

Health Technology Assessment

Volume 25 • Issue 69 • November 2021

ISSN 1366-5278

Antidepressant medication to prevent depression relapse in primary care: the ANTLER RCT

Larisa Duffy, Caroline S Clarke, Gemma Lewis, Louise Marston, Nick Freemantle, Simon Gilbody, Rachael Hunter, Tony Kendrick, David Kessler, Michael King, Paul Lanham, Dee Mangin, Michael Moore, Irwin Nazareth, Nicola Wiles, Faye Bacon, Molly Bird, Sally Brabyn, Alison Burns, Yvonne Donkor, Anna Hunt, Jodi Pervin and Glyn Lewis



Antidepressant medication to prevent depression relapse in primary care: the ANTLER RCT

Larisa Duffy^{ID},^{1*} Caroline S Clarke^{ID},^{2,3} Gemma Lewis^{ID},¹
Louise Marston^{ID},^{2,3} Nick Freemantle^{ID},^{3,4}
Simon Gilbody^{ID},⁵ Rachael Hunter^{ID},^{2,3} Tony Kendrick^{ID},⁶
David Kessler^{ID},⁷ Michael King^{ID},^{2,3} Paul Lanham^{ID},¹
Dee Mangin^{ID},^{8,9} Michael Moore^{ID},⁶ Irwin Nazareth^{ID},^{2,3}
Nicola Wiles^{ID},⁷ Faye Bacon^{ID},¹ Molly Bird^{ID},¹
Sally Brabyn^{ID},⁵ Alison Burns^{ID},⁷ Yvonne Donkor^{ID},¹
Anna Hunt^{ID},⁶ Jodi Pervin^{ID},⁵ and Glyn Lewis^{ID},¹

¹Division of Psychiatry, University College London, London, UK

²Research Department of Primary Care and Population Health, University College London, London, UK

³PRIMENT Clinical Trials Unit, University College London, London, UK

⁴Comprehensive Clinical Trials Unit, University College London, London, UK

⁵Department of Health and Social Care Sciences, University of York, York, UK

⁶Primary Care Population Sciences and Medical Education, University of Southampton, Southampton, UK

⁷Centre for Academic Mental Health, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

⁸Department of Family Medicine, McMaster University, Hamilton, ON, Canada

⁹Department of General Practice, University of Otago, Christchurch, New Zealand

*Corresponding author

Declared competing interests of authors: Simon Gilbody reports serving on a number of funding committees: Health Technology Assessment (HTA) Efficient Study Designs 2 (2015–16), HTA End of Life Care and Add-on Studies (2014–16), HTA Funding Committee Policy Group (formerly CSG) (2017–20), HTA Clinical Evaluation and Trials Committee (2008–14), and HTA Commissioning Committee (2016–20). Tony Kendrick reports grants from the National Institute for Health Research (NIHR) during the conduct of the study. Glyn Lewis reports grants from University College London during the conduct of the study; personal fees from Fortitude Law (London, UK), outside the submitted work; and being a member of the Efficacy and Mechanism Evaluation Funding Committee (2011–16). Michael Moore reports grants from NIHR during the conduct of the study. Irwin Nazareth reports, as director of PRIMMENT Clinical Trials Unit, core support funds from NIHR (that will be provided until 31 August 2022) for the Clinical Trials Unit, and was a member of the HTA Commissioning Sub-Board (2009–17), the HTA Primary Care Themed Call board (2013–14) and the HTA Commissioning Committee (2010–17).

Published November 2021

DOI: 10.3310/hta25690

This report should be referenced as follows:

Duffy L, Clarke CS, Lewis G, Marston L, Freemantle N, Gilbody S, *et al.* Antidepressant medication to prevent depression relapse in primary care: the ANTLER RCT. *Health Technol Assess* 2021;**25**(69).

Health Technology Assessment is indexed and abstracted in *Index Medicus/MEDLINE*, *Excerpta Medica/EMBASE*, *Science Citation Index Expanded (SciSearch®)* and *Current Contents®/Clinical Medicine*.

Health Technology Assessment

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.014

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, the Cochrane Library and Clarivate Analytics Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 13/115/48. The contractual start date was in May 2016. The draft report began editorial review in November 2020 and was accepted for publication in July 2021. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care.

Copyright © 2021 Duffy *et al.* This work was produced by Duffy *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaptation in any medium and for any purpose provided that it is properly attributed. See: <https://creativecommons.org/licenses/by/4.0/>. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

NIHR Journals Library Editor-in-Chief

Professor Ken Stein Professor of Public Health, University of Exeter Medical School, UK

NIHR Journals Library Editors

Professor John Powell Chair of HTA and EME Editorial Board and Editor-in-Chief of HTA and EME journals. Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK, and Professor of Digital Health Care, Nuffield Department of Primary Care Health Sciences, University of Oxford, UK

Professor Andrée Le May Chair of NIHR Journals Library Editorial Group (HS&DR, PGfAR, PHR journals) and Editor-in-Chief of HS&DR, PGfAR, PHR journals

Professor Matthias Beck Professor of Management, Cork University Business School, Department of Management and Marketing, University College Cork, Ireland

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Eugenia Cronin Senior Scientific Advisor, Wessex Institute, UK

Dr Peter Davidson Consultant Advisor, Wessex Institute, University of Southampton, UK

Ms Tara Lamont Senior Scientific Adviser (Evidence Use), Wessex Institute, University of Southampton, UK

Dr Catriona McDaid Senior Research Fellow, York Trials Unit, Department of Health Sciences, University of York, UK

Professor William McGuire Professor of Child Health, Hull York Medical School, University of York, UK

Professor Geoffrey Meads Emeritus Professor of Wellbeing Research, University of Winchester, UK

Professor James Raftery Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

Dr Rob Riemsma Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

Professor Helen Roberts Professor of Child Health Research, UCL Great Ormond Street Institute of Child Health, UK

Professor Jonathan Ross Professor of Sexual Health and HIV, University Hospital Birmingham, UK

Professor Helen Snooks Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Professor Ken Stein Professor of Public Health, University of Exeter Medical School, UK

Professor Jim Thornton Professor of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, University of Nottingham, UK

Please visit the website for a list of editors: www.journalslibrary.nihr.ac.uk/about/editors

Editorial contact: journals.library@nihr.ac.uk

Abstract

Antidepressant medication to prevent depression relapse in primary care: the ANTLER RCT

Larisa Duffy^{ID},^{1*} Caroline S Clarke^{ID},^{2,3} Gemma Lewis^{ID},¹
Louise Marston^{ID},^{2,3} Nick Freemantle^{ID},^{3,4} Simon Gilbody^{ID},⁵
Rachael Hunter^{ID},^{2,3} Tony Kendrick^{ID},⁶ David Kessler^{ID},⁷ Michael King^{ID},^{2,3}
Paul Lanham^{ID},¹ Dee Mangin^{ID},^{8,9} Michael Moore^{ID},⁶ Irwin Nazareth^{ID},^{2,3}
Nicola Wiles^{ID},⁷ Faye Bacon^{ID},¹ Molly Bird^{ID},¹ Sally Brabyn^{ID},⁵
Alison Burns^{ID},⁷ Yvonne Donkor^{ID},¹ Anna Hunt^{ID},⁶ Jodi Pervin^{ID}⁵
and Glyn Lewis^{ID}¹

¹Division of Psychiatry, University College London, London, UK

²Research Department of Primary Care and Population Health, University College London, London, UK

³PRIMENT Clinical Trials Unit, University College London, London, UK

⁴Comprehensive Clinical Trials Unit, University College London, London, UK

⁵Department of Health and Social Care Sciences, University of York, York, UK

⁶Primary Care Population Sciences and Medical Education, University of Southampton, Southampton, UK

⁷Centre for Academic Mental Health, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

⁸Department of Family Medicine, McMaster University, Hamilton, ON, Canada

⁹Department of General Practice, University of Otago, Christchurch, New Zealand

*Corresponding author larisa.duffy@ucl.ac.uk

Background: There has been a steady increase in the number of primary care patients receiving long-term maintenance antidepressant treatment, despite limited evidence of a benefit of this treatment beyond 8 months.

Objective: The ANTLER trial investigated the clinical effectiveness and cost-effectiveness of antidepressant medication in preventing relapse in UK primary care.

Design: This was a Phase IV, double-blind, pragmatic, multisite, individually randomised parallel-group controlled trial, with follow-up at 6, 12, 26, 39 and 52 weeks. Participants were randomised using minimisation on centre, type of antidepressant and baseline depressive symptom score above or below the median using Clinical Interview Schedule – Revised (two categories). Statisticians were blind to allocation for the outcome analyses.

Setting: General practices in London, Bristol, Southampton and York.

Participants: Individuals aged 18–74 years who had experienced at least two episodes of depression and had been taking antidepressants for ≥ 9 months but felt well enough to consider stopping their medication. Those who met an *International Statistical Classification of Diseases and Related Health Problems*, Tenth Revision, diagnosis of depression or with other psychiatric conditions were excluded.

Intervention: At baseline, participants were taking citalopram 20 mg, sertraline 100 mg, fluoxetine 20 mg or mirtazapine 30 mg. They were randomised to either remain on their current medication or discontinue medication after a tapering period.

Main outcome measures: The primary outcome was the time, in weeks, to the beginning of the first depressive episode after randomisation. This was measured by a retrospective Clinical Interview Schedule – Revised that assessed the onset of a depressive episode in the previous 12 weeks, and was conducted at 12, 26, 39 and 52 weeks. The depression-related resource use was collected over 12 months from medical records and patient-completed questionnaires. Quality-adjusted life-years were calculated using the EuroQol-5 Dimensions, five-level version.

