521–534 (2021).
- Cox, E. Q., Sowa, N. A., Meltzer-Brody, S. E. & Gaynes, B. N. The perinatal depression treatment cascade: baby steps toward improving outcomes. *J. Clin. Psychiatry* **77**, 20901 (2016).
- Kazdin, A. E. Addressing the treatment gap: expanding the scalability and reach of treatment. *J. Consult. Clin. Psychol.* **91**, 3–5 (2023).
- Barbui, C. et al. Efficacy of psychosocial interventions for mental health outcomes in low-income and middle-income countries: an umbrella review. *Lancet Psychiatry* **7**, 162–172 (2020).
- Barnett, M. L., Lau, A. S. & Miranda, J. Lay health worker involvement in evidence-based treatment delivery: a conceptual model to address disparities in care. *Annu. Rev. Clin. Psychol.* **14**, 185–208 (2018).
- Singla, D. R. et al. Implementation and effectiveness of nonspecialist-delivered interventions for perinatal mental health in high-income countries: a systematic review and meta-analysis. *JAMA Psychiatry* **78**, 498–509 (2021).
- Singla, D. R. et al. Scaling up mental healthcare for perinatal populations: is telemedicine the answer? *Curr. Psychiatry Rep.* **24**, 881–887 (2022).
- Nair, U., Armfield, N. R., Chatfield, M. D. & Edirippulige, S. The effectiveness of telemedicine interventions to address maternal depression: a systematic review and meta-analysis. *J. Telemed. Telecare* **24**, 639–650 (2018).
- Stentzel, U. et al. Mental health-related telemedicine interventions for pregnant women and new mothers: a systematic literature review. *BMC Psychiatry* **23**, 292 (2023).
- Ciharova, M. et al. Amount and frequency of psychotherapy as predictors of treatment outcome for adult depression: a meta-regression analysis. *J. Affect. Disord.* **359**, 92–99 (2024).
- Dimidjian, S. et al. A pragmatic randomized clinical trial of behavioral activation for depressed pregnant women. *J. Consult. Clin. Psychol.* **85**, 26 (2017).

22. Patel, V. et al. The Healthy Activity Program (HAP), a lay counsellor-delivered brief psychological treatment for severe depression, in primary care in India: a randomised controlled trial. *Lancet* **389**, 176–185 (2017).
23. Jordans, M. J. et al. Effectiveness of psychological treatments for depression and alcohol use disorder delivered by community-based counsellors: two pragmatic randomised controlled trials within primary healthcare in Nepal. *Br. J. Psychiatry* **215**, 485–493 (2019).
24. Ekers, D. et al. Behavioural activation for depression; an update of meta-analysis of effectiveness and sub group analysis. *PLoS ONE* **9**, e100100 (2014).
25. Stein, A. T., Carl, E., Cuijpers, P., Karyotaki, E. & Smits, J. A. Looking beyond depression: a meta-analysis of the effect of behavioral activation on depression, anxiety, and activation. *Psychol. Med.* **51**, 1491–1504 (2021).
26. Dimidjian, S. et al. Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *J. Consult. Clin. Psychol.* **74**, 658–670 (2006).
27. Dobson, K. S. et al. Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the prevention of relapse and recurrence in major depression. *J. Consult. Clin. Psychol.* **76**, 468–477 (2008).
28. Richards, D. A. et al. Cost and outcome of behavioural activation versus cognitive behavioural therapy for depression (COBRA): a randomised, controlled, non-inferiority trial. *Lancet* **388**, 871–880 (2016).
29. Cox, J. L., Holden, J. M. & Sagovsky, R. Detection of postnatal depression: development of the 10-item Edinburgh Postnatal Depression Scale. *Br. J. psychiatry* **150**, 782–786 (1987).
30. Spitzer, R. L., Kroenke, K., Williams, J. B. & Löwe, B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch. Intern. Med.* **166**, 1092–1097 (2006).
31. Kounali, D. et al. How much change is enough? Evidence from a longitudinal study on depression in UK primary care. *Psychol. Med.* **52**, 1875–1882 (2022).
32. Kakuma, R. et al. Human resources for mental health care: current situation and strategies for action. *Lancet* **378**, 1654–1663 (2011).
33. Cuijpers, P., Karyotaki, E., Harrer, M. & Stikkelbroek, Y. Individual behavioral activation in the treatment of depression: a meta analysis. *Psychother. Res.* **33**, 886–897 (2023).
34. Surkan, P. J. et al. Anxiety-focused cognitive behavioral therapy delivered by non-specialists to prevent postnatal depression: a randomized, phase 3 trial. *Nat. Med.* **30**, 675–682 (2024).
35. Ghimire, S., Martinez, S., Hartvigsen, G. & Gerdes, M. Virtual prenatal care: a systematic review of pregnant women’s and healthcare professionals’ experiences, needs, and preferences for quality care. *Int. J. Med. Inform.* **170**, 104964 (2023).
36. Kim, J. et al. Telehealth utilization and associations in the United States during the third year of the COVID-19 pandemic: population-based survey study in 2022. *JMIR Public Health Surveill.* **10**, e51279 (2024).
37. Spaulding, E. M. et al. Prevalence and disparities in telehealth use among US adults following the COVID-19 pandemic: national cross-sectional survey. *J. Med. Internet Res* **26**, e52124 (2024).
38. McCoyd, J. L., Curran, L., Candelario, E. & Findley, P. ‘There is just a different energy’: changes in the therapeutic relationship with the telehealth transition. *Clin. Soc. Work J.* **50**, 325–336 (2022).
39. Mehrotra, A. & Uscher-Pines, L. Informing the debate about telemedicine reimbursement—what do we need to know. *N. Engl. J. Med.* **387**, 1821–1823 (2022).
40. Clark, D. M. et al. Transparency about the outcomes of mental health services (IAPT approach): an analysis of public data. *Lancet* **391**, 679–686 (2018).
41. Karyotaki, E. et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. *JAMA Psychiatry* **78**, 361–371 (2021).
42. Cuijpers, P. et al. The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: a meta-analysis. *Acta Psychiatr. Scand.* **144**, 288–299 (2021).
43. Van Ballegooijen, W. et al. Adherence to internet-based and face-to-face cognitive behavioural therapy for depression: a meta-analysis. *PLoS ONE* **9**, e100674 (2014).
44. *Maternal Mental Health in Canada, 2018/2019* (Statistics Canada, 2019).
45. 2020 Pregnancy risk assessment monitoring system (PRAMS) detailed data tables. *Illinois Department of Public Health* <https://dph.illinois.gov> (2022).
46. Bradshaw, H., Riddle, J. N., Salimgaraev, R., Zhaunova, L. & Payne, J. L. Risk factors associated with postpartum depressive symptoms: a multinational study. *J. Affect Disord.* **301**, 345–351 (2022).
47. Dharma, C. et al. Risk for maternal depressive symptoms and perceived stress by ethnicities in Canada: from pregnancy through the preschool years. *Can. J. Psychiatry* **64**, 190–198 (2019).
48. Getahun, D. et al. Trends in postpartum depression by race/ethnicity and pre-pregnancy body mass index. *Am. J. Obstet. Gynecol.* **228**, S122–S123 (2023).
49. *Census Profile, 2021 Census of Population* (Statistics Canada, 2023).
50. *Census 2021: Families, Households, Marital Status and Income* (City of Toronto, 2022).
51. *Focus on Geography Series, 2021 Census of Population, Toronto* (Statistics Canada, 2022).
52. Provisional natality, 2023 through last month request. *CDC* <https://wonder.cdc.gov> (2024).
53. *Birth Statistics* (Illinois Department of Public Health, accessed 9 October, 2024); <https://dph.illinois.gov>
54. *Illinois* (United States Census Bureau, 2023).
55. *North Carolina* (United States Census Bureau, 2023).
56. *HDPulse: An Ecosystem of Minority Health and Health Disparities Resources*. *North Carolina* (National Institute on Minority Health and Health Disparities, 2022).
57. Cuzick, J. & Sasieni, P. Interpreting the results of noninferiority trials—a review. *Br. J. Cancer* **127**, 1755–1759 (2022).
58. Tweed, C. D. et al. Exploring different objectives in non-inferiority trials. *Br. Med. J.* **385**, e078000 (2024).
59. Mao, F. et al. Sensitivity to change and minimal clinically important difference of Edinburgh Postnatal Depression Scale. *Asian J. Psychiatry* **66**, 102873 (2021).
60. Attkisson, C. C. & Greenfield, T. K. *The UCSF Client Satisfaction Scales: I. The Client Satisfaction Questionnaire-8* (Lawrence Erlbaum Associates, 2004).
61. Herdman, M. et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual. Life Res.* **20**, 1727–1736 (2011).
62. Zimet, G. D., Dahlem, N. W., Zimet, S. G. & Farley, G. K. The multidimensional scale of perceived social support. *J. Personal. Assess.* **52**, 30–41 (1988).
63. Singla, D. R. et al. Multiple mediation analysis of the peer-delivered Thinking Healthy Programme for perinatal depression: findings from two parallel, randomised controlled trials. *Br. J. Psychiatry* **218**, 143–150 (2021).
64. Munder, T., Wilmers, F., Leonhart, R., Linster, H. W. & Barth, J. Working Alliance Inventory–Short Revised (WAI–SR): psychometric properties in outpatients and inpatients. *Clin. Psychol. Psychother.* **17**, 231–239 (2010).

