Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Lewis G, Marston L, Duffy L, et al. Maintenance or discontinuation of antidepressants in primary care. N Engl J Med 2021;385:1257-67. DOI: 10.1056/NEJMoa2106356

Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to Gemma Lewis PhD,* Louise Marston, PhD,* Larisa Duffy BSc, Nick Freemantle PhD, Simon Gilbody PhD, Rachael Hunter MSc, Tony Kendrick MD, David Kessler MD, Dee Mangin FRNZCGP, Michael King PhD, Paul Lanham BA, Michael Moore FRCGP, Irwin Nazareth PhD, Nicola Wiles PhD, Faye Bacon BSc, Molly Bird MSc, Sally Brabyn MSc, Alison Burns BSc, Caroline S Clarke PhD, Anna Hunt MSc, Jodi Pervin BSc, Glyn Lewis PhD. Trial of discontinuation vs maintenance of antidepressants in primary care

Table of Contents

Methods

Full exclusion criteria:

Bipolar disorder, psychotic illness, dementia, terminal illness, could not understand questionnaires in English, had contraindications to the medication or placebo ingredients, were taking monoamine oxidase inhibitors or were enrolled in another clinical trial. Women were excluded if they were pregnant, planning pregnancy or breast feeding.

We stopped recruiting people on citalopram before those on other antidepressants as this medication was exhausted before the end of recruitment. Likewise, 21 participants who were randomised to sertraline or sertraline discontinuation towards the end of recruitment did not receive the intended 12 months' medication due to limited supply (1 received 11 months, 18 received 10 months and 2 received 8 months' medication). These events did not differ by randomized group.

Baseline CIS-R

Due to a programming error, the maximum score on the panic section was 2 rather than the usual 4, that applied to all other sections, so the range was 0-18. The panic section was used to create the anxiety variable which was used in subgroup analyses.

Measurement and reliability and validity of primary outcome

When describing the longitudinal course of depression there has been a suggestion to distinguish remission, relapse, recovery and recurrence.[1] In particular, relapse (occurring within 6 months of an index episode) has been distinguished from recurrence. These

definitions are based upon a categorical approach to illness course that is rarely supported empirically nor by personal accounts of illness.[2]

As there was no existing simple and short structured interview that retrospectively assessed time to relapse for depression, we developed a modified retrospective version of the CIS-R (the rCIS-R) in which we used the original CIS-R sections relevant to depression (depression, depressive ideas, fatigue, concentration, sleep) but over a longer 12-week time period. The original CISR has been judged as a valid and reliable assessment of depression and we used the same questions in the rCIS-R.[3,4] The rCIS-R is a fully structured assessment that was self-administered on a computer. It asks the initial mandatory questions from the original CIS-R but asks patients if they had experienced depressive symptoms over the past 12 weeks (rather than the past week). If participants answer positively to the mandatory questions, the subsequent questions in each section ask about the worst week during the past 12 weeks. The rCIS-R was completed at every in-person follow-up except 6 weeks. Participants were asked to identify the number of weeks since the previous assessment when these symptoms began, in order to estimate date of onset. We also used the rCIS-R to identify participants who had met criteria for an ICD-10 depressive episode and present these as posthoc analyses.

We assessed test-retest reliability of the rCIS-R. Of the 478 participants recruited to the trial, 396 completed the rCIS-R twice at one of the follow-ups, separated by about 30 minutes completing the other questionnaires. We found excellent test-retest agreement for the rCIS-R definition of relapse (kappa 0.84 (95%CI 0.71 to 0.97), for the individual sections and for time of relapse (Intraclass Correlation Coefficient 0.94 (95%CI 0.92 to 0.95)).

We also assessed construct validity against a global rating question about worsening of mood, patients stopping their study medication, and Patient Health Questionnaire (PHQ-9) scores. The odds of feeling worse (compared to the same or better) were 5.55 (95% CI: 3.44 to 8.95) times greater in those who relapsed compared with those who did not at 12 weeks. Of those who relapsed, 20% stopped study medication and returned to their usual antidepressant by 12 weeks, compared to 3% of those who did not relapse. There was a strong correlation (r=0.73) between the PHQ-9 and rCIS-R at 12 weeks.

Adherence

We used a 5-item self-report measure of adherence as used in the COBALT and MIR trials.[5,6] adapted from a four-item version. The original four-item version contained the following questions (with yes/no response options) were: do you ever forget to take your medicine?; are you careless at times about taking your medicine?; when you feel better do you sometimes stop taking your medicine?; sometimes if you feel worse when you take the medicine, do you stop taking it? Given the relatively long half-life of antidepressants, individuals who had forgotten to take one or two tablets were not excluded. We established this by adding the question "In the last 4 weeks did you miss 2 or more days of your medication in a row?" Our criteria therefore defined people as adherent if (1) they scored zero on all four questions (2) they scored one and said "no" to the extra question (3) they scored 2 because of the "forget" and "careless" questions and said "no" to the extra question.

Number of depressive episodes

To determine eligibility based on number of prior depressive episodes at the telephone screening phase, patients were asked:

Apart from your current antidepressants, have you been offered antidepressant medication from your GP before?

OR

Before the current treatment, have you had any problems with depression?

In order for a patient to be eligible, they need to answer yes to at least one of the above questions, or have been taking antidepressants for at least 2 years. We found that participants had difficulty remembering discrete episodes so interpreted continuous antidepressant treatment of over 2 years as equivalent to a previous episode.

At the baseline assessment, participants were also asked how many previous episodes they had experienced with the following question (which was used as a pre-specified subgroup variable):

"Many people report having periods of feeling sad, low or depressed, while they are fairly well in between. How many periods of feeling sad, low or depressed have you had?" With responses: 1,2,3,4,5 more than 5.

Secondary outcomes

Global changes to mood were assessed with the question, “compared to when we last saw you, how have your moods and feelings changed?” Responses were “I feel a lot better”, “I

feel slightly better", "I feel about the same", "I feel slightly worse", or "I feel a lot worse." We created a binary variable, feeling the same or better versus feeling worse.

To measure withdrawal symptoms, we used a modified self-report version of the 43-item Discontinuation-Emergent Signs and Symptoms (DESS) scale created by Rosenbaum and colleagues[7] to ask about signs and symptoms associated with discontinuation of SSRI treatment. The original 43-item list was based on signs and symptoms reported in the literature. In discussion with a group of patients who had experienced withdrawal symptoms, we chose 14 of the most commonly endorsed symptoms from the Rosenbaum study. We added an additional 15th item on brain zaps, after our involvement work with patients informed us that they were commonly experienced after SSRI discontinuation.

Serious adverse events

As this was a phase IV trial of licensed medications used within their licensed indication, with a well-established safety profile, adverse events (AEs) were not recorded apart from those AEs of special interest. AEs of special interest were assessed at each follow-up using the Toronto side effects scale and DESS withdrawal symptoms scale.

All serious adverse events (SAEs) were recorded by researchers using an SAE recording and reporting log created by the sponsor (the Joint Research Officer, UCL). The primary care clinical trials unit (PRIMENT) was also given some sponsor duties. The principal investigator at each site rated the severity, outcome and relatedness of the SAE to the study medication, using the recording and reporting log (see below for the rating scales). All SAEs were reported to the Sponsor within 24 hours of the researchers knowing of the event. The Chief Investigator and trial manager were also informed.

The rating scales were:

Serious type: 1=Resulted in Death, 2=life Threatening, 3=required inpatient or prolonged existing hospitalisation, 4=resulted in persistent or significant disability/incapacity, 5=resulted in congenital anomaly/birth defect, 6= Important Medical Event.