Results: Between 9 March 2017 and 1 March 2019, we randomised 238 participants to antidepressant continuation (the maintenance group) and 240 participants to antidepressant discontinuation (the discontinuation group). The time to relapse of depression was shorter in the discontinuation group, with a hazard ratio of 2.06 (95% confidence interval 1.56 to 2.70; $p < 0.0001$). By 52 weeks, relapse was experienced by 39% of those who continued antidepressants and 56% of those who discontinued antidepressants. The secondary analysis revealed that people who discontinued experienced more withdrawal symptoms than those who remained on medication, with the largest difference at 12 weeks. In the discontinuation group, 37% (95% confidence interval 28% to 45%) of participants remained on their randomised medication until the end of the trial. In total, 39% (95% confidence interval 32% to 45%) of participants in the discontinuation group returned to their original antidepressant compared with 20% (95% confidence interval 15% to 25%) of participants in maintenance group. The health economic evaluation demonstrated that participants randomised to discontinuation had worse utility scores at 3 months (-0.037, 95% confidence interval -0.059 to -0.015) and fewer quality-adjusted life-years over 12 months (-0.019, 95% confidence interval -0.035 to -0.003) than those randomised to continuation. The discontinuation pathway, besides giving worse outcomes, also cost more [extra £2.71 per patient over 12 months (95% confidence interval -£36.10 to £37.07)] than the continuation pathway, although the cost difference was not significant.

Conclusions: Patients who discontinue long-term maintenance antidepressants in primary care are at increased risk of relapse and withdrawal symptoms. However, a substantial proportion of patients can discontinue antidepressants without relapse. Our findings will give patients and clinicians an estimate of the likely benefits and harms of stopping long-term maintenance antidepressants and improve shared decision-making. The participants may not have been representative of all people on long-term maintenance treatment and we could study only a restricted range of antidepressants and doses. Identifying patients who will not relapse if they discontinued antidepressants would be clinically important.

Trial registration: Current Controlled Trials ISRCTN15969819 and EudraCT 2015-004210-26.

Funding: This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 25, No. 69. See the NIHR Journals Library website for further project information.

Contents

List of tables	xi
List of figures	xiii
List of boxes	xv
List of supplementary material	xvii
List of abbreviations	xix
Plain English summary	xxi
Scientific summary	xxiii
Chapter 1 Introduction	1
Description of the health problem	1
Treatment options for depression	1
Economic consequences of depression	1
Current evidence on the effectiveness of maintenance treatment	1
Current evidence on withdrawal symptoms	2
Aim and objectives	2
Chapter 2 Methods	3
Trial design	3
Ethics approval and research governance	3
Participants	3
<i>Inclusion criteria</i>	3
<i>Exclusion criteria</i>	5
Recruitment	5
<i>Record search method</i>	5
<i>In-consultation recruitment method</i>	5
Intervention	5
<i>Choice of medication</i>	5
<i>Treatment of participants</i>	7
Subsequent assessments	7
Outcomes	7
<i>Primary outcome</i>	7
<i>Secondary outcomes</i>	8
<i>Mechanistic outcomes</i>	9
Sample size	10
Randomisation and blinding	12
Statistical methods	12
<i>Primary outcome and secondary analysis</i>	12
Reliability of the retrospective Clinical Interview Schedule – Revised	13
<i>Why we needed a new measure to assess relapse</i>	13
<i>The aim of the reliability study</i>	14
<i>Testing the algorithm</i>	15

CONTENTS

<i>Analysis</i>	16
<i>Results of the reliability study of the retrospective Clinical Interview Schedule – Revised</i>	17
Chapter 3 Results	21
Flow of participants in the trial	21
Baseline comparability	21
Primary outcome	21
Secondary outcomes	21
Missing data	27
Adherence to trial medication	29
Subgroup analyses	29
Post hoc analyses	34
Chapter 4 Economic evaluation	39
Introduction	39
Methods	39
<i>Outcome measures</i>	39
<i>Resource use and costs</i>	39
<i>Utilities and quality-adjusted life-years</i>	41
Economic evaluation analytical methods	41
<i>Health economics secondary and sensitivity analyses</i>	42
Health economics results	42
Costs	42
<i>Utility scores and quality-adjusted life-years</i>	46
<i>Cost-utility analysis</i>	47
<i>Secondary and sensitivity analyses</i>	49
Chapter 5 Discussion and conclusions	51
Summary of findings	51
Summary of economic evaluation findings	51
Strength and limitations	52
Implications for health care	53
Recommendation for research	54
Acknowledgements	55
References	57

List of tables

TABLE 1 Participating general practice characteristics	4
TABLE 2 Full schedule of questionnaires used in the ANTLER trial	11
TABLE 3 Baseline characteristics of the sample who completed the rCIS-R twice at the same follow-up appointment	17
TABLE 4 Level of agreement of relapse in depression and individual symptoms between the first and the second completions of the rCIS-R	18
TABLE 5 Table indicating agreement for relapse in depression between the first and the second completion of the rCIS-R	18
TABLE 6 Baseline demographic and clinical characteristics by treatment group	23
TABLE 7 Hazard ratios for relapse in the discontinuation group compared with the maintenance group	24
TABLE 8 Means and estimated mean difference for secondary outcomes in the discontinuation group compared with the maintenance group	25
TABLE 9 Secondary outcomes controlling for factors associated with missingness	28
TABLE 10 Subgroup analysis: time to relapse	30
TABLE 11 Subgroup analysis: PHQ-9	30
TABLE 12 Subgroup analysis: GAD-7	32
TABLE 13 Subgroup analysis: Global Rating Question – feel worse	35
TABLE 14 Unit costs applied to resource use collected from patient electronic medical records at GP surgeries, and collected directly from participants via the modified CSRI questionnaire	40
TABLE 15 Mean raw costs of health-care resource use and estimated adjusted differences between the discontinuation group and the maintenance group: primary care	43
TABLE 16 Health-care resource use: CSRI	44
TABLE 17 Health-care resource use: antidepressants medication costs, including the ANTLER trial medication	44
TABLE 18 Health-care resource use: total costs	45
TABLE 19 Mean raw utilities at baseline and all follow-up points and mean 12-month QALYs, and estimated adjusted differences between the discontinuation group and the maintenance group	47

LIST OF TABLES

TABLE 20 Summary of the bootstrapped incremental costs and QALYs, and the ICERs, using the health-care and social care cost perspective	49
TABLE 21 Overall incremental costs and QALYs using the wider societal perspective	49

List of figures

FIGURE 1 Flow chart outlining the stages of recruitment	6
FIGURE 2 Schedule of assessments	8
FIGURE 3 Difference in the total rCIS-R score against the mean total score	19
FIGURE 4 The CONSORT flow diagram for the ANTLER trial	22
FIGURE 5 Kaplan–Meier plot for the primary outcome	25
FIGURE 6 Raw, unadjusted complete-case mean utility scores at each time point by treatment group for each of the three methods	48
FIGURE 7 Cost-effectiveness plane using multiple imputation for utilities and CSRI costs, showing 1000 bootstrapped results from the seemingly unrelated regression of costs (NHS and PSS perspective) and QALYs (EQ-5D-5L TTO)	48
FIGURE 8 Cost-effectiveness acceptability curve generated from the CEP in <i>Figure 7</i> showing the likelihood of the discontinuation group being sufficiently cost-effective compared with the maintenance group, at a range of values for the cost-effectiveness threshold from zero to £10,000 per QALY	50

List of boxes

BOX 1 Concentration section from the rCIS-R

16

List of supplementary material

- Report Supplementary Material 1** Screening questionnaire
- Report Supplementary Material 2** Baseline questionnaire
- Report Supplementary Material 3** Six-week follow-up questionnaire
- Report Supplementary Material 4** Twelve-, 26-, 39- and 52-week follow-up questionnaire
- Report Supplementary Material 5** Health-care resource use questionnaire
- Report Supplementary Material 6** Medication from patients' notes template
- Report Supplementary Material 7** Exit questionnaire
- Report Supplementary Material 8** Health Economics post hoc analysis
- Report Supplementary Material 9** Tree diagrams
- Report Supplementary Material 10** Subgroup analysis
- Report Supplementary Material 11** Comparison of the age and gender of the participants who were invited with those who participated
- Report Supplementary Material 12** Missingness analysis

Supplementary material can be found on the NIHR Journals Library report page (<https://doi.org/10.3310/hta25690>).

Supplementary material has been provided by the authors to support the report and any files provided at submission will have been seen by peer reviewers, but not extensively reviewed. Any supplementary material provided at a later stage in the process may not have been peer reviewed.

List of abbreviations

ANTLER	ANTidepressants to prevent reLapse in dEpReSSion	ISRCTN	International Standard Randomised Controlled Trial Number
CEAC	cost-effectiveness acceptability curve	NICE	National Institute for Health and Care Excellence
CEP	cost-effectiveness plane	NIHR	National Institute for Health Research
CI	confidence interval	NRES	National Research Ethics Service
CIDI	Composite International Diagnostic Interview	PHQ-9	Patient Health Questionnaire-9 items
CIS-R	Clinical Interview Schedule – Revised	PPI	public and patient involvement
CONSORT	Consolidated Standards of Reporting Trials	PSSRU	Personal Social Services Research Unit
CRN	Clinical Research Network	QALY	quality-adjusted life-year
CSRI	Client Service Receipt Inventory	rCIS-R	Clinical Interview Schedule – Revised
DESS	Discontinuation-Emergent Signs and Symptoms	RCT	randomised controlled trial
EQ-5D-5L	EuroQol-5 Dimensions, five-level version	SCID	Structured Clinical Interview Disorder
GAD-7	Generalised Anxiety Disorder-7	SD	standard deviation
GP	general practitioner	SE	standard error
HR	hazard ratio	SF-6D	Short Form questionnaire-6 Dimensions
HTA	Health Technology Assessment	SF-12	Short Form questionnaire-12 items
ICD-10	<i>International Statistical Classification of Diseases and Related Health Problems, Tenth Revision</i>	SSRI	selective serotonin reuptake inhibitor
ICER	incremental cost-effectiveness ratio	SURF	Service Users Research Forum
		TTO	time trade-off

Plain English summary

Antidepressants are used to treat depression when someone is unwell, but are also used as maintenance treatment to prevent the reoccurrence of depression. There has been a large increase in the use of long-term maintenance antidepressant treatment, but the evidence for the benefits of maintenance beyond 8 months is very poor.

The ANTidepressants to prevent reLapse in dEpReSSION (ANTLER) trial was a randomised controlled trial that examined the effectiveness of long-term maintenance treatment with antidepressants. The participants were well enough to consider stopping antidepressant medication, were recruited from primary care and had taken antidepressants for ≥ 9 months. In total, 238 participants were randomised to continue taking antidepressants and 240 were randomised to receive a visually identical tablet that contained no active ingredients after a period when the antidepressants were gradually reduced. Neither the participants nor those interviewing them knew which group they had been placed in, and they were followed up for 1 year.