65. Üstün, T. B. et al. Developing the World Health Organization disability assessment schedule 2.0. *Bull. World Health Organ.* **88**, 815–823 (2010).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the

Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

<sup>1</sup>Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Ontario, Canada. <sup>2</sup>Department of Psychiatry, Sinai Health, Toronto, Ontario, Canada. <sup>3</sup>Lunenfeld-Tanenbaum Research Institute, Toronto, Ontario, Canada. <sup>4</sup>Department of Psychiatry, Temerty Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada. <sup>5</sup>Department of Obstetrics and Gynecology, Endeavor Health (formerly Northshore University Health System), Evanston, IL, USA. <sup>6</sup>Department of Obstetrics and Gynecology, Pritzker School of Medicine, University of Chicago, Chicago, IL, USA. <sup>7</sup>Department of Psychiatry, Women's College Hospital, Toronto, Ontario, Canada. <sup>8</sup>Department of Psychiatry, School of Medicine, University of North Carolina, Chapel Hill, NC, USA. <sup>9</sup>Sunnybrook Research Institute, Toronto, Ontario, Canada. <sup>10</sup>Department of Psychology, Vanderbilt University, Nashville, TN, USA. <sup>11</sup>Lawrence S. Bloomberg Faculty of Nursing, University of Toronto, Toronto, Ontario, Canada. <sup>12</sup>Waypoint Centre for Mental Health Care, Penetanguishene, Ontario, Canada. <sup>13</sup>Department of Experimental Psychology, University of Oxford, Oxford, UK. <sup>14</sup>Postpartum Support International, Portland, OR, USA. <sup>15</sup>Department of Epidemiology, Gillings School of Public Health, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. <sup>16</sup>Department of Obstetrics and Gynaecology, Sinai Health, Toronto, Ontario, Canada. <sup>17</sup>Department of Obstetrics and Gynaecology, University of Toronto, Toronto, Ontario, Canada. <sup>18</sup>Department of Global Health and Social Medicine, Harvard Medical School, Boston, MA, USA. <sup>19</sup>Department of Obstetrics and Gynecology, School of Medicine, University of North Carolina, Chapel Hill, NC, USA. <sup>20</sup>Department of Maternal-Child Health, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA. <sup>21</sup>Life With A Baby, Thornhill, Ontario, Canada. <sup>22</sup>Department of Global Health and Population, Harvard Chan School of Public Health, Boston, MA, USA. <sup>23</sup>These authors contributed equally: J. Jo Kim, Laura M. La Porte, Paula Ravitz, Crystal E. Schiller. <sup>24</sup>These authors jointly supervised this work: Samantha Meltzer-Brody, Richard K. Silver, Benoit H. Mulsant, Vikram Patel.

✉ e-mail: [daisy.singla@utoronto.ca](mailto:daisy.singla@utoronto.ca)

## Methods

### Study design

SUMMIT was a pragmatic, multisite, four-arm, individual randomized, noninferiority trial conducted in Chapel Hill (North Carolina), Evanston/Chicago (Illinois) and Toronto (Canada) at university-affiliated healthcare settings. In North Carolina, recruitment was conducted at three clinical sites affiliated with the University of North Carolina (UNC) Women's and Neuroscience Hospitals. In Illinois, recruitment was conducted at 14 obstetric and family medicine clinics affiliated with Endeavor Health (formerly known as Northshore University Health System) and the University of Chicago. In Toronto, recruitment was conducted at Mount Sinai Hospital, Women's College Hospital and St. Michael's Hospital, all of which are affiliated with the University of Toronto. The study received ethical approvals from the following three institutional review boards (IRBs): UNC Biomedical IRB (19-1786), Endeavor Health IRB (EH18-129) and Clinical Trials Ontario (1895). All participants provided written informed consent before enrollment. An independent DSMB supervised the trial (Supplementary Table 15). The trial design, protocol and oversight have been described previously<sup>66</sup>.

### Participants

Full eligibility criteria and recruitment procedures are detailed in the published protocol<sup>66</sup>. In brief, inclusion criteria included being a pregnant ( $\leq 36$  weeks) or postpartum (4–30 weeks) adult ( $\geq 18$  years; inclusive of gender identities and birthing persons) with a score  $\geq 10$  on the EPDS<sup>29</sup> and speaking English or Spanish. A cutoff score of EPDS  $\geq 10$  was used to encompass both minor and major depression<sup>67</sup>. This threshold has been used in antenatal, postnatal and community-based populations with high sensitivity<sup>68</sup>, particularly for specific cultural groups (for example, ref. 69). Exclusion criteria included active suicidal intent; active symptoms of psychosis, mania or substance misuse; change in psychotropic medication within 2 weeks of beginning treatment; ongoing psychotherapy; and severe fetal anomalies, stillbirth or infant death.

Between January 2020 and October 2023, participants were recruited through self, internal or external referrals. Internal referrals were elicited from clinicians from site hospitals and satellite clinics (that is, obstetrical, mental health and family departments) who sent patient contact information directly to the research team. External referrals were received from clinicians at sites that were not affiliated with the trial. Finally, recruitment materials with contact information for the research team (that is, brochures and posters) were available in clinics and interested individuals contacted the team for more information (self-referrals). The study was introduced to potential participants by either their clinical provider (for internal and external referrals) or a trained research assistant at their respective sites (for self-referrals). Before contacting potential participants, research staff were trained in obtaining informed consent to meet ethical standards. In addition, all research staff interacting directly with participants engaged in role plays where interpersonal skills such as empathy and a nonjudgemental stance were emphasized. After providing consent, participants were screened for eligibility. Informed consent was obtained in-person or via telephone, and most occurred virtually using a combination of phone or Zoom.

### Intervention

All participants were offered the same manualized BA intervention over six to eight individual weekly sessions. BA relies on increasing enjoyable or fulfilling activities that align with one's values and targeting key mechanisms of patient activation and avoidant coping<sup>70</sup>. The intervention manual was adapted from two well-established manuals: the Healthy Activity Program from India<sup>22</sup> and the Alma Program for perinatal populations in Colorado. As described elsewhere<sup>66,71</sup>, key strategies included psychoeducation, behavioral assessment, values-based activity monitoring and structuring, and problem solving and interpersonal effectiveness, all through a culturally sensitive lens<sup>71</sup>.

**Intervention arms.** Participants were randomized to one of four arms: nonspecialist telemedicine, specialist telemedicine, nonspecialist in-person or specialist in-person.