Causal relationship: a= definitely, b=probably, c=possibly, d=unlikely, e= not related, f=not assessable

Severity grade: 1= Mild, 2 = Moderate, 3 = Severe

Outcome: 1= Resolved, 2 = Resolved with sequelae, 3 = Unresolved, 4= Worsening, 5 = Fatal, 6= not assessable

Minimisation algorithm

Minimisation was conducted by Sealed Envelope and the minimisation variables were site, medication and CISR score. The median CISR score was calculated at the time of each new randomisation. Then the numbers in <=median and >median were assessed and allocated on 70:30 biased coin basis in the direction of minimising the difference of the number of subjects in each group. The first six participants who were enrolled in the trial underwent randomisation with the use of a simple 1:1 randomisation method. As the median CIS-R was dynamic, and there was not a median before the study commenced, this allowed us to establish a median. As subsequent participants were included in the study, the median changed over time.

Statistical analyses

Continuous secondary outcomes

The precise statistical model used to analyse each continuous secondary outcome was as follows. We used linear mixed models for repeated measures with two levels, with higher level participant. The repeated measures outcome included two data points, one for baseline and one for follow-up. The baseline values for the randomised group variable were all set to 0 and at follow up the maintenance group was coded 0 and the discontinuation group 1. This method takes account of the baseline value of the outcome.[8] The model included the randomised group variable and a dichotomous variable for baseline or follow-up as fixed effects. We ran a separate mixed model for each follow-up time-point. Analyses for the 12, 26, 39 and 52-week time-points were pre-specified.

As a pre-specified sensitivity analysis of the secondary outcomes, we treated all the follow-up points (6, 12, 26, 39 and 52 weeks) as a repeated measures outcome (Table S3). As a further posthoc analysis of the secondary outcomes, we used the 6-week outcome, and the same model specification as described above (Table S3).

Predictors of missingness

For secondary outcomes, except time to stopping study medication, and time points, we conducted sensitivity analyses including predictors of missingness identified using univariable logistic regression. For this, the outcomes were whether the measure was missing or not at each time point separately. Baseline variables were considered as possible

predictors of missingness. Those that were associated and were statistically significant at the 5% level for each outcome and time point were adjusted for that outcome and time point using similar models to the main secondary outcome models. If no variables were statistically significant we did not perform adjustments. For the models including data from all time points, all the baseline predictors of missingness for the given outcome were included in the model. For the SF-12, this analysis was only carried out once for each time point since the same questions were used to calculate the physical and mental component scores; thus the same predictors of missingness were included for these outcomes. There was no missing data for the outcome time to stopping ANTLER medication.

Results for pre-specified analyses

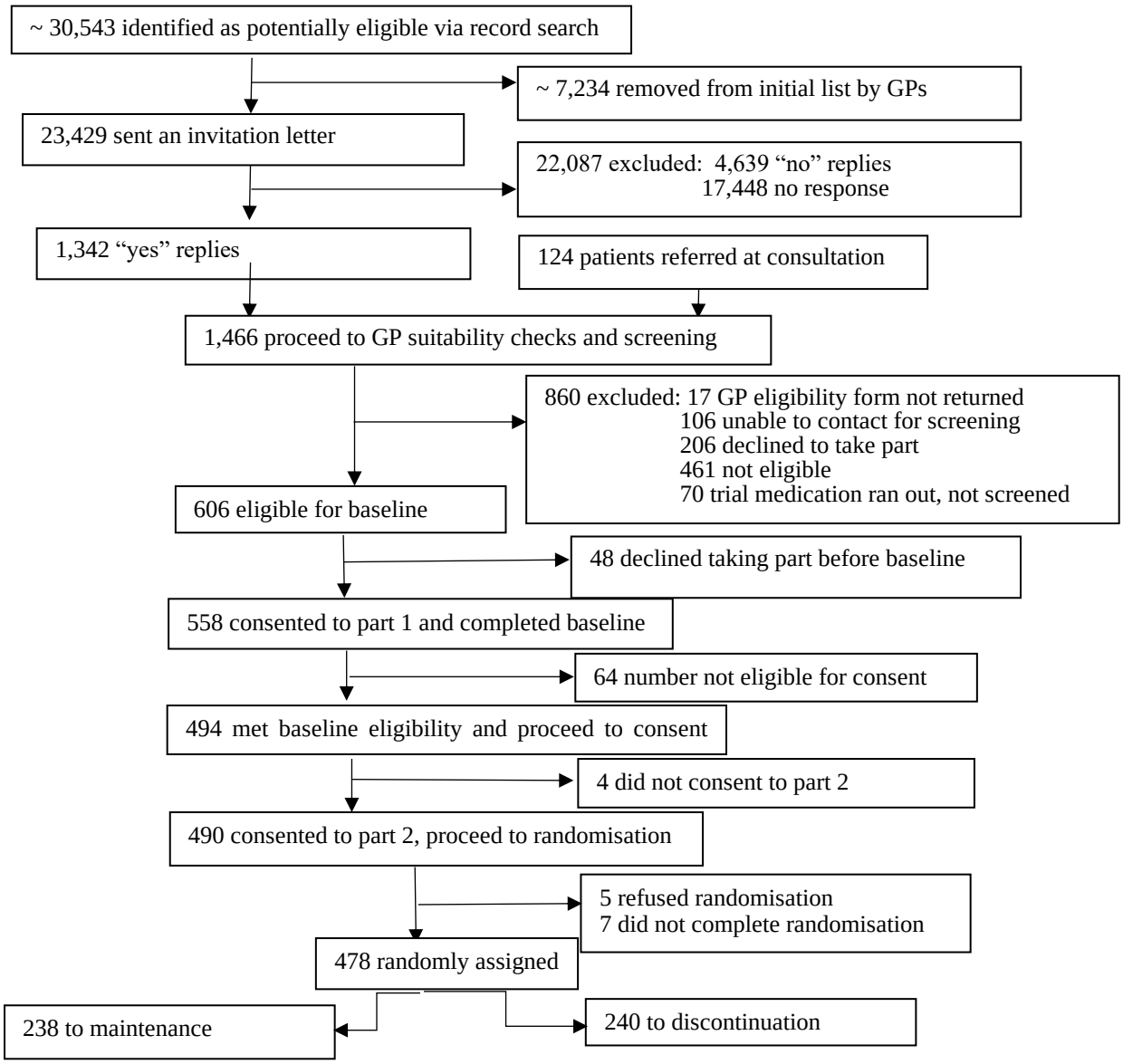

Figure S1: Consort diagram showing flow of participants through the study up to randomisation. See main manuscript for flow of participants for follow-up.