Participants who discontinued antidepressants were more likely to experience relapse than those who continued antidepressants. By 52 weeks, 39% of those who continued antidepressants had experienced a relapse, compared with 56% in the group that discontinued antidepressants. In other words, over a 52-week period, one in every six patients who stopped antidepressants would experience a relapse that may not have occurred if they had remained on their antidepressants. Patients in the discontinuation group reported more symptoms of anxiety and depression and experienced more withdrawal symptoms than those in the maintenance group, mostly in the first 3–4 months after stopping the antidepressants. Participants in the discontinuation group also reported lower quality of life than those in the maintenance group but both groups used similar amounts of health-care and social care resources over the 12-month period. About one-third of participants who were allocated to the discontinuation group in the ANTLER trial decided to restart their antidepressants. However, another one-third of participants in that group remained on trial medication for 12 months and managed without antidepressants.

Long-term maintenance treatment with antidepressants is effective in reducing the rate of relapses. For those who are considering stopping their antidepressant, our findings will provide estimates of the likely benefits and harms, to improve shared decision-making and support the regular review of long-term antidepressant prescription.

Scientific summary

Background

Depression is the leading cause of ill health and disability worldwide, with > 300 million people now living with depression. Depression causes marked emotional distress and interferes with daily function not only of the individual, but also of society. It has been estimated that depression reduces England's national income (gross national product) by over 4% (approximately £80M) every year. This reduction results from increased unemployment, a larger number of sick days and reduced productivity. It is also accompanied by increased welfare expenditure.

Antidepressants are often a first-line treatment for depressive symptoms and are also used for maintenance treatment, that is, to prevent relapse once an individual has recovered. It has been estimated that 90% of antidepressant prescriptions in the UK were used for maintenance between 1993 and 2005. A more recent UK study has also demonstrated a steady increase in the duration of long-term treatment between 2001 and 2012. In a Scottish study involving 78 practices, 8.6% of the registered population were prescribed an antidepressant, and half had been taking antidepressants for > 2 years (Johnson CF, MacDonald HJ, Atkinson P, Buchanan AI, Downes N, Dougall N. Reviewing long-term antidepressants can reduce drug burden: a prospective observational cohort study. *Br J Gen Pract* 2012;**62**:e773–9). This trend is similar in other high-income countries.

The National Institute for Health and Care Excellence (NICE) in England recommends that antidepressant maintenance treatments should continue to be used for 2 years for those at risk of relapse. However, NICE also recognises the uncertainty about the benefit of long-term maintenance treatment and has recommended further research into its psychological and pharmacological effects. Continuing maintenance treatment in the first few months after remission has been extensively studied and reduces relapse rates. However, the evidence for a treatment period longer than 8 months is insufficient to justify long-term maintenance treatment. The studies were conducted during the 1980s or early 1990s in secondary care by pharmaceutical companies for regulatory purposes; one was investigating tricyclic antidepressants that are no longer used for depression. All were very small studies ($n < 20$ participants in total) and had a poor follow-up rate.

Aim

The overall aim of the ANTLER trial was to answer the following research question: 'What is the clinical effectiveness and cost-effectiveness in UK primary care of continuing on long-term maintenance antidepressants compared with a placebo in preventing relapse of depression in those who have taken antidepressants for more than 9 months and who are now well enough to consider stopping maintenance treatment?'. The trial was embedded in primary care and had broad inclusion criteria to increase the generalisability to the population currently receiving maintenance antidepressants.

Methods

The trial was a Phase IV, double-blind, pragmatic, multisite, individually randomised parallel-group controlled trial, with follow-up at 6, 12, 26, 39 and 52 weeks.

We recruited primary care patients who were taking one of four of the most commonly used antidepressant medications. At the point of recruitment, the patients were well enough to consider stopping their medication. Participants were recruited from 150 primary care practices in four UK sites: London, Bristol, Southampton and York. We recruited the first participant in March 2017, and within 2 years we randomised 238 participants to antidepressant continuation and 240 participants to discontinuation. The participants were individuals aged 18–74 years who had experienced at least two episodes of depression and had been taking antidepressants for ≥ 9 months but felt well enough to consider stopping their medication. Participants were excluded if they met *International Statistical Classification of Diseases and Related Health Problems*, Tenth Revision, criteria for depressive illness assessed with the clinical interview schedule – Revised (CIS-R) (see below); had bipolar disorder, psychotic illness, dementia or 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 breastfeeding. At baseline, participants were taking citalopram 20 mg, sertraline 100 mg, fluoxetine 20 mg or mirtazapine 30 mg. They were randomised to either remain on their current medication or switch to placebo after a tapering period of 1 (fluoxetine) or 2 (citalopram, sertraline or mirtazapine) months. The trial compared continuing antidepressant medication with discontinuing medication by replacing it with an identical placebo after a tapering period.

The primary outcome was the time, in weeks, to the beginning of the first depressive episode after randomisation. This was measured by a shortened CIS-R that assessed the onset of a depressive episode in the previous 12 weeks and was conducted at 12, 26, 39 and 52 weeks. We called the new assessment the retrospective CIS-R (rCIS-R) and conducted a test–retest reliability study, which was nested within the ANTLER trial. The ANTLER trial participants were asked to complete the rCIS-R twice, at the beginning and the end of one of the follow-up appointments, and 396 participants provided data. The depression-related resource use was collected over 12 months and for the 6 months preceding baseline from medical records and patient-completed questionnaires. Quality-adjusted life-years (QALYs) were calculated using the EuroQol-5 Dimensions, five-level version.

Results

Recruitment began on 9 March 2017 and the last participant was randomised on 1 March 2019. The general practitioner record search identified 23,429 potentially eligible patients, who were sent an invitation letter. Another 124 potentially eligible patients were referred during general practitioner consultation, resulting in 1466 patients wanting to take part. Of these patients, 606 were eligible. In total, 478 participants were randomised: 238 to maintenance antidepressant and 240 to placebo. A total of 390 participants (82%) completed the trial. All participants provided data on whether or not they relapsed; however, 10 (maintenance group, $n = 6$; discontinuation group, $n = 4$) participants did not provide data on timing of relapse, so could not be included in the analysis of the primary outcome.

The groups were well balanced at baseline, with a mean age of 54 years [standard deviation (SD) 13 years] in the maintenance group and 55 years (SD 12 years) in the discontinuation group. Just over 40% of participants were recruited from London, 20% from each of Bristol and Southampton and 17% from York. Under half were taking citalopram, one-third fluoxetine, one-sixth sertraline and under one-twentieth mirtazapine. Almost three-quarters of participants had taken antidepressants for > 3 years, with over one-third taking them for ≥ 6 years.

The hazard ratio for relapse in the discontinuation group was 2.06 [95% confidence interval (CI) 1.56 to 2.70; $p < 0.0001$], with 39% of those who continued antidepressants and 56% who discontinued experiencing relapse. In other words, over a 52-week period, one in every six patients who stopped antidepressants would experience a relapse that may not have occurred if they had remained on their antidepressants. A similar pattern was observed for the secondary outcomes, for which participants in the

discontinuation group at 12 weeks had more depressive and anxious symptoms (coefficients Patient Health Questionnaire-9 items 2.12, 95% CI 1.37 to 2.86; Generalised Anxiety Disorder-7 2.24, 95% CI 1.59 to 2.89) and poorer mental health quality of life (coefficient Short Form questionnaire-12 items -4.61, 95% CI -6.42 to -2.81) than those in the maintenance group, and had more than twice the odds of feeling worse using the Global Rating Question (odds ratio 2.88, 95% CI 1.90 to 4.38). In the discontinuation group, 37% (95% CI 28% to 45%) of participants remained on their randomised medication until the end of the trial, and 39% (95% CI 32% to 45%) returned to their original antidepressant compared with 20% (95% CI 15% to 25%) of participants in maintenance group.

People who discontinued antidepressants experienced more withdrawal symptoms than those who remained on medication, with the largest difference at 12 weeks (coefficient 1.81, 95% CI 1.40 to 2.23) and reducing thereafter. Among those who relapsed in the discontinuation group, 53% of participants chose to return to an antidepressant prescribed by their doctor.

The health economic evaluation found that participants randomised to discontinuation had worse utility scores at 3 months (-0.037, 95% CI -0.059 to -0.015) and fewer QALYs over 12 months (-0.019, 95% CI -0.035 to -0.003) than those randomised to the maintenance group. The discontinuation group was dominated by the maintenance group in that the discontinuation pathway, besides giving worse outcomes, also cost more (extra £2.71 per patient over 12 months, 95% CI -£36.10 to £37.07), although the cost difference was not statistically significant. There was only a 12% probability that discontinuing antidepressants was sufficiently cost-effective compared with maintenance therapy at a cost-effectiveness threshold of £30,000 per QALY gained.

Conclusions

Our trial found that primary care patients who discontinue long-term maintenance antidepressants were at increased risk of relapse and withdrawal symptoms, particularly in the first few months after ending antidepressants. The results of our economic evaluation suggest that discontinuing antidepressant medication would not be recommended by national decision-makers on the grounds of cost-effectiveness. Some individuals may choose to taper and stop antidepressants to see if they can manage without antidepressants. For example, a substantial proportion of patients are still able to discontinue even if they relapse. Our findings will give patients and clinicians an estimate of the likely benefits and harms of stopping long-term maintenance antidepressants and improve shared decision-making. It supports the policy that there should be regular medical review of long-term maintenance antidepressant medication.

Trial registration

This trial is registered as ISRCTN15969819 and EudraCT 2015-004210-26.

Funding

This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 25, No. 69. See the NIHR Journals Library website for further project information.