**Providers.** Details on provider recruitment, training and supervision have been published<sup>66,72</sup>. Briefly, providers were recruited through listservs advertising the casual position of a SUMMIT treatment provider, and by direct contact and word of mouth through stakeholders. Nonspecialists were healthcare workers (that is, registered nurses, midwives and doulas) without any formal mental health training or experience delivering psychotherapy, ascertained through resume review and the hiring interview process. In Canada and the United States, these healthcare workers are already embedded in pregnancy and postpartum care<sup>16</sup>. Specialists were professionals with formal mental health training (that is, psychiatrists, psychologists or social workers) and  $\geq 5$  years of experience delivering psychotherapies.

**Training phase:** both nonspecialists and specialists received the same BA training, provided by a minimum of two clinical leads (expert clinicians). The workshops utilized observation and didactics with interactive educational strategies including role play, games and homework. Nonspecialists and specialists meeting competency standards<sup>73,74</sup> were selected for the 8 week internship phase of the trial during which they implemented BA treatment with up to two participants, with supervision from the site-level clinical lead. Only nonspecialists and specialists who achieved competence, as assessed by standardized role plays and therapy quality assessments<sup>72,73</sup>, were selected to deliver BA during the trial.

**Trial phase:** once the internship phase was complete, specialists attended weekly to monthly peer supervision and contacted the clinical lead on an ad hoc basis. Nonspecialists also participated in weekly to bimonthly supervision with their site clinical lead, as well as monthly measurement-based supervision where clinical leads and nonspecialists rated individual audio-recorded sessions for therapy quality. Audio-recorded treatment sessions were rated using the Q-SUMMIT, a brief, validated 20-item therapy quality checklist that examined the extent to which an individual treatment provider exhibited treatment-specific skills (for example, establishing an agenda) and general skills (for example, using collaboration and nonjudgemental stance) on a scale of zero to four. Using the same Q-SUMMIT measure, independent raters evaluated intervention fidelity on at least 5% of all treatment session audiotapes at each site, with good to excellent interrater reliability ( $\kappa = 0.75$ – $0.85$ ). Both nonspecialists and specialists who did not reach the cutoff for specific items received refresher training by the hub clinical leads. The Q-SUMMIT is freely available on the SUMMIT website (<https://thesummittrial.com/external-resources/bamanual-and-materials/>).

**Modality.** Telemedicine was implemented via WebEx in Chapel Hill and Zoom in Evanston/Chicago and Toronto, in compliance with local privacy laws. Participants attended in-person BA sessions at their outpatient clinical site or telemedicine BA sessions at their preferred private location. If needed, telemedicine participants were provided with study tablets and internet access. Seven telemedicine participants (or 0.57%) required study tablets or internet access.

### Involvement of stakeholders

The SUMMIT trial involved multiple and diverse stakeholders to inform the development, implementation and dissemination strategies of the study. From the study's inception, we consulted individuals with lived experience, family members, patient advocates, clinical experts, healthcare administrators, payers and policy members, and these stakeholders were integral members of our team, alongside researchers and clinicians. In addition, we also conducted qualitative research with a range of SUMMIT stakeholders, including (but not limited to) participant representation from each of the four arms, significant

others (including spouses or partners of participating mothers) and a diverse range of other stakeholders to inform key decisions about the trial including the cadre of nonspecialist provider<sup>75</sup>, training and supervision processes<sup>76</sup>, culturally sensitive care<sup>71</sup> and resuming in-person sessions during the COVID-19 pandemic<sup>77</sup>. Insights from our diverse stakeholder group were vital to informing patient-centered delivery and implementation of the SUMMIT trial.

### Randomization and masking

Individual participants were randomized within REDCap to one of four study arms and randomization was stratified by perinatal period and site. An independent biostatistician generated the randomization sequence through computer-generated lists, with random blocks of four and eight, and stratified by perinatal period and site. The randomization sequence was concealed until all 3 month data were analyzed. Several strategies were used to reduce bias. First, all team members were masked to study arm allocation except for the data site coordinator, participants, treatment providers and clinical leads (who did not have access to aggregate data<sup>66</sup>). Second, all treatment providers introduced themselves as a 'SUMMIT treatment provider' to reduce potential bias of provider preference among study participants. All data were self-reported through REDCap and reviewed by independent data staff to reduce the number of staff who were unmasked. Third, differences in clinical and demographic characteristics between self-referrals compared with other referral sources were balanced. Trial recruitment began in January 2020 and participants were initially randomized 1:1:1:1 to each arm (Extended Data Fig. 1). Due to the COVID-19 pandemic, in-person sessions were paused, and participants were randomized 1:1 to the two telemedicine arms between March 2020 and July 2021. Once healthcare institutions allowed in-person visits, participants were randomized in a weighted 3:1 scheme favoring in-person conditions (July 2021 to January 2022). Finally, we returned to 1:1:1:1 randomization in January 2022 (see the Statistical Analysis Plan (SAP)<sup>78</sup> or Extended Data Fig. 1 for more details).

### Database and handling

Quantitative data were collected through standardized REDCap<sup>TM</sup> databases, securely hosted on institutional servers at each of the three participating hubs. Deidentified data from the United States hubs were extracted, encrypted and transferred to Toronto, where they were subsequently uploaded into the REDCap system. The data were protected by stringent measures, including restricted access, 24 h surveillance and continuous monitoring. Upon providing informed consent, participants were assigned unique study IDs and received email links to complete their baseline, session-wise and outcome assessments linked to their study ID profiles in REDCap. Personal information associated with study IDs was kept in a separate password-protected, encrypted file on secure institutional servers. Email addresses were stored separately from study data on a secure backend server within REDCap, ensuring protection from unauthorized access and data export.

### Procedures

Participants completed a baseline assessment and were then randomized by trained research assistants using REDCap. All reported outcome data were collected via self-report through REDCap at 3 months post-randomization. All staff and providers were trained to follow an established and detailed safety guide. This included the monitoring of SAEs, which were defined as fatal or life-threatening events to the participant, fetus or child; disability or permanent damage; hospitalization; and other serious medical events and AEs, which included active suicide intent (regardless of hospitalization) or development of psychosis and mania. When potential SAEs or AEs were identified, they were reviewed by the site coordinator, trial psychiatrist and site principal investigator. All SAEs and AEs were documented by the site-level principal investigator and research staff, followed by an expedited review by the DSMB.

**Trial changes due to the COVID-19 pandemic.** The COVID-19 pandemic required (1) changing the randomization scheme (see above), (2) treating 21 participants who were initially randomized to the in-person modality via telemedicine during the peaks of the pandemic due to institutional policies and, consequently, (3) temporarily ceasing in-person recruitment and changing the noninferiority margin for modality (in-person versus telemedicine) from 10% to 13% (see SAP), which reduced the required sample size for in-person arms and the overall sample size from 1,368 to 1,226 (see below). These modifications were approved by our DSMB and research ethics committees across sites. Research has demonstrated that the minimal clinically important difference (MCID) on our primary outcome, the EPDS, is typically four points (95% CI 1.4–6.5)<sup>59,79</sup>. In line with noninferiority guidelines, the revised noninferiority margin of 13% was determined to consider the lower bound of this MCID (that is, EPDS 1.4). In addition, we conducted a post hoc sensitivity analysis using the (pre-COVID-19 original) lower noninferiority margin of 10% for in-person versus telemedicine. Finally, the study team decided to simplify the study protocol by (1) aligning secondary outcomes with predefined secondary aims. In doing so, several outcomes (trauma symptoms, perceived support, activation levels and quality of life) in the protocol were redefined as exploratory outcomes. With the exception of trauma symptoms, the analyses of these exploratory outcomes at 3 months post-randomization are presented in the current study; and (2) defining AE/SAE according to the SAP and approved ethics protocols to ensure clarity when assessing active suicidal intent. In short, when the published trials protocol and the SAP are discrepant, the text and analysis of the paper are congruent with the SAP (which was completed after the published protocol and refined some analyses).