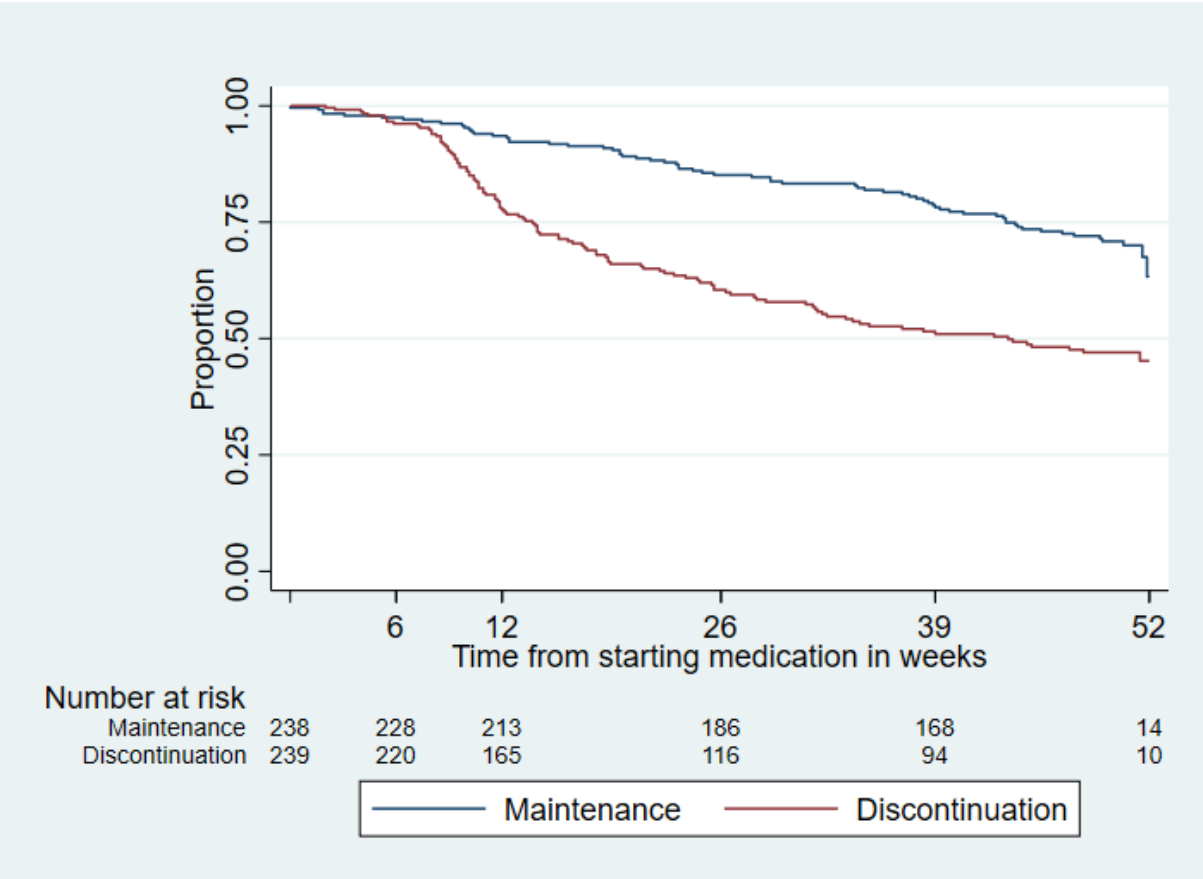

Figure S2: Kaplan Meier plot for time to stopping study medication as secondary outcome.

Table S1: Additional baseline characteristics

Characteristic	Maintenance (n=238)		Discontinuation (n=240)	
Courses of antidepressants in the past n/N %				
0	102/238	43	92/239	38
1	40/238	17	39/239	16
2 or more	96/238	40	108/239	45
In current antidepressant treatment, time since improvement in depression? n/N %				
Up to a year ago	84/238	35	82/239	34
2 to 4 years ago	75/238	32	77/239	32
5 or more years ago	75/238	32	76/239	32
Never	4/238	2	4/239	2
Attempted suicide in the past n/N %	31/238	13	16/239	7
At least one symptom on the Toronto Side Effects Scale n/N %	217/235	92	218/239	91

Table S2: Sensitivity analysis of the primary outcome

Outcome	Hazard ratio	95% CI
Time to first depression relapse*	2.04	(1.55, 2.68)
Time to first depression relapse including minimisation variables N=468†	2.07	(1.57, 2.72)
Time to first depression relapse, good outcome in maintenance, bad outcome in placebo N=478‡	2.12	(1.61, 2.78)
Time to first depression relapse, those censored in the maintenance group had a relapse N=468	1.61	(1.25, 2.09)
Time to first depression relapse, those censored in the withdrawal group had a relapse N=468	2.71	(2.09, 3.52)

*Including dichotomous above or below the median depression at baseline
†Study centre (4 groups), antidepressant medication (4 groups) and severity of depressive symptoms (above or below the median, 2 groups).
‡Antidepressant group: censored at the date of last follow-up or withdrawal (good outcome, no relapse). Placebo group: relapse on the day before last follow-up or withdrawal (bad outcome, relapse

Table S3: Secondary outcomes including predictors of missingness and 6-week outcomes.

	Maintenance		Discontinuation		Main models		With predictors of missingness	
	Mean	(SD)	Mean	(SD)	Estimate	95% CI	Estimate	95%CI
PHQ-9 (difference in means)*								
Baseline	3.9	(3.5)	3.8	(3.6)				
6 weeks N=478	4.1	(3.8)	4.4	(4.0)	0.30	(-0.26, 0.87)	0.31	(-0.25, 0.88)
12 weeks N=477	4.1	(3.8)	6.3	(5.1)	2.16	(1.47, 2.84)		
26 weeks N=477	4.2	(3.7)	5.0	(4.6)	0.72	(0.02, 1.42)	0.80	(0.05, 1.56)
39 weeks N=477	3.8	(3.9)	4.4	(4.2)	0.55	(-0.14, 1.24)	0.64	(-0.05, 1.33)
52 weeks N=477	3.7	(3.7)	4.0	(4.5)	0.38	(-0.32, 1.07)	0.38	(-0.32, 1.08)
Over all time points N=478					0.84	(0.38, 1.29)	1.02	(0.65, 1.39)
GAD-7 (difference in means)†								
Baseline	3.2	(3.1)	2.8	(3.0)				
6 weeks N=478	3.2	(3.6)	3.6	(3.7)	0.50	(-0.03, 1.03)	0.50	(-0.03, 1.03)
12 weeks N=477	3.1	(3.3)	5.3	(4.6)	2.40	(1.81, 2.99)		
26 weeks N=477	3.4	(3.8)	4.1	(4.4)	0.79	(0.13, 1.45)	1.13	(0.42, 1.85)
39 weeks N=477	2.9	(3.5)	3.8	(4.1)	0.99	(0.36, 1.62)	1.03	(0.41, 1.67)
52 weeks N=477	3.0	(3.7)	3.1	3.0	0.27	(-0.36, 0.89)	0.28	(-0.34, 0.91)
Over all time points N=478					1.00	(0.58, 1.42)	1.19	(0.74, 1.64)
Modified Toronto Side Effects Scale (difference in means)‡								
Baseline	4.2	(2.7)	3.7	(2.7)				
6 weeks N=478	3.7	(2.7)	4.0	(2.8)	0.53	(0.13, 0.92)	0.54	(0.16, 0.92)
12 weeks N=477	4.2	(2.9)	4.6	(3.0)	0.68	(0.25, 1.11)		
26 weeks N=477	4.0	(2.6)	3.9	(2.8)	0.20	(-0.26, 0.66)	0.25	(-0.28, 0.77)
39 weeks N=476	3.8	(2.5)	3.7	(2.6)	0.16	(-0.30, 0.62)	0.21	(-0.24, 0.66)
52 weeks N=475	3.7	(2.6)	3.5	(2.8)	0.04	(-0.41, 0.49)	0.09	(-0.34, 0.53)
Over all time points N=478					0.36	(0.06, 0.65)	0.47	(0.16, 0.77)
Number of new or worsening symptoms using modified DESS (difference in means)§								
Baseline	1.0	(1.4)	0.6	(1.0)				
6 weeks N=478	1.1	(2.0)	1.5	(2.5)	0.51	(0.17, 0.84)	0.51	(0.18, 0.84)