Chapter 1 Introduction

Description of the health problem

Depression is a leading cause of disability, with more than 300 million people having depressive illness worldwide.¹ According to the *International Statistical Classification of Diseases and Related Health Problems*, Tenth Revision (ICD-10), depression is defined as a state of mental health that can be characterised by symptoms of low mood, loss of interest or pleasure, tearfulness, feelings of guilt or low self-worth, social withdrawal, disturbed sleep or appetite, poor concentration, tiredness and diminished activity.²

Owing to difficulties in distinguishing between clinically significant and 'normal' mood changes, it is now generally accepted that depressive symptoms are a continuum of severity.³ The severity of depressive illness can range from mild to severe. Even mild to moderate depression can impair people's ability to function and cope with daily life. The diagnosis of depression is based not only on severity of symptoms but also on their duration and the degree of social and functional impairment. Depression is no longer thought of as a time-limited disorder with complete recovery after 4–6 months of treatment; it is now accepted that depression is often chronic (lasting for ≥ 2 years) or recurrent. Although it has been accepted that depression exists as a continuum, researchers and clinicians still use terms such as episode, relapse and chronicity to guide the diagnosis, and to inform and monitor treatment.

Treatment options for depression

People with depressive symptoms are mainly treated in primary care, and antidepressants are usually the first-line treatment; they are used to treat acute depressive symptoms and for maintenance treatment, that is to prevent relapse once an individual has recovered. The number of prescriptions for antidepressant medication has risen dramatically in high-income countries in recent decades, largely because of an increase in the number of patients receiving long-term treatment.^{4,5} Psychological treatments, such as cognitive-behavioural therapy, are also effective treatments for depression, including for those who have not responded to antidepressants.⁶

Economic consequences of depression

Depression not only causes marked emotional distress and interferes with daily function for the individual, but also has substantial negative social and financial impact on the wider community.⁷ Every year, it has been estimated that depression reduces England's national income (gross national product) by over 4% (approximately £80M). This reduction results from increased unemployment, a larger number of sick days and reduced productivity. It is also accompanied by increased welfare expenditure.^{8,9}

Current evidence on the effectiveness of maintenance treatment

A substantial proportion of the burden of depression arises from relapses, recurrence and chronicity. However, in contrast to the large number of drug trials in acute depression,¹⁰ there are relatively few studies assessing antidepressant efficacy in preventing relapse of depression during maintenance treatment. A number of systematic reviews and meta-analyses^{11–14} report a reduced risk of relapse rate in patients receiving antidepressant medication by between 50% and 70% compared with patients receiving placebo. The constituent studies had several limitations. These studies recruited patients from secondary care during an acute depressive episode, treated the patients with an open-label antidepressant and reported that only those who met criteria for recovery were eligible for randomisation to either

remain on double-blind antidepressant or switch to placebo. Most studies were conducted during the 1980s or early 1990s in secondary care by pharmaceutical companies for regulatory purposes, using tricyclic antidepressants that are no longer widely used for depression. Most studies had either short (≤ 2 months) or intermediate (3–5 months) pre-randomisation treatment. It is difficult to generalise these findings to people currently receiving antidepressants in primary care, many of whom have been on maintenance treatment for some years.^{15,16} Studies of the effectiveness of maintenance treatment for patients receiving antidepressants for longer than 8 months are rare and have several limitations.^{17–19} All were very small ($n < 20$ participants in total) and had a poor follow-up rate.

NICE guidance on the treatment and management of depression in adults²⁰ recommends that patients who have recovered from an episode of depression should stay on their antidepressant medication for at least 6 months after remission. The guidelines also suggest that the medication should be continued for 2 years after remission in those ‘at risk of relapse’; the risk is defined as ‘two or more episodes, residual symptoms or severe or prolonged episodes’ (© NICE 2010. *Depression: Management of Depression in Primary and Secondary Care*. Available from www.nice.org.uk/guidance/cg23. All rights reserved. Subject to Notice of rights. NICE guidance is prepared for the National Health Service in England. All NICE guidance is subject to regular review and may be updated or withdrawn. NICE accepts no responsibility for the use of its content in this product/publication).²⁰ However, there is no evidence that these proposed factors (number of episodes, residual symptoms and severity of previous episodes) affected the difference between antidepressant and placebo maintenance treatment.¹³ NICE recommends that antidepressant maintenance treatments should continue to be used for 2 years for those at risk of relapse; however, NICE also recognise the uncertainty about the benefit of long-term maintenance treatment and recommend further research into its psychological and pharmacological effects.

At present, there is little evidence to support the use of long-term maintenance antidepressant treatment, despite its widespread use. Given the costs of treatment and medical supervision, the side effects of the medication and patients’ wish to be medication free, it is important to investigate this question further.

Health economic considerations are discussed and presented in *Chapter 4*.

Current evidence on withdrawal symptoms

There is also uncertainty about the frequency and severity of withdrawal symptoms after antidepressants are discontinued. Randomised, double-blind, placebo-controlled trials are required to provide robust evidence on withdrawal symptoms. A recent systematic review of such trials found evidence of withdrawal symptoms after antidepressant discontinuation, but the studies were too heterogeneous to establish the frequency, severity and timing of withdrawal effects. Almost all of these trials were rated as being of low quality because of the high risk of selection bias, attrition and incomplete outcome data.²¹ Withdrawal symptoms are an important risk to evaluate when considering stopping maintenance treatment. They raise an important clinical question of the benefits of antidepressants compared with the possible risks.

Aim and objectives

The ANTLER trial was funded by the National Institute for Health Research (NIHR) as part of its Health Technology Assessment (HTA) programme. The overall aim of the ANTLER trial was to answer the following research question: ‘What is the clinical effectiveness and cost-effectiveness in UK primary care of continuing on long-term maintenance antidepressants compared with a placebo in preventing relapse of depression in those who have taken antidepressants for more than 9 months and who are now well enough to consider stopping maintenance treatment?’.

The trial was pragmatic, embedded in primary care and had broad inclusion criteria to increase the generalisability to the population currently receiving maintenance antidepressants.

Chapter 2 Methods

Trial design

The ANTLER trial was a Phase IV, double-blind, pragmatic, multisite, individually randomised parallel-group controlled trial that was funded by the NIHR HTA programme. We recruited primary care patients who were taking one of four of the most commonly used antidepressant medications. At the point of recruitment, the patients were well enough to consider stopping their medication. Participants were recruited from primary care practices in four UK sites: London, Bristol, Southampton and York.

The trial compared maintenance treatment with antidepressant (citalopram 20 mg, sertraline 100 mg, fluoxetine 20 mg or mirtazapine 30 mg) treatment by replacing the medication with an identical placebo after a tapering period. We chose these antidepressant doses because they are the most common for long-term maintenance treatment in UK primary care and they simplified the manufacture of placebo and the conduct of the trial. In addition, there is no evidence that higher doses increase effectiveness.²² The trial intervention was for 52 weeks and the participants were followed up at 6, 12, 26, 39 and 52 weeks.

Ethics approval and research governance

Ethics approval was obtained from the National Research Ethics Service (NRES) committee, East of England – Cambridge South (reference 16/EE/0032) and the Health Research Authority. A number of subsequent communications were sent to both the NRES and the Health Research Authority either seeking approval for substantial amendments or informing committees of minor changes. Clinical trial authorisation was given by Medicines and Healthcare products Regulatory Agency. The trial sponsor was University College London.

The ANTLER trial was registered on the Current Controlled Trials International Standard Randomised Controlled Trial Number (ISRCTN) registry (ISRCTN15969819; 21 September 2015) and also received a EudraCT number (2015-004210-26). As part of the NIHR Evaluation, Trials and Studies Co-ordinating Centre research portfolio, the trial was adopted and listed on the portfolio.

Participants

The trial recruited patients from 150 general practices across four research sites (London, Bristol, Southampton and York). *Table 1* provides a summary of the main characteristics of the participating practices.

Inclusion criteria

Patients were considered for inclusion if they:

- Had had at least two episodes of depression (because participants find it difficult to remember previous episodes and depressive symptoms are on a continuum, we used a pragmatic approach and considered those who had been treated for over 2 years as having two episodes).
- Were aged 18–74 years (we excluded older people because different assessments for depression are used in the older age groups).
- Had been taking antidepressants for ≥ 9 months and were taking citalopram 20 mg, sertraline 100 mg, fluoxetine 20 mg or mirtazapine 30 mg.

METHODS

- Had satisfactory adherence to medication – the ANTLER trial used a five-item self-report measure of compliance, as adapted for the MIR²³ and CoBaT trials.⁶ Given the relatively long half-life of antidepressant medication, individuals who had forgotten to take 1 or 2 days' worth of medication were excluded, and this was established with an extra question: 'Did you forget to take 2 days of your medication in a row?'. Therefore, the criteria defined people as adherent if they (1) scored zero on the first four questions, (2) scored 1 and said 'no' to the extra question or (3) scored 2 because of 'forget' and 'careless' questions and said 'no' to the extra question.
- Were considering stopping their antidepressant medication.

Comparison of the age and gender of the participants who were invited with those who participated can be found in *Report Supplementary Material 11*.

TABLE 1 Participating general practice characteristics

Characteristic	Category	Per cent of category (n = 150)
Centre	Bristol	20
	London	55
	York	10
	Southampton	15
Geographical location ^a	Urban	85
	Rural	15
List size ^b	1-4999	5
	5000-9999	27
	10,000-14,999	33
	≥ 15,000	35
Number of GPs employed	0-5	24
	6-10	51
	11-15	19
	≥ 16	6
Number of randomised participants	0-4	76
	5-10	19
	11-15	5
Index of Multiple Deprivation ^c	1-10	22
	11-20	43
	21-30	26
	≥ 31	9

GP, general practitioner.

a Based on the 2011 rural-urban classification for output areas in England.

b The number of patients enrolled in the 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.

Exclusion criteria

Patients were excluded if they:

- met internationally agreed (ICD-10) criteria for a depressive illness
- had bipolar disorder, psychotic illness, dementia, alcohol or substance dependence or a terminal illness
- were unable to complete self-administered questionnaires in English
- had contraindications to any of the prescribed medication
- were pregnant or intended to get pregnant with the next 12 months
- were using monoamine oxidase inhibitors
- had allergies to placebo excipients
- were enrolled in another clinical trial of an investigational medicinal product.

The screening questionnaire is in *Report Supplementary Material 1*.

Recruitment

Recruitment started in March 2017. Within 2 years, 478 participants were recruited from general practices across our four research centres using two methods: record search and in-consultation recruitment. *Figure 1* outlines a detailed flow chart of the stages of recruitment.