### Outcomes

The predefined primary outcome was the EPDS<sup>29</sup> total score at 3 months post-randomization. As the treatment was offered on a weekly basis and typically takes up to 3 months to complete, 3 months post-randomization reflects a post-treatment outcome and is commonly used in other psychological treatment trials<sup>42</sup>. The EPDS is a ten-item, Likert scale (range 0–30) that is freely available in multiple languages. Consistent with the pragmatic trial design of SUMMIT, the ten-item EPDS scale was selected for the primary outcome, rather than a diagnostic interview, for several reasons: its high sensitivity, it is the most common measure of parental depressive symptoms in clinical and research settings, and it is feasible to implement at scale<sup>80</sup>. The secondary outcome was the GAD-7 (ref. 30) total score. Exploratory outcomes included patient-reported quality of life (as assessed through the World Health Organization Disability Assessment Schedule 2.0 (ref. 65) and EQ-5D 5-Level<sup>61</sup>), client satisfaction (Client Satisfaction Questionnaire-8 (ref. 60)), therapeutic alliance (Working Alliance Inventory–Short Form<sup>64</sup>), activation levels (Premium Abbreviated Activation Scale<sup>63</sup>) and perceived social support (Multidimensional Scale of Perceived Social Support<sup>62</sup>). All measures are global measures and were selected because of their robust psychometric properties across diverse settings. With regards to minor discrepancies between the listed secondary and exploratory outcomes in the published protocol and SAP, the text and analysis of the paper are congruent with the SAP (which was completed after the published protocol and refined some analyses). The current study focused on outcomes assessed at 3 months post-randomization. Data collection for all additional analyses related to sustained outcomes, child development outcomes, economic evaluation and qualitative data are ongoing and will therefore be published separately.

### Sample size

Using a noninferiority margin of 10%, an EPDS mean estimate of 7.93 (s.d. 4.68) (ref. 81), 80% power and alpha ( $\alpha$ ) of 0.05, the comparison of provider (nonspecialist versus specialist) required 431 participants in each of the two conditions. To account for 10% dropout, the sample size was inflated to  $N = 958$ . As described in our detailed SAP and

based on recent noninferiority guidelines, we did not adjust our two primary hypotheses for multiplicity because they did not involve different endpoints (that is, both hypotheses test total EPDS scores at 3 months post-randomization). Specifically, when multiple hypotheses test a similar underlying outcome, no adjustment for multiplicity is required<sup>82,83</sup>. As described above, and due to COVID-19, we increased the noninferiority margin for the modality comparison (telemedicine versus in-person) from 10% to 13% and used the same EPDS estimate, 80% power and  $\alpha = 0.05$ . The comparison of telemedicine to in-person delivery required an additional 268 participants, yielding a target sample size of 1,226. While the study is a multicenter one, randomization occurred within each individual site. Thus, no cluster randomization was carried out and the sample size was not inflated to account for an intracluster correlation. The sample size calculation was run using PASS version 12 (Power Analysis and Sample Size Software; NCSST LLC).

### Statistical analysis

Baseline variables included demographic characteristics (for example, self-reported age, race/ethnicity, sex and gender identity and marital status). All baseline variables of interest were summarized using means and two-sided 95% CIs for continuous measures, and percentages for categorical measures.

The primary analysis was a noninferiority *t*-test, using a one-sided, 95% CI around the difference in EPDS scores based on the actual primary endpoint data. Noninferiority was determined if the upper limit of the 95% CI of the mean score was below the noninferiority margin.

We selected predetermined noninferiority margins of 10% for provider and 13% for modality, which were calculated based on the actual primary endpoint data<sup>58</sup>. These noninferiority margins were selected to ensure clinically relevant conclusions of noninferiority between conditions were met. Specifically, and in line with guidelines for noninferiority trials, we selected these noninferiority margins to correspond with the MCID and to ensure that they were lower than any potential superiority effect between groups. Compared with other psychotherapy trials, which typically use noninferiority margins of 24–60%, our prespecified margins were intentionally conservative to ensure that noninferiority conclusions between providers and modalities could be ascertained. The same noninferiority margins were used to examine noninferiority for secondary and exploratory outcomes.

All primary and secondary noninferiority analyses were run as both ITT (based on randomization arms) without imputed data and then with imputed data (Supplementary Table 9), and a PP analysis based on the condition (provider: nonspecialist versus specialist, or modality: telemedicine or in-person) that the participant actually received. A sensitivity analysis omitting protocol deviations was also conducted (Supplementary Table 6). The key noninferiority analyses were based on noninferiority *t*-tests across the entire patient sample (under the assumption that randomization will serve to balance out potential confounders). Each *t*-test looked at the CI around the difference in outcome based on the actual data (for example, EPDS scores between nonspecialist and specialist or between telemedicine and in-person) to see whether the upper bound contained the noninferiority margin (upper bound for the case in which we take the difference to be nonspecialist minus specialist (or telemedicine minus in-person) and where higher scores indicate a worse outcome). In addition, we tested for a modality by provider interaction for our primary and secondary outcomes.

Secondary analyses repeated the same analyses with GAD-7 scores at 3 months post-randomization. Linear regression models were used as an additional analysis to assess noninferiority while accounting for preselected conceptually and empirically driven baseline covariates (Supplementary Table 7), such as study site, perinatal period (antenatal versus postnatal), education levels, treatment preference for provider or modality and baseline severity. Linear mixed models assessed the moderating effects of severity at baseline, with participants as a random effect and a treatment-by-time interaction. Exploratory analyses

examined whether noninferiority held for additional exploratory outcomes. Further, we examined a potential pandemic effect, testing for an interaction between providers (specialist and nonspecialist) and timing (1:1 randomization to nonspecialist and specialist in telemedicine only versus 3:1 randomization favoring in-person) in relation to change in EPDS and GAD-7 scores in a linear regression model. We also conducted a sensitivity analysis on our primary outcome (EPDS scores at 3 months post-randomization), which excluded participants recruited during pre-COVID-19 1:1:1:1 randomization (Supplementary Table 16). An ICC (and its associated CI) was calculated to ensure that clustering at the level of the treatment provider would not be a concern for the key outcomes of interest.

Multiple imputation methods were conducted to impute missing data using fully conditional specification methods with Proc MI and Proc MIANALYZE in SAS. Fifteen imputed datasets were created to align with the percent missingness, with model results averaged across the 15 iterations<sup>84</sup>. Sensitivity analyses were carried out since missing data lead to the use of multiple imputation methods. These analyses compared the results of the models on the imputed data with the ones with the actual missing data included. Additional sensitivity analyses (for example, psychotropic medications and participants without BA sessions) were conducted (Supplementary Table 17). Outliers were defined as scores that were greater than  $1.5 \times$  the interquartile range + the third quartile OR lower than the first quartile –  $1.5 \times$  the interquartile range. The trial biostatistician conducted all analyses using SAS version 9.4 (SAS Institute).

The study was registered at [clinicaltrials.gov](https://clinicaltrials.gov) (NCT04153864) before study recruitment on 6 November 2019.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Two years after publication, all individual participant deidentified data will be shared with researchers who submit a written proposal to: [summittrial@sinahealth.ca](mailto:summittrial@sinahealth.ca). Data will be available indefinitely. This timeline allows the study team to collect long-term study outcomes and analyze and potentially publish the results of the secondary aims involving sustained outcomes at 12 months post-randomization as indicated in our detailed SAP. Researchers requesting data will receive a response from the study team within three business days.

### Code availability

Two years after publication, analytic code will be shared with researchers who submit a formal written proposal to [summittrial@sinahealth.ca](mailto:summittrial@sinahealth.ca). Code will be available indefinitely. This timeline allows the study team to analyze and potentially publish the results of the secondary aims involving sustained outcomes at 12 months post-randomization as indicated in our detailed SAP. Researchers requesting the code will receive a response from the study team within three business days.

### References

66. Singla, D. R. et al. Scaling up maternal mental healthcare by increasing access to treatment (SUMMIT) through non-specialist providers and telemedicine: a study protocol for a non-inferiority randomized controlled trial. *Trials* **22**, 186 (2021).
67. Wisner, K. L. et al. Onset timing, thoughts of self-harm, and diagnoses in postpartum women with screen-positive depression findings. *JAMA Psychiatry* **70**, 490–498 (2013).
68. Levis, B., Negeri, Z., Sun, Y., Benedetti, A. & Thombs, B. D. Accuracy of the Edinburgh Postnatal Depression Scale (EPDS) for screening to detect major depression among pregnant and postpartum women: systematic review and meta-analysis of individual participant data. *Br. Med. J.* **371**, m4022 (2020).