12 weeks N=478	1.3	(2.4)	3.1	(3.5)	1.87	(1.46, 2.28)		
26 weeks N=478	1.4	(2.3)	1.9	(2.9)	0.50	(0.12, 0.89)	0.53	(0.12, 0.95)
39 weeks N=478	0.8	(1.6)	1.7	(2.7)	0.94	(0.60, 1.28)	0.95	(0.61, 1.29)
52 weeks N=478	0.8	(1.8)	1.1	(2.5)	0.32	(-0.02, 0.65)	0.31	(-0.02, 0.65)
Over all time points N=478					0.86	(0.62, 1.11)	0.96	(0.70, 1.22)
SF-12 physical (difference in means)‖								
Baseline	48	(11)	50	(9)				
12 weeks N=476	48	(10)	50	(9)	0.44	(-0.91, 1.78)		
26 weeks N=476	48	(10)	49	(10)	0.15	(-1.33, 1.62)	0.16	(-1.46, 1.77)
39 weeks N=476	48	(11)	51	(10)	1.49	(-0.06, 3.04)	1.65	(0.11, 3.19)
52 weeks N=476	49	(10)	49	(11)	-0.59	(-2.09, 0.92)	-0.44	(-1.94, 1.05)
Over all time points N=476					0.44	(-0.60, 1.48)	0.27	(-0.84, 1.37)
SF-12 mental (difference in means)‖								
Baseline	47	(9)	48	(9)				
12 weeks N=476	46	(10)	41	(11)	-4.86	(-6.44, -3.29)		
26 weeks N=476	46	(11)	44	(11)	-2.56	(-4.35, -0.77)	-2.91	(-4.78, -1.04)
39 weeks N=476	48	(10)	45	(11)	-3.07	(-4.84, -1.31)	-3.05	(-4.80, -1.30)
52 weeks N=476	47	(10)	46	(11)	-1.59	(-3.43, 0.25)	-1.68	(-3.51, 0.15)
Over all time points N=476					-3.02	(-4.23, -1.81)	-3.13	(-4.39, 1.88)
Global rating Question (OR)	n/N	%	n/N	%				
Baseline N=476								
Feeling the same or better	224/237	95	230/239	96				
Feeling worse	13/237	5	9/239	4				
6 weeks N=446								
Feeling the same or better	182/223	82	182/223	82	1.00		1.00	
Feeling worse	41/223	18	41/223	18	1.00	(0.62, 1.61)	1.03	(0.63, 1.66)
12 weeks N=444								
Feeling the same or better	180/228	79	122/216	56	1.00			
Feeling worse	48/228	21	94/216	44	2.88	(1.90, 4.38)		
26 weeks N=403								
Feeling the same or better	164/210	78	151/193	78	1.00		1.00	

Feeling worse	46/210	22	42/193	22	0.99	(0.62, 1.59)	1.08	(0.63, 1.85)
39 weeks N=396								
Feeling the same or better	185/212	87	153/184	83	1.00		1.00	
Feeling worse	27/212	13	31/184	17	1.39	(0.79, 2.43)	1.39	(0.79, 2.43)
52 weeks N=391								
Feeling the same or better	181/210	86	154/181	85	1.00		1.00	
Feeling worse	29/210	14	27/181	15	1.09	(0.62, 1.93)	1.12	(0.63, 1.98)
Time to stopping study medication (HR) n/N (%) stopped ANTLER medication	68/225	(30)	111/230	(48)	2.28	(1.68, 3.08)		

Missing data: Additional analyses were performed only with statistically significant predictors of missingness from the individual time points for the given outcome so when there were no significant predictors then no further adjustments were made. There were no significant predictors of missingness at 12 weeks for any outcome.

* PHQ-9 - Patient Health Questionnaire 9, range 0 to 27. Predictors of missingness: 6 weeks – site, 12 weeks - none, 26 weeks – age at randomization, able to replace worn out furniture, able to buy new clothes, stop taking medication if feeling worse, 39 weeks – age at randomization, SF12 physical component score at baseline, 52 weeks – age at randomization, site

† GAD-7 – Generalised anxiety disorder 7, range 0 to 21. Predictors of missingness: 6 weeks – site, 12 weeks – none, 26 weeks – age at randomization, able to replace worn out furniture, stop taking medication if feeling worse, 39 weeks - age at randomization, SF12 physical component score at baseline, 52 weeks – age at randomization, site.

‡ Modified Toronto Side Effects Scale (count of side effects), range 0 to 13. Predictors of missingness: 6 weeks – site, CIS-R score at randomization, 12 weeks – none, 26 weeks - able to buy new clothes, 39 weeks – age at randomisation, SF12 physical component score at baseline, 52 weeks – age at randomisation, CIS-R score at randomization.

§ Modified DESS - Discontinuation-emergent signs and symptoms (DESS) checklist, range 0 to 15. Predictors of missingness: 6 weeks – site, 12 weeks – none, 26 weeks – age at randomisation, able to buy new clothes, 39 weeks – age at randomization, 52 weeks – age at randomization, site.

‖ SF-12 – Short form 12 questions, range 0 to 100. Predictors of missingness: 12 weeks – none, 26 weeks – age at randomization, able to buy new clothes, CIS-R score at randomization, stop taking medication if feeling worse, 39 weeks – age at randomization, 52 weeks – age at randomization, site.

OR = odds ratio, HR = hazard ratio

Table S4: Analyses according to pre-specified subgroups for the primary outcome

Subgroups	HR	95% CI
Sertraline	2.41	(1.20, 4.82)
Citalopram	2.14	(1.44, 3.18)
Fluoxetine	1.70	(1.05, 2.76)
CIS-R depression score below the median*	2.34	(1.56, 3.52)
CIS-R depression score above the median*	1.80	(1.24, 2.61)
CIS-R anxiety score below the median†	2.08	(1.49, 2.89)
CIS-R anxiety score above the median†	1.95	(1.20, 3.18)
2 previous episodes of depression	0.96	(0.26, 3.62)
3 or more previous episodes of depression	2.15	(1.63, 2.85)
Age at onset of depression below median‡	2.66	(1.84, 3.85)
Age at onset of depression above median‡	1.43	(0.94, 2.16)

*CIS-R depression dichotomized at <3 versus 3+.
†CIS-R anxiety dichotomized at <2 versus 2+
‡Age when became aware of depression dichotomized at <32 versus 32+
HR = hazard ratio