Record search method

General practitioner (GP) electronic patient records were searched to identify potential participants. These individuals were sent an initial letter and the patient information sheet by the GP surgery, followed by a reminder invitation letter if there was no response. Those patients who replied positively to the invitation letter were reviewed by their GP, who informed the local research team on inclusion/exclusion criteria from the patients' medical notes. The GP could also decide that the person was unsuitable to take part in the trial on any other grounds.

In-consultation recruitment method

General practitioners could introduce the trial to suitable patients at a consultation, give them the patient information sheet to read at home and ask for their permission for release of their contact details to the local trial team. The information was sent by secure nhs.net e-mail or fax to the trial team.

Intervention

Choice of medication

The objective of the ANTLER trial was to provide a valid and generalisable estimate of the clinical effectiveness and cost-effectiveness of long-term maintenance treatment with antidepressants in UK primary care. The trial population was patients who were taking long-term maintenance treatment but who were well enough to consider stopping their antidepressant medication. The intervention that was studied by the ANTLER trial was taking patients off their antidepressant medication rather than starting antidepressant medication.

The choice of medication was guided by the pragmatics of recruitment and carrying out the trial. The ANTLER trial medication was citalopram 20 mg, sertraline 100 mg, fluoxetine 20 mg and mirtazapine 30 mg. We selected these doses because they are the most commonly prescribed doses in primary care (Professor Irene Petersen, University College London, 2013, personal communication; based on The Health Improvement Network electronic health records data). At the time of developing the ANTLER trial protocol in 2013, selective serotonin reuptake inhibitors (SSRIs) were the most commonly prescribed medication, followed by citalopram (32% of antidepressant prescriptions in England after excluding

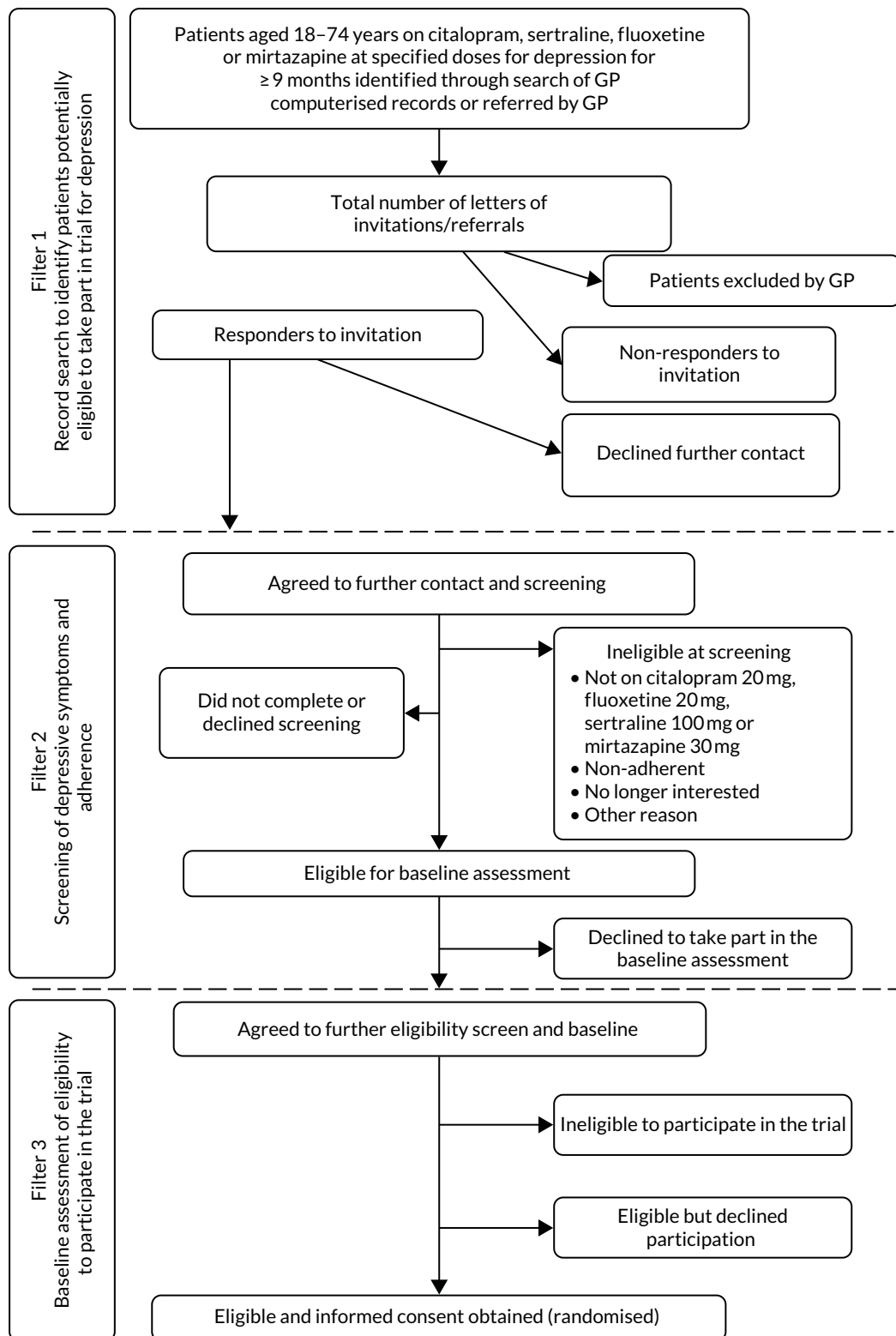


FIGURE 1 Flow chart outlining the stages of recruitment.

amitriptyline, which is mainly used to manage pain and sleep), sertraline (15%) and fluoxetine (14%). Mirtazapine accounted for 13% of prescriptions in England in 2013. Mirtazapine is a different type of antidepressant, a noradrenergic and specific serotonergic antidepressant; however, the net effect of its action is similar to that of SSRIs, to increase serotonergic transmission.

Together, these medications account for about 75% of all long-term antidepressant prescriptions in England (Professor Irene Petersen, personal communication), and all are licensed for the treatment of depression. According to recent data from openprescribing.net,²⁴ these are still the four most commonly prescribed antidepressants.

We excluded escitalopram because it is not widely used in primary care, paroxetine because prescription rates are dropping and it leads to more marked withdrawal symptoms, and venlafaxine because it also causes withdrawal symptoms and most clinical guidelines recommend it as second-line treatment only. We, therefore, recruited patients taking maintenance treatment with citalopram, sertraline, fluoxetine or mirtazapine. These medications account for the vast majority of all antidepressant prescriptions.

Treatment of participants

At baseline, participants were taking citalopram 20 mg, sertraline 100 mg, fluoxetine 20 mg or mirtazapine 30 mg. They were randomised either to remain on their current medication (maintenance group) or to discontinue medication after a tapering period. In the first month, those in the discontinuation group took their usual medication at half of the dose (citalopram 10 mg, sertraline 50 mg or mirtazapine 15 mg). In the second month, they took either half the dose of their usual medication or placebo, on alternate days. From the third month to the end of the trial, they took only the placebo. As fluoxetine is not available as a 10-mg capsule, in the first month those taking fluoxetine at baseline who were allocated to the discontinuation group alternated between a 20-mg capsule and the placebo. During the second month and subsequent months, they took the placebo because fluoxetine has a long half-life.

The active medication was encapsulated and the placebo was an identical capsule filled with an inert excipient. All capsules exactly matched in dimensions and appearance, so that allocation concealment and blinding were maintained.

Subsequent assessments

The follow-up assessments were carried out at 6, 12, 26, 39 and 52 weeks after randomisation. Participants were invited to follow-up appointments unless they had withdrawn from the trial. Participants were followed up even if they stopped taking the trial medication. The follow-up appointments took place at the participant's home, their general practice or university premises. *Figure 2* describes the baseline and follow-up assessments. The baseline questionnaire can be found in *Report Supplementary Material 2*. The 6-week follow-up questionnaire can be found in *Report Supplementary Material 3*. The 12-, 26-, 39- and 52-week follow-up questionnaire can be found in *Report Supplementary Material 4*. The exit questionnaire can be found in *Report Supplementary Material 7*.

Outcomes

Primary outcome

The time in weeks to the beginning of the first episode of depression after randomisation (we call relapse) was measured using the rCIS-R, which is based on the CIS-R,²⁵ which asks about the previous 12 weeks at all follow-up points, except at 6 weeks. Only the five sections (i.e. depression, depressive ideas, concentration, sleep and fatigue) that are used for a depression diagnosis were asked, along with questions that asked about symptoms. Further questions were asked to determine the time to the nearest week when the score was ≥ 2 . The rCIS-R and the precise description of relapse are described further in *Reliability of the retrospective Clinical Interview Schedule – Revised*.