69. Salehi, A., Zhang, M., Kithulegoda, N., Vigod, S. & Ivers, N. Validation of the culturally adapted Edinburgh Postpartum Depression Scale among east Asian, southeast Asian and south Asian populations: a scoping review. *Int J. Ment. Health Nurs.* **32**, 1616–1635 (2023).
70. Dimidjian, S., Barrera, M. Jr, Martell, C., Muñoz, R. F. & Lewinsohn, P. M. The origins and current status of behavioral activation treatments for depression. *Annu. Rev. Clin. Psychol.* **7**, 1–38 (2011).
71. Singla, D. R. et al. Culturally sensitive psychotherapy for perinatal women: a mixed methods study. *J. Consult. Clin. Psychol.* **90**, 770–786 (2022).
72. Singla, D. R. et al. Scaling up quality-assured psychotherapy: the role of therapist competence on perinatal depression and anxiety outcomes. *Gen. Hosp. Psychiatry* **83**, 101–108 (2023).
73. Singla, D. R. et al. Improving the scalability of psychological treatments in developing countries: an evaluation of peer-led therapy quality assessment in Goa, India. *Behav. Res. Ther.* **60**, 53–59 (2014).
74. Restivo, J. L. et al. Assessing health worker competence to deliver a brief psychological treatment for depression: development and validation of a scalable measure. *J. Behav. Cogn. Ther.* **30**, 253–266 (2020).
75. Singla, D. R. et al. Implementing psychological interventions through nonspecialist providers and telemedicine in high-income countries: qualitative study from a multistakeholder perspective. *JMIR Ment. Health* **7**, e19271 (2020).
76. Singla, D. R. et al. Adapting behavioral activation for perinatal depression and anxiety in response to the COVID-19 pandemic and racial injustice. *J. Affect. Disord.* **299**, 180–187 (2022).
77. Andrejek, N. et al. Barriers and facilitators to resuming in-person psychotherapy with perinatal patients amid the COVID-19 pandemic: a multistakeholder perspective. *Int. J. Environ. Res. Public Health* **18**, 12234 (2021).
78. Scaling up maternal mental healthcare by increasing access to treatment (SUMMIT): a pragmatic, non-inferiority trial for perinatal depression and anxiety. *ClinicalTrials.gov* <https://clinicaltrials.gov/study/NCT04153864> (2024).
79. Matthey, S. Calculating clinically significant change in postnatal depression studies using the Edinburgh Postnatal Depression Scale. *J. Affect. Disord.* **78**, 269–272 (2004).
80. El-Den, S., O'Reilly, C. L. & Chen, T. F. A systematic review on the acceptability of perinatal depression screening. *J. Affect. Disord.* **188**, 284–303 (2015).
81. Dennis, C.-L. et al. Effect of peer support on prevention of postnatal depression among high risk women: multisite randomised controlled trial. *BMJ* **338**, a3064 (2009).
82. Food and Drug Administration. *Multiple Endpoints in Clinical Trials Guidance for Industry* (Center for Drug Evaluation and Research, 2017).
83. *Points to Consider on Multiplicity Issues in Clinical Trials* (Committee for Proprietary Medicinal Products, 2002).
84. Rubin, D. B. Multiple imputation after 18+ years. *J. Am. Stat. Assoc.* **91**, 473–489 (1996).

## Acknowledgements

We thank the patients and treatment providers who participated in this trial, members of the independent data and safety management board (S. Goodman, PhD (chair), R. Gibbons, PhD, C.E. Monk, PhD and T. Oberlander, MD), our stakeholders (A. Beck, J. Burkhard, S. Dimidjian, A. Griffen, B. Gunyon-Meyer, R. Lamba, W. Lopez, S. Lye, J. Maggi,

R. Maunder, F. Meffe, M. Sabini, S. Sockalingham, M. Sermer, P. Tomasi, K. Tully, L. Wiesenfeld and J. Woodgett), and research staff involved in the trial. We also thank the Patient-Centered Outcomes Research Institute award (PCS-2018C1-10621, to D.R.S.) who funded SUMMIT. The Patient-Centered Outcomes Research Institute had no role in the writing of this paper or the decision to submit it for publication.

## Author contributions

D.R.S. conceptualized the study, with initial input from V.P. D.R.S. designed the study methodology, with contributions from V.P., C.-L.D., A.K., S.M.-B., R.K.S. and S.N.V. D.R.S. and A.S.L. led efforts in funding acquisition. Alongside data staff, T.S.B. cleaned the data; J.E.G., T.S.B., N.S.-M. and H.A.K. performed data curation; and A.K. conducted the formal analysis. D.R.S., S.M.-B., R.K.S., S.N.V., N.S.-M., S.R.K., T.S.B., A.S.L., H.A.K. and J.E.G. managed and coordinated responsibility for the research activity planning and execution. D.R.S., S.M.-B., R.K.S. and S.N.V. provided oversight and leadership for the research activity planning and execution, with oversight from N.S.-M., T.S.B., H.A.K. and J.E.G.; J.J.K., L.M.L.P., P.R. and C.E.S. provided oversight of clinical intervention delivery. A.K. verified raw study data and is responsible for replication/reproducibility of results and other research outputs. D.R.S., A.S.L. and N.S.-M. wrote the original draft of this paper, with significant inputs from B.H.M., V.P., S.M.-B., R.K.S., S.N.V. and S.D.H. All authors reviewed and edited subsequent drafts. All authors had full access to all the data in the study, read the final paper, and had final responsibility for the decision to submit for publication.

## Competing interests

B.N.G. reports compensation for authorship/reviewer for UpToDate chapters on depression. M.L. receives book royalties from Hachette Books and Norton Book. S.M.-B. receives research funding to UNC for clinical trials sponsored by Sage Therapeutics, Electromedical Products International and Sirtsei Pharmaceuticals for clinical trials. She serves as a clinical advisor and professional corporation owner for Modern Health, a digital health company. She also serves as a scientific advisor to EmbarkNeuro and Seaport Therapeutics. B.H.M. holds and receives support from the Labatt Family Chair in Biology of Depression in Late-Life Adults at the University of Toronto. He also receives compensation from the Department of Psychiatry, University of Toronto, Toronto, Ontario; and the Centre for Addiction and Mental Health, Toronto, Ontario. S.N.V. reports royalties from UpToDate Inc. for authorship of materials on depression and pregnancy.

## Additional information

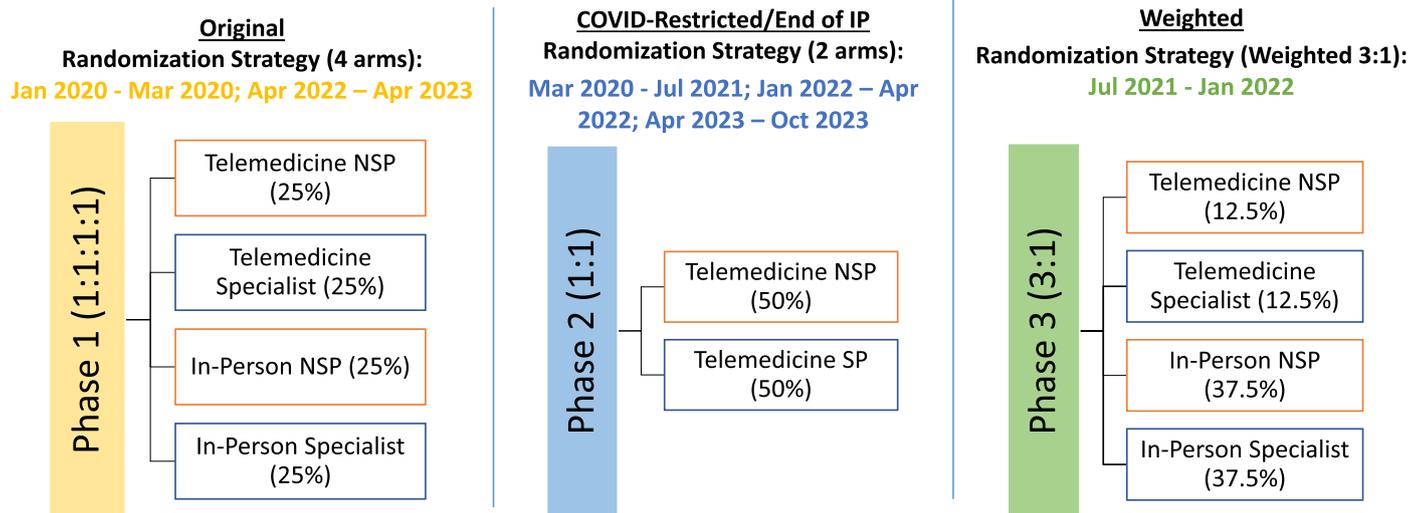
**Extended data** is available for this paper at <https://doi.org/10.1038/s41591-024-03482-w>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41591-024-03482-w>.