Table S5: Analyses according to pre-specified subgroups for PHQ-9*

Subgroups	Coefficient	95% CI
6 weeks		
Sertraline	0.74	(-0.76, 2.25)
Citalopram	0.56	(-0.28, 1.41)
Fluoxetine	-0.10	(-1.00, 0.80)
CIS-R depression score below the median†	0.29	(-0.23, 0.81)
CIS-R depression score above the median†	0.45	(-0.50, 1.40)
CIS-R anxiety score below the median‡	0.49	(-0.09, 1.07)
CIS-R anxiety score above the median‡	0.01	(-1.19, 1.21)
2 previous episodes of depression	1.23	(-0.21, 2.67)
3 or more previous episodes of depression	0.27	(-0.32, 0.87)
Age when became aware of depression below median§	0.52	(-0.30, 1.34)
Age when became aware of depression above median§	0.06	(-0.71, 0.83)
12 weeks		
Sertraline	4.74	(2.93, 6.55)
Citalopram	1.61	(0.67, 2.55)
Fluoxetine	1.73	(0.60, 2.86)
CIS-R depression score below the median	2.25	(1.45, 3.04)
CIS-R depression score above the median	2.16	(1.14, 3.17)
CIS-R anxiety score below the median	2.05	(1.29, 2.81)
CIS-R anxiety score above the median	2.49	(1.19, 3.80)
2 previous episodes of depression	1.69	(0.06, 3.31)
3 or more previous episodes of depression	2.22	(1.50, 2.95)
Age when became aware of depression below median	3.01	(2.02, 3.99)
Age when became aware of depression above median	1.14	(0.23, 2.06)
26 weeks		
Sertraline	3.73	(1.84, 5.61)
Citalopram	0.61	(-0.42, 1.65)
Fluoxetine	-0.41	(-1.48, 0.67)
CIS-R depression score below the median	0.79	(-0.05, 1.63)
CIS-R depression score above the median	0.66	(-0.33,1.64)
CIS-R anxiety score below the median	1.09	(0.31, 1.87)
CIS-R anxiety score above the median	0.06	(-1.22, 1.33)
2 previous episodes of depression	0.57	(-0.84, 1.98)
3 or more previous episodes of depression	0.79	(0.05, 1.52)
Age when became aware of depression below median	1.34	(0.35, 2.34)
Age when became aware of depression above median	0.07	(-0.88, 1.03)
39 weeks		
Sertraline	1.23	(-0.57, 3.03)
Citalopram	0.47	(-0.55, 1.48)
Fluoxetine	0.02	(-1.09, 1.14)
CIS-R depression score below the median	0.33	(-0.44, 1.10)

CIS-R depression score above the median	0.82	(-0.20, 1.85)
CIS-R anxiety score below the median	0.45	(-0.30, 1.21)
CIS-R anxiety score above the median	0.92	(-0.41, 2.24)
2 previous episodes of depression	-1.70	(-2.90, -0.49)
3 or more previous episodes of depression	0.77	(0.04, 1.50)
Age when became aware of depression below median	1.06	(0.13, 2.00)
Age when became aware of depression above median	0.01	(-1.01, 1.03)
52 weeks		
Sertraline	1.72	(-0.10, 3.54)
Citalopram	-0.04	(-1.07, 0.99)
Fluoxetine	-0.01	(-1.07, 1.05)
CIS-R depression score below the median	0.13	(-0.69, 0.95)
CIS-R depression score above the median	0.67	(-0.34, 1.67)
CIS-R anxiety score below the median	0.48	(-0.30, 1.26)
CIS-R anxiety score above the median	0.22	(-1.09, 1.52)
2 previous episodes of depression	-0.89	(-2.53, 0.75)
3 or more previous episodes of depression	0.48	(-0.26, 1.22)
Age when became aware of depression below median	0.28	(-0.65, 1.21)
Age when became aware of depression above median	0.42	(-0.63, 1.48)
All time points		
Sertraline	2.26	(0.97, 3.55)
Citalopram	0.72	(0.07, 1.37)
Fluoxetine	0.22	(-0.48, 0.93)
CIS-R depression score below the median	0.74	(0.21, 1.27)
CIS-R depression score above the median	1.00	(0.34, 1.67)
CIS-R anxiety score below the median	0.88	(0.37, 1.40)
CIS-R anxiety score above the median	0.84	(-0.01, 1.70)
2 previous episodes of depression	0.62	(-0.55, 1.78)
3 or more previous episodes of depression	0.90	(0.43, 1.38)
Age when became aware of depression below median	1.27	(0.62, 1.92)
Age when became aware of depression above median	0.35	(-0.28, 0.98)

*PHQ-9 - Patient Health Questionnaire 9, range 0 to 27.
†CIS-R depression dichotomized at <3 versus 3+.
‡CIS-R anxiety dichotomized at <2 versus 2+
§Age when became aware of depression dichotomized at <32 versus 32+

Table S6: Analyses according to pre-specified subgroups for GAD-7*

Subgroup	Coefficient	95% CI
6 weeks		
Sertraline	0.11	(-1.32, 1.54)
Citalopram	0.99	(0.24, 1.74)
Fluoxetine	-0.24	(-1.15, 0.67)
CIS-R depression score below the median†	0.54	(0.00, 1.08)
CIS-R depression score above the median†	0.48	(-0.42, 1.39)
CIS-R anxiety score below the median‡	1.03	(0.52, 1.55)
CIS-R anxiety score above the median‡	-0.52	(-1.67, 0.63)
2 previous episodes of depression	0.45	(-0.64, 1.54)
3 or more previous episodes of depression	0.56	(-0.01, 1.13)
Age when became aware of depression below median§	0.83	(0.04, 1.62)
Age when became aware of depression above median§	0.13	(-0.58, 0.83)
12 weeks		
Sertraline	3.66	(2.18, 5.15)
Citalopram	2.37	(1.59, 3.15)
Fluoxetine	1.63	(0.58, 2.68)
CIS-R depression score below the median	1.97	(1.29, 2.66)
CIS-R depression score above the median	2.84	(1.92, 3.76)
CIS-R anxiety score below the median	2.37	(1.72, 3.01)
CIS-R anxiety score above the median	2.53	(1.42, 3.65)
2 previous episodes of depression	0.17	(-1.18, 1.51)
3 or more previous episodes of depression	2.56	(1.93, 3.19)
Age when became aware of depression below median	3.26	(2.39, 4.14)
Age when became aware of depression above median	1.32	(0.57, 2.08)
26 weeks		
Sertraline	1.96	(0.00, 3.91)
Citalopram	0.70	(-0.22, 1.63)
Fluoxetine	0.30	(-0.77, 1.38)
CIS-R depression score below the median	0.93	(0.14, 1.73)
CIS-R depression score above the median	0.59	(-0.41, 1.59)
CIS-R anxiety score below the median	1.02	(0.31, 1.74)
CIS-R anxiety score above the median	0.38	(-0.85, 1.60)
2 previous episodes of depression	-0.85	(-2.83, 1.12)
3 or more previous episodes of depression	0.96	(0.27, 1.65)
Age when became aware of depression below median	0.98	(0.04, 1.92)
Age when became aware of depression above median	0.61	(-0.31, 1.52)
39 weeks		
Sertraline	1.37	(-0.29, 3.04)
Citalopram	0.82	(-0.10, 1.73)
Fluoxetine	0.75	(-0.24, 1.75)
CIS-R depression score below the median	0.30	(-0.46, 1.05)

CIS-R depression score above the median	1.68	(0.73, 2.63)
CIS-R anxiety score below the median	0.75	(0.11, 1.40)
CIS-R anxiety score above the median	1.52	(0.25, 2.80)
2 previous episodes of depression	-1.19	(-2.96, 0.57)
3 or more previous episodes of depression	1.18	(0.52, 1.85)
Age when became aware of depression below median	1.33	(0.50, 2.16)
Age when became aware of depression above median	0.66	(-0.30, 1.62)
52 weeks		
Sertraline	0.81	(-0.52, 2.15)
Citalopram	0.35	(-0.59, 1.28)
Fluoxetine	-0.54	(-1.52, 0.43)
CIS-R depression score below the median	-0.21	(-0.97, 0.55)
CIS-R depression score above the median	0.74	(-0.20, 1.67)
CIS-R anxiety score below the median	-0.04	(-0.72, 0.63)
CIS-R anxiety score above the median	0.96	(-0.19, 2.10)
2 previous episodes of depression	-1.67	(-3.20, -0.14)
3 or more previous episodes of depression	0.41	(-0.25, 1.07)
Age when became aware of depression below median	-0.04	(-0.90, 0.82)
Age when became aware of depression above median	0.61	(-0.30, 1.52)
All time points		
Sertraline	1.56	(0.38, 2.73)
Citalopram	1.12	(0.53, 1.71)
Fluoxetine	0.30	(-0.38, 0.99)
CIS-R depression score below the median	0.69	(0.20, 1.19)
CIS-R depression score above the median	1.35	(0.70, 1.99)
CIS-R anxiety score below the median	1.04	(0.58, 1.50)
CIS-R anxiety score above the median	0.98	(0.18, 1.78)
2 previous episodes of depression	-0.38	(-1.49, 0.73)
3 or more previous episodes of depression	1.14	(0.69, 1.58)
Age when became aware of depression below median	1.31	(0.71, 1.90)
Age when became aware of depression above median	0.64	(0.05, 1.22)