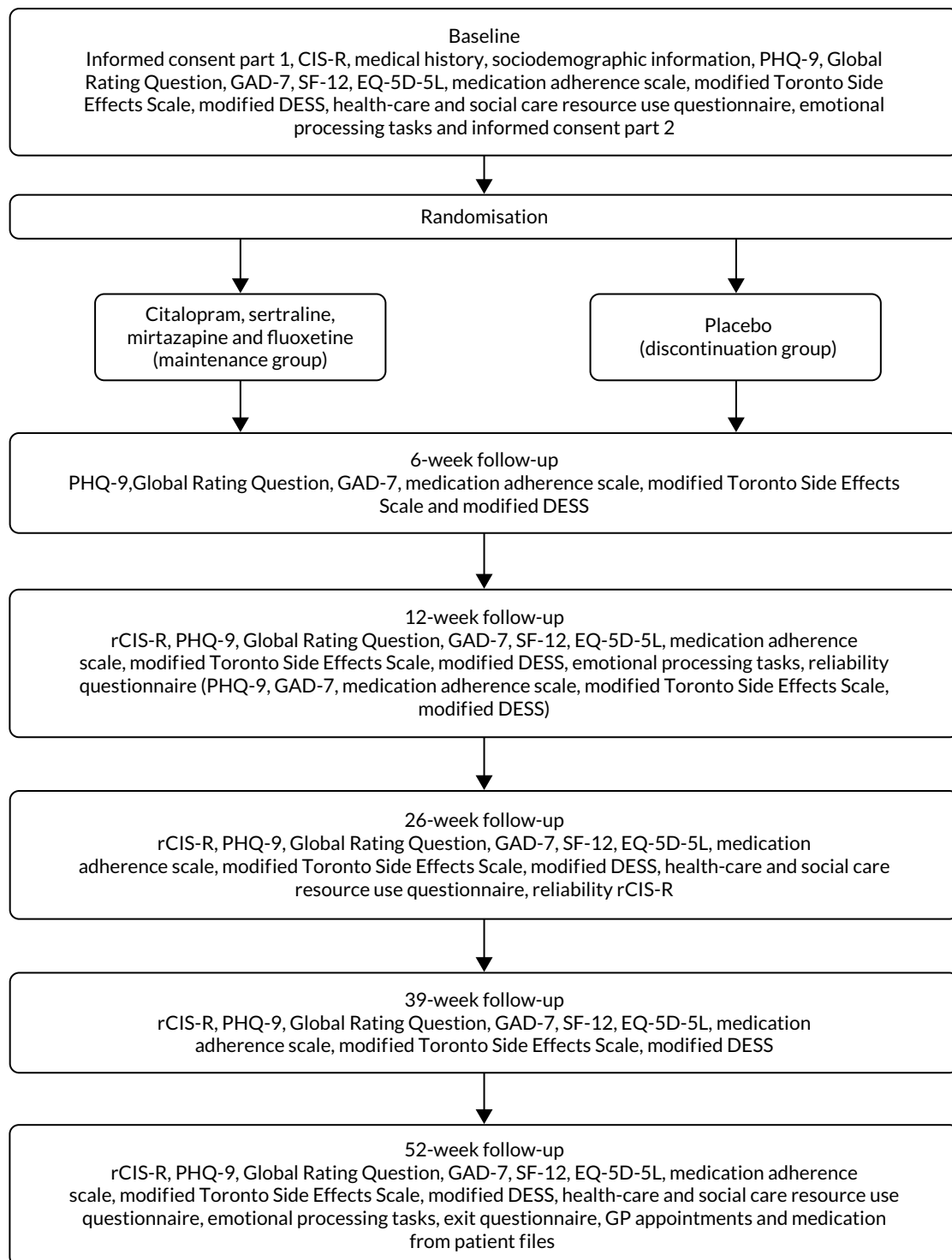


FIGURE 2 Schedule of assessments. DESS, Discontinuation-Emergent Signs and Symptoms; EQ-5D-5L, EuroQol-5 Dimensions, five-level version; GAD-7, Generalised Anxiety Disorder-7; PHQ-9, Patient Health Questionnaire-9 items; SF-12, Short Form questionnaire-12 items.

Secondary outcomes

Depressive symptoms were assessed using the Patient Health Questionnaire-9 items (PHQ-9).^{26,27} This is a nine-item questionnaire. Each item has four responses that range from not at all (0) to nearly every day (3). The score from each item is added to give a total that ranges from 0 to 27. If there are one or two items missing from a participant's questionnaire, the items are replaced by the mean of the items present. If there are more than two items missing, the questionnaire is considered missing for that participant.

Anxiety symptoms were assessed using the Generalised Anxiety Disorder-7 (GAD-7) questionnaire.²⁸ This is a seven-item questionnaire. Each item has four possible responses that range from not at all (0) to nearly every day (3). The score from each item is added to give a total that ranges from 0 to 21. If there are one or two items missing from a participant's questionnaire, the items are replaced by the mean of the items present. If there are more than two items missing, the questionnaire is considered missing for that participant.

The adverse effects of antidepressants were measured using a modified Toronto Side Effects Scale.^{29,30} This is a 13-item measure for males and females and an open-ended item to report any other side effects. Following a consultation with patient groups, we included a question on electric sensations in the brain (brain zaps); this resulted in a 15-item scale. For each item, the scale asks how often the side effect has been present in the past 2 weeks: never, on several days, on more than half of the days or nearly every day. Scores from each item are added to give an overall score between 13 and 52.

Health-related quality of life was measured using the Short Form questionnaire-12 items (SF-12).³¹ The physical and mental component scores were analysed separately.

Withdrawal symptoms were based on Rosenbaum *et al.*;³² participants are asked about the 15 most common symptoms of depression, and a score ranging from 0 to 15 is calculated by summing the number of 'new symptom[s]' and the number of 'old symptom[s] but worse'.

To measure the time to stopping trial medication, the exact date on which the trial medication was stopped is recorded for those who stopped early. For those who completed their course of medication, the date is the date of the last interview or the date that they took the last dose of trial medication, whichever was later.

For the Global Rating Question, participants were asked at baseline and at each subsequent follow-up point, 'Compared to when we last saw you, how have your moods and feelings changed?'. The possible responses were 'I feel a lot better', 'I feel slightly better', 'I feel about the same', 'I feel slightly worse' and 'I feel a lot worse'.³³ We created a dichotomous variable: feeling worse (1) and feeling the same or better (0).

We also examined the test-retest reliability of the PHQ-9, GAD-7, rCIS-R, adverse effects, withdrawal symptoms and adherence questionnaires by asking participants to complete again these questionnaires at one of the follow-up appointments. We included the results of the test-retest reliability of the rCIS-R in *Reliability of the retrospective Clinical Interview Schedule - Revised*.

The measures used in the economic evaluation are discussed in *Chapter 4*.

Mechanistic outcomes

The mechanistic outcomes are not reported in the report or in the main trial paper that reports the clinical outcomes; they will be reported in separate paper(s).³⁴ However, we provide their description below.

Face recognition task

In this task, prototypical 'happy' and 'sad' composite images are generated from 20 individual male faces showing a happy facial expression and the same individuals showing a sad expression from the Karolinska emotional face set (CD ROM available from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, Stockholm), using established techniques.³⁵ These are used as the end points of a linear morph sequence, which consists of 15 images that change in displayed emotion incrementally from unambiguously 'happy', through ambiguity, to unambiguously 'sad'.

The procedure comprises 45 trials, with each stimulus in the sequence presented three times. Images are presented sequentially, in a random order, for 150 ms. Stimuli are preceded by a fixation cross, presented for a random duration ranging from 1500 to 2500 ms. Following presentation, a 250-ms backward mask of visual noise prevents processing of afterimages. Participants are prompted to judge whether the face was 'happy' or 'sad'. As responses change monotonically from one emotion to the other, this allows the calculation of a balance point: the continuum frame at which participants shifted from perceiving primarily happiness to perceiving primarily sadness.

Word recall task

The word recall task³⁶ tests the memory of socially rewarding and socially critical information. The participant is presented with 20 likeable (e.g. cheerful and honest) and 20 dislikeable (e.g. untidy and hostile) personality characteristic words on a laptop screen in a random order for 500 ms. Words are matched according to length, usage frequency and meaningfulness, and they differ at each time point. After each word, participants indicate whether they would 'like' or 'dislike' to hear someone describing them in this way by pressing a key on the keyboard. At the end of the task, participants are asked to recall as many words as possible in 2 minutes. This is a surprise recall task (at baseline) to test incidental memory. The number of positive and negative words accurately recalled (i.e. hits) and the number of false responses (i.e. intrusions) are also recorded.

Go/no-go task

In the go/no-go task,³⁷ each trial includes three events: the presentation of a fractal image, the presentation of a target and the presentation of a probabilistic outcome. At the beginning of each trial, one of four possible fractal images is presented on a computer screen, which indicates whether the best choice in a subsequent target detection task is a go (pressing a key on the keyboard) or a no-go (withholding a response to the target). The fractal also indicates the valence of any outcome dependent on the participant's behaviour (reward/no reward or punishment/no punishment). The meaning of the fractal images (go to win, no-go to win, go to avoid punishment, no-go to avoid punishment) is randomised across the participants, and participants have to learn these by trial and error. Participants are informed that the correct choice for each fractal image is either a go (button press) or a no-go (withhold button press). Actions are required in response to a target circle that follows the fractal image. After a brief delay, the outcome is presented (an upwards arrow indicates a win, a downwards arrow indicates a loss and a horizontal bar indicates the absence of a win or a loss). In go to win trials, a button press is rewarded. In go to avoid punishment, a button press avoids punishment. In no-go to win, withholding a button press is rewarded. In no-go to avoid losing trials, withholding a button press avoids punishment. The task consists of a total of 240 trials (60 trials per condition). The participant could win between £1 and £10.

Table 2 lists the schedule of assessments used in the trial.

Sample size

The sample size estimation was based on the evidence from systematic reviews available when the trial was designed. The reduction in the odds of relapse in an active group compared with a placebo group was estimated to be 70% in a systematic review by Geddes *et al.*,¹¹ 65% by Kaymaz *et al.*¹³ and Glue *et al.*¹² and 50% by NICE.²⁰ Between 15% and 22% of those taking the active drug relapsed in 12 months. To detect the difference between relapse rates of 15% (maintenance group) and 30% (discontinuation group) (hazard ratio 0.46), or between relapse rates of 20% (maintenance group) and 35% (discontinuation group) (hazard ratio 0.52), we estimated that the required sample sizes were 333 and 383 participants, respectively, for 90% power at the 5% significance level. Allowing for 20% attrition, we therefore proposed to recruit 479 participants.³⁸ Analyses are expressed with the discontinuation group as reference, equating to a hazard ratio of 1.92 for the power calculation.

TABLE 2 Full schedule of questionnaires used in the ANTLER trial

Time point	Trial period							Close-out
	Screening	Baseline	Post allocation				After 52 weeks	
	-t ₁	0	6 weeks	12 weeks	26 weeks	39 weeks		52 weeks
Enrolment								
Eligibility screen	X							
Informed consent		X						
Eligibility determination		X						
Randomisation		X						
Intervention								
Sertraline, citalopram, fluoxetine or mirtazapine								
Matching placebo								
Assessments								
Medical history		X						
Sociodemographic information		X						
CIS-R		X						
rCIS-R				X	X	X	X	
PHQ-9	X	X	X	X	X	X	X	
GAD-7		X	X	X	X	X	X	
EQ-5D-5L		X		X	X	X	X	
SF-12		X		X	X	X	X	
Toronto Side Effects Scale (modified)		X	X	X	X	X	X	
Medication adherence	X	X	X	X	X	X	X	
DESS scale (modified)		X	X	X	X	X	X	
Global Rating Question		X	X	X	X	X	X	
Placebo/active question			X	X	X	X	X	
Pill count				X	X	X	X	
Emotional processing		X		X			X	
Health-care and social care resource use		X			X		X	
GP appointments and medication							X	

DESS, Discontinuation-Emergent Signs and Symptoms; EQ-5D-5L, EuroQol-5 Dimensions, five-level version.