**Correspondence and requests for materials** should be addressed to Daisy R. Singla.

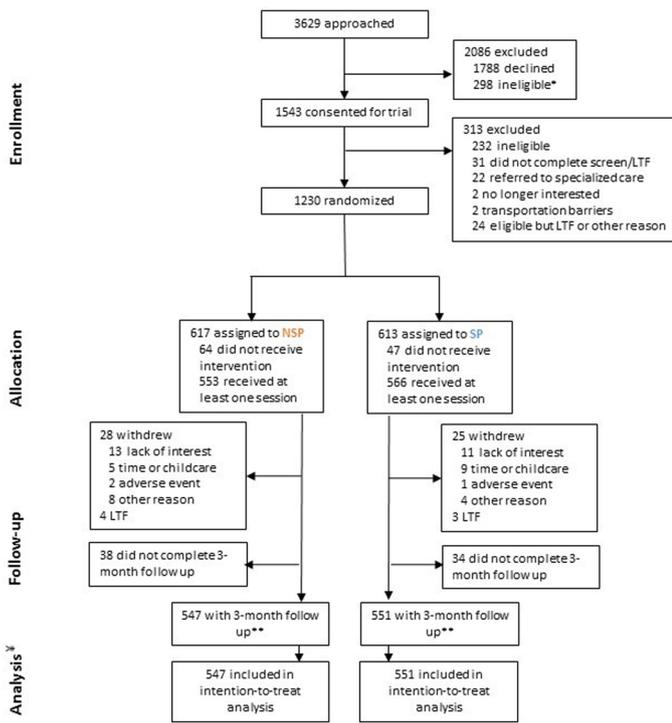
**Peer review information** *Nature Medicine* thanks Zong-Chao Liu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Ming Yang, in collaboration with the *Nature Medicine* team.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

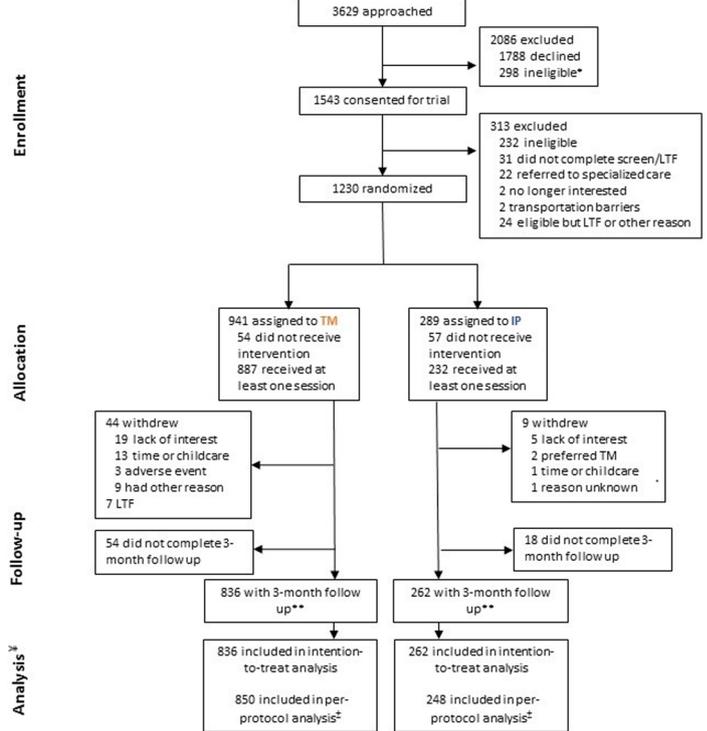


**Extended Data Fig. 1 | Randomisation strategy.** Note. In-Person (IP). Randomisation shifted to the two telemedicine (TM) arms based on site specific, COVID-related institutional restrictions, and once enrollment to in-person arms was complete.