*GAD-7 – Generalised anxiety disorder 7, range 0 to 21
†CIS-R depression dichotomized at <3 versus 3+.
‡CIS-R anxiety dichotomized at <2 versus 2+
§Age when became aware of depression dichotomized at <32 versus 32+

Table S7: Analyses according to pre-specified subgroups for Global Rating Questionnaire (feeling worse)

Subgroup	OR	95% CI
6 weeks		
Sertraline	1.07	(0.36, 3.17)
Citalopram	1.43	(0.66, 3.10)
Fluoxetine	0.62	(0.28, 1.38)
CIS-R depression score below the median*	1.21	(0.59, 2.48)
CIS-R depression score above the median*	0.87	(0.45, 1.67)
CIS-R anxiety score below the median†	1.12	(0.61, 2.06)
CIS-R anxiety score above the median†	0.86	(0.39, 1.90)
2 previous episodes of depression	§	
3 or more previous episodes of depression	1.01	(0.62, 1.64)
Age when became aware of depression below median‡	1.22	(0.67, 2.22)
Age when became aware of depression above median‡	0.66	(0.28, 1.54)
12 weeks		
Sertraline	3.88	(1.37, 10.98)
Citalopram	3.36	(1.80, 6.27)
Fluoxetine	1.96	(0.97, 3.97)
CIS-R depression score below the median	3.53	(1.97, 6.33)
CIS-R depression score above the median	2.30	(1.27, 4.19)
CIS-R anxiety score below the median	2.65	(1.59, 4.42)
CIS-R anxiety score above the median	3.54	(1.71, 7.31)
2 previous episodes of depression	2.12	(0.43, 10.52)
3 or more previous episodes of depression	2.93	(1.90, 4.51)
Age when became aware of depression below median	3.01	(1.71, 5.28)
Age when became aware of depression above median	2.67	(1.43, 4.98)
26 weeks		
Sertraline	1.52	(0.51, 4.53)
Citalopram	0.90	(0.46, 1.76)
Fluoxetine	1.00	(0.40, 2.49)
CIS-R depression score below the median	1.53	(0.79, 2.95)
CIS-R depression score above the median	0.62	(0.31, 1.24)
CIS-R anxiety score below the median	1.25	(0.69, 2.27)
CIS-R anxiety score above the median	0.68	(0.31, 1.52)
2 previous episodes of depression	0.36	(0.03, 4.50)
3 or more previous episodes of depression	1.06	(0.65, 1.71)
Age when became aware of depression below median	1.15	(0.61, 2.18)
Age when became aware of depression above median	0.84	(0.41, 1.70)
39 weeks		
Sertraline	1.18	(0.28, 4.92)

Citalopram	1.49	(0.66, 3.39)
Fluoxetine	0.98	(0.37, 2.60)
CIS-R depression score below the median	1.45	(0.64, 3.27)
CIS-R depression score above the median	1.35	(0.63, 2.92)
CIS-R anxiety score below the median	1.96	(0.94, 4.07)
CIS-R anxiety score above the median	0.86	(0.34, 2.16)
2 previous episodes of depression	0.46	(0.06, 3.35)
3 or more previous episodes of depression	1.53	(0.85, 2.75)
Age when became aware of depression below median	1.56	(0.70, 3.47)
Age when became aware of depression above median	1.28	(0.58, 2.80)
52 weeks		
Sertraline	1.20	(0.29, 5.02)
Citalopram	1.28	(0.55, 2.97)
Fluoxetine	0.66	(0.24, 1.82)
CIS-R depression score below the median	1.02	(0.44, 2.36)
CIS-R depression score above the median	1.18	(0.55, 2.56)
CIS-R anxiety score below the median	1.04	(0.53, 2.05)
CIS-R anxiety score above the median	1.22	(0.44, 3.39)
2 previous episodes of depression	0.85	(0.05, 15.16)
3 or more previous episodes of depression	1.08	(0.60, 1.93)
Age when became aware of depression below median	1.30	(0.59, 2.90)
Age when became aware of depression above median	0.84	(0.37, 1.94)

*CIS-R depression dichotomized at <3 versus 3+.
†CIS-R anxiety dichotomized at <2 versus 2+
‡Age when became aware of depression dichotomized at <32 versus 32+
§Perfect prediction
OR=Odds ratio

Results for posthoc analyses

Most primary care surgeries were large and in urban areas with deprivation scores at the lower to moderate end of the index of multiple deprivation (Table S8). Compared to those who did not participate, trial participants were older and slightly more were female (Table S9). At baseline, 18 (4%) participants exceeded the recommended score of 10 or more[9] on the GAD-7.[9] Participants exceeding GAD-7 cut-off scores were well balanced according to randomised group.

For the binary outcome of relapse, the number needed to harm was 6 (95% CI 3, 19). So, if six (95% CI 3, 19) people stopped their medication, one would experience a relapse who would not have experienced a relapse if they had remained on maintenance treatment. Outcomes for each group are shown in Figure 2 of the supplement, according to relapse status. By the end of the study, of those who in the discontinuation group, 49/134 (37%; 95% CI 28%, 45%) remained on study medication (i.e. placebo), 71/134 (53%; 95% CI 44%, 62%) returned to a known antidepressant and 14/134 (10%; 95% CI 6%, 17%) were not on any antidepressant. Of those who relapsed in the maintenance group, 46/89 (52%; 95% CI 41%, 62%) remained on study medication, 32/89 (36%; 26%, 47%) returned to a known antidepressant and 11/89 (12%; 95% CI 6%, 21%) were not on any antidepressant. In the discontinuation group, 89/204 (44%; 95% CI 37%, 51%) and 201/225 (89%; 95% CI 85%, 93%) in the maintenance group were taking an antidepressant at the end of the trial.

Fifty-nine percent (141/240) were unblinded in the discontinuation group either through withdrawal (24%; 58/240) from the study or emergency code break (35%; 83/240). This was much lower in the maintenance group (29%; 68/236, withdrawal 11%; 25/236, emergency code break 18%; 43/236). Over the course of the study, 71% (162/228) in the discontinuation

group and 47% (108/232) in the maintenance group correctly guessed their randomised group at any time before being unblinded. This finding is consistent with prior studies[10,11] suggesting that patients can distinguish placebo from active treatment. In our study, participants may have guessed correctly because they experienced relapse or withdrawal symptoms. In principle this could influence the outcome, but it might also occur after the outcome and as a result of relapse.