Randomisation and blinding

Following completion of the baseline assessment, eligible participants who consented were randomised using the automated randomisation service provided by Sealed Envelope (London, UK; <https://sealedenvelope.com>). The randomisation was minimised by the four study centres, the four medications and the severity of depressive symptoms at baseline (two categories measured using the CIS-R). The dispensing pharmacy (University Hospitals Bristol Pharmacy) was informed of the randomised allocation and posted the medication by recorded delivery to either the participant's home or the GP surgery at 8-week intervals. Trial participants, clinicians and all members of the research team were blinded to the trial treatment allocation. Statisticians analysed the data blind to allocation. Health economists were aware of the allocation so that they could cost the trial medications. Participants were free to withdraw from the medication at any time.

Together with trial medication, participants were posted a contact card so that any treating clinician could be unblinded to treatment allocation in case of a medical emergency (emergency unblinding) or to enable treatment decisions (early unblinding). If unblinding was required, a formal request by a clinician was made to the trial pharmacy (through the 24-hour contact number provided on the contact card) that had a list of the participants' treatment allocations. The treating physician managed the medical emergency as appropriate on receipt of the treatment allocation.

The researcher recorded any breaking of the code and the reasons for doing so in the unblinding log. When possible, members of the research team remained blinded. Those participants who did not require emergency or early unblinding were unblinded on completion of the trial (routine unblinding). This information was provided to their GP by the pharmacy; the participant was encouraged to see their GP to discuss any further treatment during that consultation. The trial team remained blind to this information.

Statistical methods

Primary outcome and secondary analysis

The statistical analysis plan was agreed in advance with the Trial Steering Committee and the independent Data Monitoring Committee (uploaded to <https://discovery.ucl.ac.uk/id/eprint/10089782/>; accessed 27 September 2021) and includes the health economics analysis plan. All statistical analyses were complete case and conducted using intention to treat. That is, data were analysed in the groups that they were randomised to regardless of whether or not participants maintained the allocation that they were randomised to throughout the trial.

The time to depression relapse was analysed using exact Cox proportional hazards modelling. The start date was the date of baseline data collection, and the end date was the date of relapse, date of withdrawal from the trial or date of final follow-up. Participants were asked to identify the number of weeks since the previous assessment that their symptoms began to estimate the date of onset of their relapse. If participants withdrew from the trial and did not provide further data, they were censored at the date of last data collection. The primary model adjusted for the participant's CIS-R depressive symptom score. Sensitivity analyses for the primary outcome included adjusting for the minimisation variables (centre in four categories and antidepressant in four categories, depressive symptoms above or below the median in two categories), using best- and worst-case scenarios for the 10 participants who were not included in the primary analysis. For the best- and worst-case scenarios, those in the maintenance group with missing primary outcome data were censored at the date of last follow-up or withdrawal (good outcome, no relapse) and for those in the discontinuation group with missing primary outcome data, relapse on the day before last follow-up or withdrawal (bad outcome, relapse).

The scales [PHQ-9, GAD-7, SF-12, Toronto Side Effects Scale and modified Discontinuation-Emergent Signs and Symptoms (DESS)] were analysed as if continuous. These secondary outcomes were analysed at each time point separately using mixed-effects linear regression, with two observations per participant: the baseline value and the value from the follow-up point. For these analyses, there was a fixed effect parameter for time and a parameter that was coded as follows: 1 for discontinuation group at follow up and 0 for maintenance group at both times and for discontinuation group at baseline.³⁹ These outcomes were also analysed using available data from all time points in a similar way to the analysis at each time point. The Global Rating Question was analysed using logistic regression at each time point. The time to stopping trial medication was analysed using exact Cox proportional hazards modelling. The start was taken as the latest of receiving trial medication, the date that the participant reported starting the medication or the date of randomisation. The end date was the earliest of the reported date that participants reported stopping taking trial medication or their final follow-up date. The main model included the randomisation variable and the participant's depression CIS-R score. A sensitivity analysis also included the minimisation variables.

For secondary outcomes and time points, we conducted sensitivity analyses that included predictors of missingness identified using univariable logistic regression. For this, the outcomes were whether or not the measure was missing at each time point separately. Baseline variables were considered as possible predictors of missingness. Those that were statistically significant for each outcome and time point were adjusted for using models similar to the main secondary outcome models. For the models including data from all time points, all of the baseline predictors of missingness for the given outcome were included in the model. A predictors of missingness analysis was not carried out on the outcome of time to stopping the ANTLER trial medication given that we had data on whether or not all participants had stopped taking the ANTLER trial medication.

Subgroup analyses were conducted for the outcomes of time to relapse, PHQ-9, GAD-7 and the Global Rating Question. We conducted interactions between the treatment group and the antidepressant medication (dropping mirtazapine because of small numbers), baseline CIS-R depression and anxiety scores, number of previous episodes of depression (dichotomised at two compared with three or more) and the age at which the participant became aware of depression as a continuous measure. The *p*-value for the interaction is reported. In addition, we carried out analyses for each subgroup separately and the coefficient or odds ratio for the treatment group is reported. For this analysis, the age at which the participant became aware of depression was dichotomised at the median. Similar models to the main analyses were used for the subgroup analyses. These were carried out at each time point for PHQ-9, GAD-7 and the Global Rating Question. We also conducted a post hoc analysis, which was requested by a reviewer, to investigate whether or not withdrawal symptoms in the discontinuation compared with the maintenance group differed by antidepressant class.

There were no interim analyses and no predetermined stopping rules.

Reliability of the retrospective Clinical Interview Schedule – Revised

We assessed the onset of a depressive episode and called this assessment rCIS-R and conducted a test-retest reliability study of rCIS-R. The study was nested within the ANTLER trial. The ANTLER trial participants were asked to complete the rCIS-R twice: at the beginning of one of the face-to-face follow-up appointments and at the end of the same appointment.

Why we needed a new measure to assess relapse

One of the methodological challenges of measuring relapse in depression studies has been a lack of clarity in defining relapse and how to differentiate relapse from recurrence. Frank *et al.*⁴⁰ attempted to define the course of depressive illness by defining the use of terms, such as response, remission,

recovery, relapse and recurrence, and offering conceptualisation and operational criteria for each term. However, Frank *et al.*⁴⁰ did not provide the time scale to recovery, leaving uncertainty on when the distinction should apply. Rush *et al.*⁴¹ elaborated further on the distinctions between the terms and proposed defining remission in terms of minimal symptoms over a 3-week duration and defining recovery as having at least 4 months of remission. However, to ensure that the occurrence of recovery has been accurately determined, frequent (i.e. every 2 weeks) assessments must be carried out to detect the return of the index episode. Such an approach is perhaps impractical in clinical practice. The 'minimal symptoms' definition is dependent on the measure used, producing arbitrary definitions. Given that depression is no longer seen as a time-limited disorder with episodes lasting around 4–6 months with full recovery, but is rather thought of as a 'relapsing–remitting' continuum with debilitating symptoms occurring between acute episodes,⁴² we believe that studies assessing the benefit of long-term maintenance treatment need to measure the appearance of any depressive episode and that the proposed distinction between relapse and recurrence is less important. We, therefore, use the term relapse in this report to refer to any new episode of depressive symptoms.

Another issue with assessing relapse is the variety of scales used in medical research. A considerable effort has been put into research of acute treatment; by contrast, there have been relatively few studies investigating long-term maintenance treatments (see *Chapter 1, Current evidence on the effectiveness of maintenance treatment*). To measure relapse, most studies used clinical rating scales, such as the Hamilton Rating Scale for Depression⁴³ or the Montgomery and Åsberg Depression Rating Scale,⁴⁴ at frequent intervals, typically fortnightly. Such scales are prone to observer bias because they are administered by a clinician and measure current symptoms. Self-completed questionnaires, such as the Beck Depression Inventory⁴⁵ and PHQ-9,²⁶ are often used in research in addition to rating scales by clinicians. Although they eliminate observer bias, they can be regarded as crude and might miss some symptoms and/or their intensity owing to patients interpreting the questions in different ways.⁴⁶ They also assess the current symptoms and do not determine the time to relapse.

Fully structured interviews have also been used in research on relapse; they can be administered by lay interviewers, can eliminate observer bias and, therefore, are much more economical. An example is the Composite International Diagnostic Interview (CIDI).⁴⁷ However, the CIDI has over 280 symptom questions that are accompanied by 'probe' questions to assess severity, which makes the interview extremely long, up to 3 hours, and often unacceptable for participants. In addition, the rigid rules of administration and the use of complex flow charts may lead to mistakes by the interviewer in either presenting questions or interpreting participants' responses.⁴⁸ The Structured Clinical Interview Disorder (SCID)⁴⁹ is semistructured, so interviewers need more extensive training in its use, is lengthy (taking between 2 and 6 hours to administer) and requires judgements to be made about the presence of symptoms, thus incurring the risk of introducing observer bias. Although the inter-rater reliability study⁵⁰ of SCID on 151 participants produced fair agreement on the depression scale, the use of audio tapes in this study could have improved the reliability because both raters had access to the same verbal information, but the second rater did not have non-verbal information and any interviewer-related measurement error would not have been included.

A simpler structured interview that is considerably shorter and assesses the symptoms in the last 12 weeks would be a better option to accurately assess the symptoms.

The aim of the reliability study

The aim of this study was to assess the test–retest reliability of the rCIS-R. We developed a simple measure that can be used to diagnose the reappearance of depressive symptoms after recovery. The rCIS-R is a new assessment that is based on the CIS-R,²⁵ a validated measure that has been widely used by researchers to assess the severity and duration of depression; however, the CIS-R asks about symptoms in the last 7 days. Therefore, we adapted it to assess the symptoms in the last 12 weeks in a fully structured format, that, to our knowledge, has not been carried out for other existing measures.