**A Enrollment, Randomization and Follow-up of the Participants by Provider Condition**

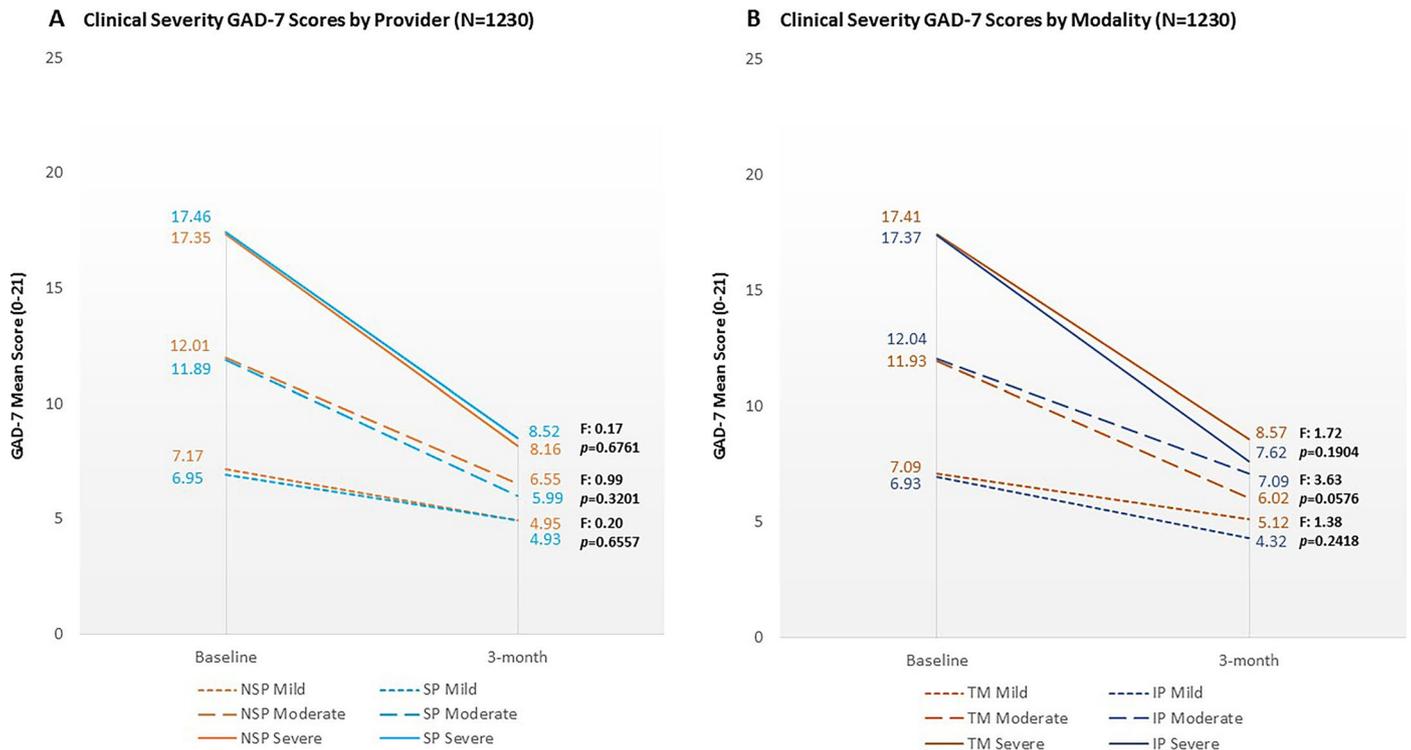


**B Enrollment, Randomization and Follow-up of the Participants by Modality Condition**



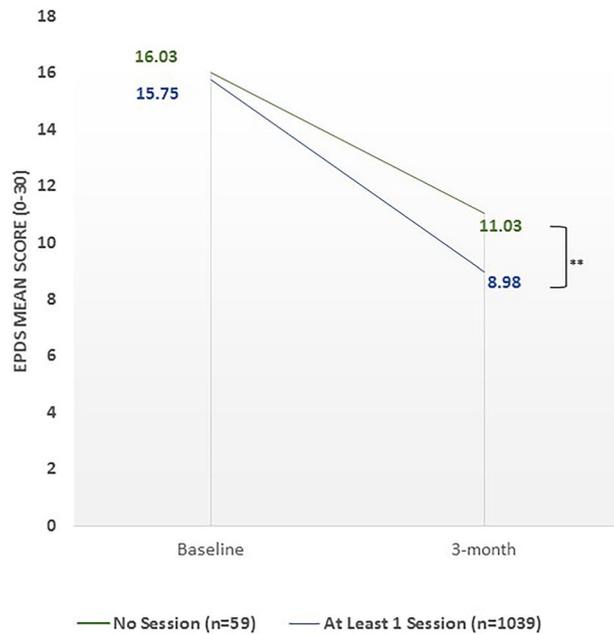
**Extended Data Fig. 2 | Enrollment, randomisation and follow-up of the participants by condition.** 1A. Participants (N=1230) were randomly assigned to the provider condition: Non-specialist provider (NSP) or Specialist provider (SP) and all were offered Behavioral Activation. 1B. Participants (N=1230) were randomly assigned to the modality condition: Telemedicine (TM) or In-person (IP) and all were offered Behavioral Activation. *Note.* Lost to follow-up (LTF).

\*Includes one participant who was a screen fail. \*\*Reflects participants who completed the primary outcome (EPDS). <sup>‡</sup>Per protocol analyses were only conducted for modality (TM vs. IP) comparisons because protocol deviations were defined as instances where participants were switched from in-person to telemedicine due to the COVID-19 pandemic. <sup>§</sup>Imputed analyses including the full sample (N=1230) were also conducted for all ITT and PP analysis.

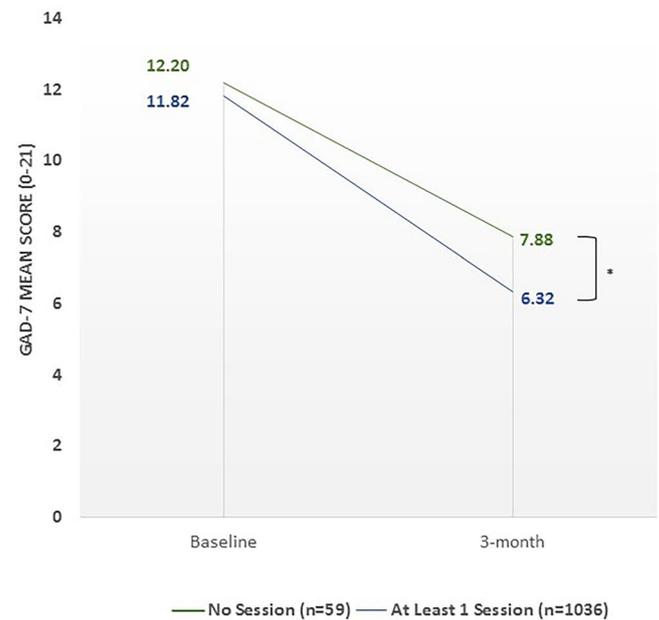


**Extended Data Fig. 3 | Comparing baseline clinical severity anxiety (GAD-7) scores by provider (Supplementary Fig. 3a) and modality (Supplementary Fig. 3b).** *Note.* Generalized Anxiety Disorder (GAD-7). Clinical severity GAD-7 scores were assessed at baseline. Change over time within each severity group

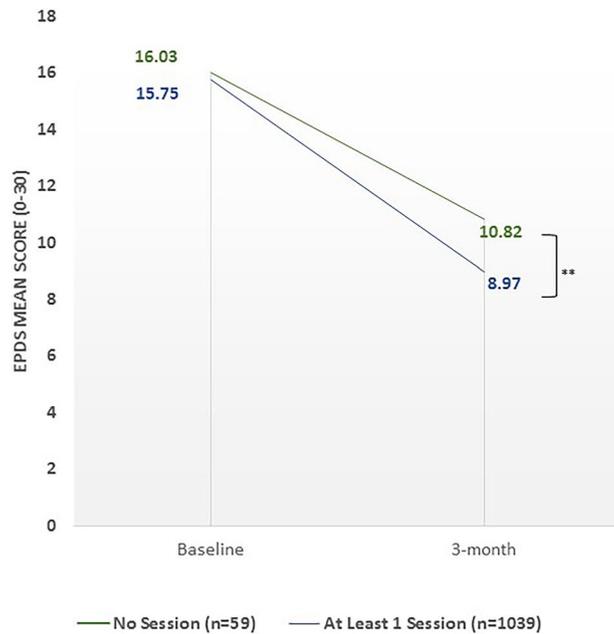
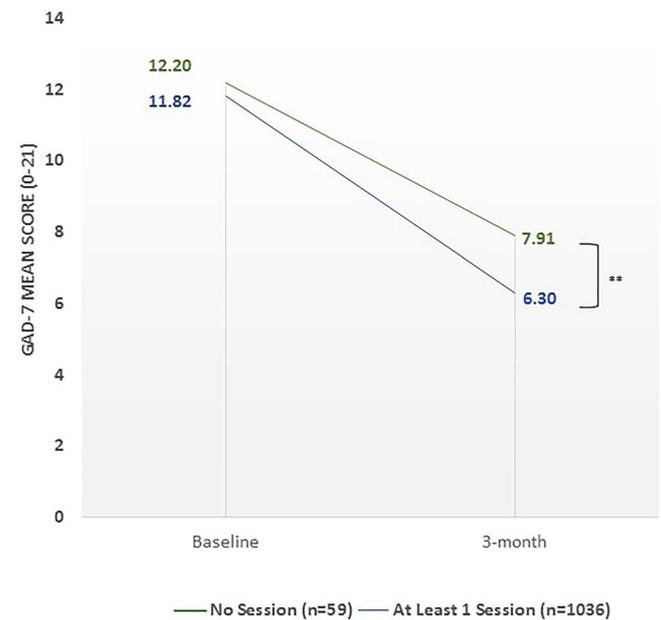
between conditions by provider (a) and modality (b) were not significant. Severity groups based on baseline GAD-7 score: Mild (5-9), Moderate (10-14), Severe (15-21). F-statistic and p-value corresponds to the time by condition interaction term from a linear mixed-model.

**A Depressive Symptoms**

**Extended Data Fig. 4 | Comparing non-imputed depressive (EPDS) and anxiety (GAD-7) symptom scores among participants who did and did not complete BA sessions.** *Note.* Edinburgh Postnatal Depression Scale (EPDS), General Anxiety Disorder-7 (GAD-7); \* $p < 0.05$ ; \*\* $p < 0.01$ . The figures depict non-imputed data. (a) Primary outcome was depressive symptom scores, as assessed by the EPDS at 3-months post-randomisation. (b) Secondary outcome was anxiety symptom

**B Anxiety Symptoms**

scores, as assessed by the GAD-7 at 3-months post-randomisation. Participants who did not complete any sessions had significantly higher (a) depressive and (b) anxiety scores at 3-months post-randomisation. p-value asterisks correspond to the two-sample t-test conducted to examine the difference in EPDS symptom scores between individuals who did and did not complete BA sessions, at 3-months.