When secondary outcome scores were log transformed, there was strong evidence that PHQ-9 and GAD-7 scores were higher in the discontinuation compared with the maintenance group at 12 weeks (adjusted proportional difference PHQ-9 1.41, 95% CI 1.21, 1.64; GAD-7 1.54, 95% CI 1.32, 1.79, Table S10). PHQ-9 and GAD-7 scores were 41% and 54% higher (respectively) in the discontinuation than maintenance group at 12 weeks. Comparing this with prior estimates of minimal clinically important differences supports a meaningful change for patients.[10]

Results when using ICD-10 criteria to define relapse diagnoses are shown in Table S11. Relapse was experienced by 33% (95% CI 27%, 39%) of participants in the maintenance group and 51% (95% CI 44%, 57%) in the discontinuation group (HR 2.23, 95% CI 1.68, 3.01, $p<0.0001$). This estimate was higher than the primary outcome analysis, but the confidence intervals overlapped indicating that the effects were similar.

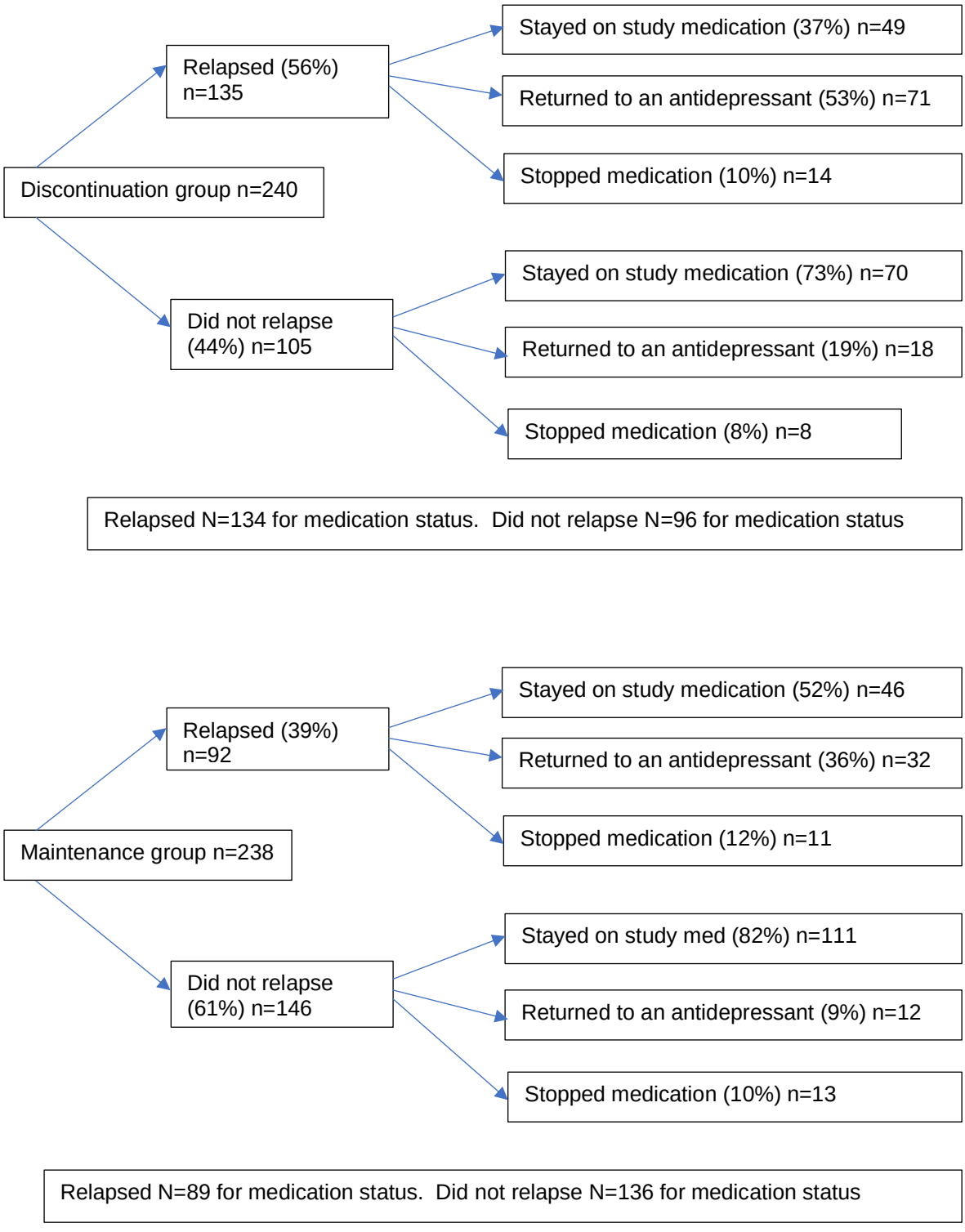

Figure S2. Tree diagrams showing outcomes for each randomised group, according to relapse.

Table S8. Summary of the main characteristics of the participating practices.

Characteristic	Category	% (n=150)
Centre	Bristol London York Southampton	20 55 10 15
Geographical location	Urban Rural	85 15
List size[b]	1-4999 5000-9999 10,000-14,999 15,000+	5 27 33 35
Number of GPs employed	0-5 6-10 11-15 16+	24 51 19 6
Number of randomised participants	0-4 5-10 11-15	76 19 5
Index of Multiple Deprivation	1-10 11-20 21-30 30+	22 43 26 9

[a]Based on the 2011 rural-urban classification for output areas in England
[b]Number of patients enrolled in practice
[c]The Index of Multiple Deprivation combines UK national census information from 38 indicators into seven domains of deprivation (income; employment; health and disability; education, skills, and training; barriers to housing and services; living environment; and crime). This results in a deprivation score for each 32,482 'lower super output area' in England, geographical units used for the reporting of neighbourhood level statistics.

Table S9. Age and gender of participants who were invited and those who participated.

	Characteristic					
	Number	Age in years		Number	Gender	
		Mean	SD		Female (n)	Female (%)
Trial participants[a]	478	55	12	478	349	73
Patients identified as potentially eligible and invited to participate[b]	20,060[c]	49	14	19,901[c]	14,106	71

[a]Included in the trial
[b]Identified as potentially eligible during the database search and sent an invitation letter. These data were provided by 120 of the 150 GP practices.
[c]A subset of the total number of participants who were identified as eligible and sent an invitation letter (23,429). This subset was comprised from the practices who returned these data.

Table S10: Log transformed depressive and anxiety symptoms as secondary outcomes.

	Maintenance		Discontinuation		Main models		With predictors of missingness	
Outcome	Mean	(SD)	Mean	(SD)	Estimate[§]	95% CI	Estimate	95% CI
PHQ-9 (coefficient)*								
Baseline	3.9	(3.5)	3.8	(3.6)				
6 weeks N=478	4.1	(3.8)	4.4	(4.0)	1.04	(0.89, 1.20)	1.00	(0.88, 1.14)
12 weeks N=477	4.1	(3.8)	6.3	(5.1)	1.41	(1.21, 1.64)	1.41	(1.21, 1.64)
26 weeks N=477	4.2	(3.7)	5.0	(4.6)	1.05	(0.90, 1.23)	1.04	(0.91, 1.20)
39 weeks N=477	3.8	(3.9)	4.4	(4.2)	1.06	(0.91, 1.24)	1.04	(0.92, 1.18)
52 weeks N=477	3.7	(3.7)	4.0	(4.5)	1.01	(0.87, 1.19)	1.00	(0.88, 1.13)
Over all time points N=478					1.08	(0.98, 1.21)	1.14	(1.01, 1.28)
GAD-7 (coefficient)†								
Baseline	3.2	(3.1)	2.8	(3.0)				
6 weeks N=478	3.2	(3.6)	3.6	(3.7)	1.07	(0.92, 1.25)	0.97	(0.85, 1.10)
12 weeks N=477	3.1	(3.3)	5.3	(4.6)	1.54	(1.32, 1.79)	1.54	(1.32, 1.79)
26 weeks N=477	3.4	(3.8)	4.1	(4.4)	1.13	(0.97, 1.33)	1.04	(0.91, 1.19)
39 weeks N=477	2.9	(3.5)	3.8	(4.1)	1.17	(1.00, 1.34)	1.02	(0.91, 1.16)
52 weeks N=477	3.0	(3.7)	3.1	3.0	1.00	(0.86, 1.18)	0.95	(0.84, 1.07)
Over all time points N=478					1.11	(0.99, 1.24)	1.16	(1.03, 1.31)