**A Depressive Symptoms****B Anxiety Symptoms**

**Extended Data Fig. 5 | Comparing imputed depressive (EPDS) and anxiety (GAD-7) symptom scores among participants who did and did not complete BA sessions.** *Note.* Edinburgh Postnatal Depression Scale (EPDS), General Anxiety Disorder-7 (GAD-7); \* $p < 0.05$ ; \*\* $p < 0.01$ . The figures depict imputed data. (a) Primary outcome was depressive symptom scores, as assessed by the EPDS at 3-months post-randomisation. (b) Secondary outcome was anxiety symptom

scores, as assessed by the GAD-7 at 3-months post-randomisation. Participants who did not complete any sessions had significantly higher (a) depressive and (b) anxiety scores at 3-months post-randomisation. p-value asterisks correspond to the two-sample t-test conducted to examine the difference in EPDS symptom scores between individuals who did and did not complete BA sessions, at 3-months on imputed data.

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- |                                     |  |
|-------------------------------------|--|
| n/a                                 | Confirmed  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of all covariates tested   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated   |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Two years after publication, all individual participant de-identified data will be shared with researchers who submit a methodologically sound written proposal to: [summittrial@sinaihealth.ca](mailto:summittrial@sinaihealth.ca). Data will be available indefinitely. This timeline allows the study team to collect long-term study outcomes, and analyze and potentially publish the results of the secondary aims involving sustained outcomes at 12-months post-randomization as indicated our detailed Statistical Analysis Plan. Researchers requesting data will receive a response from the study team within three (3) business days.

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	Pregnant or postpartum adults (inclusive of gender identities and birthing persons) were enrolled in the study. Self-reported gender identity was collected with participants predominantly identifying as: cis-women (1168/1173, 99.57%). Gender based analyses were not performed due to data sparseness.
Reporting on race, ethnicity, or other socially relevant groupings	Self-reported race/ethnicity was collected as required by the funding agency and to address mental health disparities for vulnerable populations including racialized minorities. We used the funding agency's classification system: American Indian/Alaska Native, Asian, Black/African American, Hawaiian/Pacific Islander, White, Multi-race, Hispanic (Latino/Latina), Other, Prefer not to answer. Race/ethnicity was included as a covariate in analyses.
Population characteristics	Participants' mean age was 33.27 (95% CI=33.00-33.55) years, predominantly identifying as: cis-women (1168/1173, 99.57%); Black, Indigenous, or Persons of Colour (578/1226, 47.15%); and nulliparous (668/1226, 54.49%). Most participants (1051/1226, 85.73%) reported a history of depression or anxiety prior to pregnancy, and almost one-quarter (288/1226, 23.49%) were taking psychotropic medications at enrollment.
Recruitment	Between January 2020 and October 2023 participants were recruited through self, internal, or external referrals. Internal referrals were elicited from clinicians from site hospitals and satellite clinics (i.e., obstetrical, mental health, and family departments) who sent patient contact information directly to the research team. External referrals were received from clinicians at sites that were not affiliated with the trial. Finally, recruitment materials with contact information for the research team (i.e., brochures, posters) were available in clinics and interested individuals contacted the team for more information (self-referrals). The study was introduced to potential participants by either one of their clinical providers (for internal and external referrals) or a trained researcher assistant at their respective sites (for self-referrals). Self-selection bias was reduced in the current study because: (1) the large sample size; (2) there were significantly fewer self-referrals compared to other referral pathways; and (3) we found no evidence of statistically significant in baseline or outcome variables among participants who self-referred.
Ethics oversight	The study received ethical approvals from the following three institutional review boards (IRB): UNC Biomedical IRB (19-1786); Endeavour Health IRB (EH18-129); and Clinical Trials Ontario (1895).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Pragmatic, multi-site, non-inferiority, four-arm trial that tested the non-inferiority of NSP vs. specialist providers (SP) and telemedicine vs. in-person delivery of a brief, 6-8 session behavioural activation (BA) therapy for improving perinatal depression and anxiety symptoms. The predefined primary outcome was the EPDS total score at 3-months post-randomisation. All data included in the current manuscript are quantitative.
Research sample	Pregnant ( $\leq 36$ weeks) or postpartum (4-30 weeks) adult ( $\geq 18$ years; inclusive of gender identities and birthing persons) with a score $\geq 10$ on the Edinburgh Postnatal Depression Scale (EPDS). Participants' mean age was 33.27 (95% CI=33.00-33.55) years, predominantly identifying as: cis-women (1168/1173, 99.57%); Black, Indigenous, or Persons of Colour (578/1226, 47.15%); and nulliparous (668/1226, 54.49%). The study sample was fairly representative of the populations they were recruited from. One in five women experience depression or anxiety during the perinatal period and treatment is essential, however only 10% of affected perinatal patients receive psychotherapy.
Sampling strategy	A random stratified sampling procedure was used. Sample size: Using a non-inferiority margin of 10%, and an EPDS mean estimate of 7.93 (SD=4.68), 80% power, and alpha ( $\alpha$ )=0.05, the comparison of provider (non-specialist vs. specialist) required 431 participants in each of the two conditions. To account for 10% dropout, the sample size was inflated to N=958. As described in our detailed SAP and based on recent non-inferiority guidelines, we did not adjust our two primary hypotheses for multiplicity because they did not involve different endpoints (i.e., both hypotheses test total EPDS scores at 3-months post-randomisation). Specifically, when multiple hypotheses test a similar underlying outcome, no adjustment for multiplicity is required. As described above, and due to COVID-19, we increased the non-inferiority margin for the modality comparison (telemedicine vs. in-person) from 10% to 13% and used the same EPDS estimate, 80% power, and $\alpha=0.05$ . The comparison of telemedicine to in-person delivery required an additional 268 participants, yielding a target sample size of 1226. While the study is a multi-centre one, randomisation occurred within each individual site. Thus, no cluster randomisation was carried out and the sample size was not inflated to account for an intracluster

	correlation. The sample size calculation was run using PASS Version 12 (Power Analysis and Sample Size Software). NCSS, LLC. Kaysville, Utah, USA).
Data collection	Quantitative self-report data was collected through standardized REDCap™ databases reviewed by independent data staff to reduce the number of staff who were unmasked. All team members were masked to study arm allocation except for the data site coordinator, participants, treatment providers, and clinical leads. Participants were sent unique REDCap™ links to collect their self-report data.
Timing	Three-month follow-up data was collected from March 2020 through February 2024
Data exclusions	No data were excluded from the primary analyses
Non-participation	N=60 participants dropped out from the trial after being randomized for the following reasons: lack of interest, time or childcare, adverse event, preferred telemedicine, lost to follow-up, or unknown.
Randomization	Individual participants were randomised within REDCap™ to one of four study arms and randomisation was stratified by perinatal period and site. An independent biostatistician generated the randomisation sequence through computer-generated lists, with random blocks of four and eight, and stratified by perinatal period and site. Randomization sequence was concealed until all 3-month data were analysed.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration	ClinicalTrials.Gov: NCT04153864
Study protocol	Singla, D.R., et al. Scaling Up Maternal Mental healthcare by Increasing access to Treatment (SUMMIT) through non-specialist providers and telemedicine: a study protocol for a non-inferiority randomized controlled trial. <i>Trials</i> 22, 186 (2021)
Data collection	Between January 2020 and October 2023 participants were recruited in Chapel Hill (North Carolina), Evanston/Chicago (Illinois), and Toronto (Canada) through university-affiliated healthcare settings. Three-month follow-up data was collected via standardized REDCap™ databases, securely hosted on institutional servers at each of the three participating Hubs from March 2020 through February 2024.
Outcomes	All outcomes were assessed at 3-months post-randomization. The predefined primary outcome was depressive symptoms, as assessed by the Edinburgh Postnatal Depression Scale (EPDS) total score. The secondary outcome was anxiety symptoms, as assessed by the Generalized Anxiety Disorder (GAD-7) total score. Exploratory outcomes included in this manuscript were quality of life (as assessed through the World Health Organization Disability Assessment Schedule 2.0 and EQ5D-5-Level); client satisfaction (Client Satisfaction Questionnaire-8); therapeutic alliance (Working Alliance Inventory-Short Form); activation levels (Premium Abbreviated Activation Scale); and perceived social support (Multidimensional Scale of Perceived Social Support).