* PHQ-9 - Patient Health Questionnaire 9, range 0 to 27. Predictors of missingness: 6 weeks – site, 12 weeks - none, 26 weeks – age at randomization, able to replace worn out furniture, able to buy new clothes, stop taking medication if feeling worse, 39 weeks – age at randomization, SF12 physical component score at baseline, 52 weeks – age at randomization, site.

† GAD-7 – Generalised anxiety disorder 7, range 0 to 21. Predictors of missingness: 6 weeks – site, 12 weeks – none, 26 weeks – age at randomization, able to replace worn out furniture, stop taking medication if feeling worse, 39 weeks - age at randomization, SF12 physical component score at baseline, 52 weeks – age at randomization, site.

[§]Adjusted proportional difference in outcome scores between randomised groups. These models used a log-transformed PHQ-9 score as the outcome. Adjusted proportional differences can be interpreted as the difference in scores between randomised groups expressed as a proportion with values above 1.0 indicating more symptoms in the discontinuation group.

Table S11: Depression relapse using ICD-10 depression diagnosis.

Outcome	Maintenance n/N	%	Discontinuation n/N	%	Hazard ratio	95% CI
Time to first depression relapse (HR) N=468 in model	78/238	33	124/240	51	2.16	(1.61, 2.89)
Time to first depression relapse including minimisation variables (HR) N=468*					2.21	(1.64, 2.96)
Time to first depression relapse including predictors of missingness (HR) N=468†					2.20	(1.58 to 3.05)

*Study centre (4 groups), antidepressant medication (4 groups) and severity of depressive symptoms (above or below the median, 2 groups).
†Site, age at randomization, able to replace worn out furniture, able to buy new clothes, stop taking medication if feeling worse, physical component score at baseline.
HR = hazard ratio

Table S12: ANTLER medication status over time.

	Maintenance		Discontinuation	
	n/N	%	n/N	%
6 weeks				
On ANTLER medication	216/221	98	219/224	98
Not taking ANTLER medication, taking antidepressants	2/221	1	2/224	1
Not taking ANTLER medication or antidepressants	3/221	1	3/224	1
12 weeks				
On ANTLER medication	216/228	95	184/216	85
Not taking ANTLER medication, taking antidepressants	6/228	3	21/216	10
Not taking ANTLER medication or antidepressants	6/228	3	11/216	5
26 weeks				
On ANTLER medication	184/211	87	119/193	62
Not taking ANTLER medication, taking antidepressants	11/211	5	53/193	27
Not taking ANTLER medication or antidepressants	16/211	8	21/193	11
39 weeks				
On ANTLER medication	175/212	83	96/184	52
Not taking ANTLER medication, taking antidepressants	25/212	12	69/184	38
Not taking ANTLER medication or antidepressants	12/212	6	19/184	10
52 weeks				
On ANTLER medication	148/210	70	83/180	46
Not taking ANTLER medication, taking antidepressants	30/210	14	59/180	33
Not taking ANTLER medication or antidepressants	32/210	15	38/180	21
Over the course of the study				
On ANTLER medication	157/225	70	119/230	52
Not taking ANTLER medication, taking antidepressants	44/225	20	89/230	39
Not taking ANTLER medication or antidepressants	24/225	11	22/230	10

CONSORT 2010 checklist

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	Not done
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	3-4
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	5-6
	2b	Specific objectives or hypotheses	5
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	6
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	6
Participants	4a	Eligibility criteria for participants	7-8
	4b	Settings and locations where the data were collected	6
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	8-9
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	10-11
	6b	Any changes to trial outcomes after the trial commenced, with reasons	Protocol changes document
Sample size	7a	How sample size was determined	13
	7b	When applicable, explanation of any interim analyses and stopping guidelines	14
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	Supplement
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	Supplement
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	Supplement
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	Supplement

Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	13
	11b	If relevant, description of the similarity of interventions	9
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	13-15
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	13-15
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	Figure 1
	13b	For each group, losses and exclusions after randomisation, together with reasons	Figure 1
Recruitment	14a	Dates defining the periods of recruitment and follow-up	Not reported
	14b	Why the trial ended or was stopped	Not reported
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	Table 1
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	16-17
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	18-19
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	Just relative
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	Supplement
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	19
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	20-21
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	20-21
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	19-22
Other information			
Registration	23	Registration number and name of trial registry	4
Protocol	24	Where the full trial protocol can be accessed, if available	With ms
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	27

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.

References

1 Frank E, Prien RF, Jarrett RB, *et al.* Conceptualization and rationale for consensus definitions of terms in major depressive disorder: Remission, recovery, relapse, and recurrence. *Arch Gen Psychiatry* 1991; **48**: 851–5.

2 The NICE Guideline on the Treatment and Management of Depression in Adults Updated Edition. In: Excellence NIfHaC, editor. London2010. .

3 Lewis G. Assessing psychiatric disorder with a human interviewer or a computer. *J Epidemiol Community Health* 1994; **48**: 207–10.

4 Lewis G, Pelosi AJ, Araya R, Dunn G. Measuring psychiatric disorder in the community: a standardized assessment for use by lay interviewers. *Psychol Med* 1992; **22**: 465.

5 Wiles N, Thomas L, Abel A, *et al.* Cognitive behavioural therapy as an adjunct to pharmacotherapy for primary care based patients with treatment resistant depression: results of the CoBalT randomised controlled trial. *Lancet* 2013; **381**: 375–84.

6 Kessler DS, Macneill SJ, Tallon D, *et al.* Mirtazapine added to SSRIs or SNRIs for treatment resistant depression in primary care: phase III randomised placebo controlled trial (MIR).

7 Rosenbaum JF, Fava M, Hoog SL, Ascroft RC, Krebs WB. Selective serotonin reuptake inhibitor discontinuation syndrome: A randomized clinical trial. *Biol Psychiatry* 1998; **44**: 77–87.

8 Tiwsk J, Bosman L, Hoekstra T, Rijnhart J, Welten M, Heymans M. Different ways to estimate treatment effects in randomised controlled trials. *Contemp Clin trials Commun* 2018; **10**: 80–5.

9 Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: The GAD-7. *Arch Intern Med* 2006.

10 Lewis G, Duffy L, Ades A, *et al.* The clinical effectiveness of sertraline in primary care and the role of depression severity and duration (PANDA): a pragmatic, double-blind, placebo-controlled randomised trial. *The Lancet Psychiatry* 2019; **0366**: 1–12.

11 Kranzler HR, Burleson JA, Korner P, *et al.* Placebo-controlled trial of fluoxetine as an adjunct to relapse prevention in alcoholics. *Am J Psychiatry* 1995; **152**: 391–7.