The measure

The rCIS-R was designed as a self-administered computerised questionnaire and asked about the previous 12 weeks at each follow-up point: 12, 26, 39 and 52 weeks. Only five sections (i.e. depressive mood, depressive ideas, concentration, sleep and fatigue) that are used for a depression diagnosis are asked, along with questions asking about the duration of the symptoms and the intensity of symptoms during the worst week, and questions establishing the start of the symptom(s) in the last 12 weeks. The rCIS-R begins with the two overarching mandatory questions for the depressive mood and depressive ideas sections. To progress further, the participant is required either to answer 'yes' to the first mandatory depression question, 'Almost everyone becomes low in mood or depressed at times. Has there been a time in the past three months when you had a spell of feeling sad, miserable or depressed?', or to answer 'no' to the second mandatory question, 'In the past three months, have you been able to enjoy or take an interest in things as much as you usually do?'. If the participant's answers indicate that they have experienced either low mood or anhedonia in the last 12 weeks, they are asked about the duration to establish that symptoms have been present for ≥ 2 weeks and the time when they started feeling depressed. If the symptom(s) have been present for ≥ 2 weeks, the participant is considered to be positive for that symptom and is asked 10 additional questions covering depressive symptoms during the worst week (e.g. feeling low for prolonged periods, unresponsiveness of mood, loss of sexual interest, restlessness, decreased cognitive function, feeling of guilt, lower self-esteem, hopelessness, feeling life is not worth living and suicidal thoughts).

The other three sections of the rCIS-R (i.e. concentration, sleep and fatigue) are similar in structure and they also start with mandatory question(s). If the participant's answer to the mandatory question(s) indicates that they have not experienced such symptoms, the extra questions relating to the severity of the symptom are not asked and the participant skips to the mandatory question in the next section of the rCIS-R. If the participant's answer indicates that they have experienced the symptom, they are considered positive for that section and further questions about their experience during the worst week are also asked. It is possible to score a maximum of 3 on the concentration and fatigue sections and 4 on the sleep section.

Box 1 shows the concentration section as an example of a section from the rCIS-R. The first two questions are the mandatory questions and if the answer is 'yes' to at least one, the other three questions are asked.

The assessment takes approximately 5 minutes to complete; however, if the participant does not have any symptoms, it takes as little as 2 minutes.

Testing the algorithm

A relapse of depression was defined as experiencing two or more depressive symptoms from any of the five sections during the worst week in the past 3 months (this must include at least one of the two overarching mandatory questions on depressive mood or anhedonia for ≥ 2 weeks) on the rCIS-R.

We also defined relapse in line with ICD-10 criteria and investigated the number of participants experiencing four or more depressive symptoms. In addition to defining a binary outcome of relapse, rCIS-R generates a total score for the depressive episode that occurred in the previous 12 weeks. Each section generates a maximum score between 3 and 6; higher scores indicate more symptoms and the total score can range from 0 to 21.

In September 2018, before the statistical analysis plan was finalised, we looked at some preliminary data from the 12-week follow-up. Data were available for 157 ANTLER trial participants and we used this subset to test the algorithm. Using the above definition for relapse, 29 participants (18%) were identified as having relapsed in depression by their 12-week follow-up appointment.

BOX 1 Concentration section from the rCIS-R

Has there been a period of time in the PAST THREE MONTHS, when you had problems in concentrating on what you were doing?

Has there been a period of time in the PAST THREE MONTHS, when you noticed any problems with forgetting things?

During the worst WEEK in the past three months:

1. Could you concentrate on all of the following without your mind wandering?

- A whole TV programme*
- A newspaper article*
- Talking to someone*

- 1. Yes, I could concentrate on all of them*
- 2. No, I couldn't concentrate on at least one of these things*

2. Did problems with your concentration STOP you from getting on with things you used to do or would like to do?

- 1. No*
- 2. Yes*

3. Did you forget anything important?

- 1. No*
- 2. Yes, I did forget something important*

We also explored if the number of relapses changed if the algorithm included more symptoms, so we tested the following algorithms:

- participant experienced either low mood or anhedonia for ≥ 2 weeks in the past 3 months and was positive on two of any of the five sections (depressive mood, depressive thoughts, fatigue, concentration or sleep)
- participant experienced either low mood or anhedonia for ≥ 2 weeks in the past 3 months and was positive on three of any of the five sections (depressive mood, depressive thoughts, fatigue, concentration or sleep)
- participant experienced either low mood or anhedonia for ≥ 2 weeks in the past 3 months and was positive on four of any of the five sections (depressive mood, depressive thoughts, fatigue, concentration or sleep).

The first two algorithms generated the same number of cases as the original algorithm (29 cases). However, the last algorithm generated 25 cases. We concluded that a score of ≥ 2 on the five sections of depression (including at least one of the two mandatory questions) accurately identifies participants who have relapsed and, therefore, we adopted this case definition as our primary outcome.

Analysis

The level of agreement between the first and the second completion of the rCIS-R was assessed using kappa (quadratic weighted and unweighted) statistics. Quadratic weighted and unweighted kappa

produced very similar results. Given that weighted kappa provides a ratio-scale degree of disagreement to each cell of the $\kappa \times \kappa$ table, making weighted kappa suitable as a measure of agreement. The test-retest reliability was also assessed by using a Bland–Altman plot.⁵¹ We also calculated the intraclass correlation coefficient using a single-measurement, absolute agreement, two-way mixed-effects model.

Results of the reliability study of the retrospective Clinical Interview Schedule – Revised

Of 478 participants who were recruited to the trial, 396 completed the rCIS-R twice. Two participants completed the rCIS-R at the 12-week follow-up appointment, 335 participants at 26 weeks, 42 participants at 39 weeks and 17 participants at 52 weeks. There were 106 male participants, the mean age was 55 years (SD 6 years) and 6% of the participants reported being from an ethnic minority group. The full description of the reliability study sample characteristics compared with the trial sample is in Table 3.

The Cohen's kappa for relapse in depression was 0.84 [95% confidence interval (CI) 0.71 to 0.97], which indicates excellent agreement between the first and the second completions of the rCIS-R (Tables 4 and 5). The level of agreement of the individual sections of the rCIS-R was also excellent (see Table 4). The agreement of the time of depression relapse was also assessed: both agreement of month $\kappa = 0.84$ (95% CI 0.71 to 0.97) and agreement of week $\kappa = 0.87$ (95% CI 0.74 to 1.00) of reappearance of depression.

TABLE 3 Baseline characteristics of the sample who completed the rCIS-R twice at the same follow-up appointment

Characteristic	Reliability study sample	Trial sample
Age (years), mean (SD)	55 (6)	54 (6)
Male, n/N (%)	106/396 (27)	128/478 (27)
Ethnicity, n/N (%)		
White	373/396 (94)	449/473 (95)
Not white	23/396 (6)	25/473 (5)
Highest educational qualification, n/N (%)		
Degree/higher degree	146/396 (37)	179/472 (38)
Diploma/A Levels or equivalent	127/396 (32)	148/472 (31)
GCSE or equivalent/other/none	123/396 (31)	145/472 (31)
Site, n/N (%)		
London	170/396 (43)	199/478 (42)
Bristol	80/396 (20)	102/478 (21)
Southampton	84/396 (21)	96/478 (20)
York	62/396 (16)	81/478 (17)
Antidepressant, n/N (%)		
Sertraline	62/396 (16)	78/478 (16)
Citalopram	183/396 (46)	223/478 (47)
Fluoxetine	135/396 (34)	160/478 (33)
Mirtazapine	16/396 (4)	17/478 (4)
CIS-R score at baseline, mean (SD)	5.0 (4.8)	5.1 (4.8)
PHQ-9 score at baseline, mean (SD)	3.8 (3.6)	3.8 (3.5)
Age first became aware of having depression (years), mean (SD)	32 (5)	32 (15)

A Level, Advanced Level; GCSE, General Certificate of Secondary Education.

TABLE 4 Level of agreement of relapse in depression and individual symptoms between the first and the second completions of the rCIS-R

	Frequency present at the first completion (%)	Weighted Cohen's kappa (95% CI)
Relapse	20	0.84 (0.71 to 0.97)
Symptoms		
Depression or depressive mood	54	0.87 (0.77 to 0.97)
Depressive thoughts	50	0.87 (0.77 to 0.87)
Fatigue	56	0.85 (0.75 to 0.95)
Concentration	25	0.81 (0.72 to 0.91)
Sleep	52	0.91 (0.82 to 1.00)

TABLE 5 Table indicating agreement for relapse in depression between the first and the second completion of the rCIS-R

Relapsed on the second completion	Relapse on the completion (n)		Total (n)
	Did not relapse	Relapsed	
Did not relapse	301	18	319
Relapsed	15	62	77
Total	316	80	396

The mean score for the first completion of the rCIS-R was 6.67 (SD 5.06) and the mean score for the second completion was 6.41 (SD 5.25). The percentage of participants meeting the relapse criteria at the first completion was 20% ($n = 80$) and at the second completion was 19% ($n = 77$). The mean total score difference was -0.25 (95% CI -0.43 to -0.07).

The Bland-Altman plot (Figure 3) shows the agreement between the first and the second completion of the rCIS-R. The intraclass correlation coefficient was 0.94 (95% CI 0.92 to 0.95).

We also looked at defining relapse in line with ICD-10 criteria: 18% of participants relapsed according to the ICD-10 diagnostic criteria for depression at the first ($n = 72$) and second ($n = 70$) completions. Cohen's kappa for relapse of depression according to ICD-10 criteria was 0.74 (95% CI 0.64 to 0.84).

Public and patient involvement

Paul Lanham, our public and patient involvement (PPI) representative, is a co-applicant of the proposal and has been involved in all stages of the trial for almost 5 years. He is a former chairperson (1988–93) and director (1986–2009) of Depression Alliance [now merged with Mind (London, UK)] and provided input to previous studies on depression funded by the HTA programme (TREAD,⁵² CoBaIT⁶ and PREVENT⁵³).

In the development of the proposal, Paul Lanham was supportive of examination of the effectiveness of maintenance medication and wrote 'From a user point of view this strikes me as being vitally important and I am delighted to be involved in it; I am sure that others (especially sufferers and GPs) will also welcome it and gain a great deal from its results' (reproduced with permission from Paul Lanham, London, May 2014, personal communication). Paul is currently on maintenance antidepressants himself and added 'I mistrust terms like "remission", "well", "normal" etc. People ask me if Citalopram helps; the answer is that I have no idea of what I would be like without it so I cannot judge. This [study]

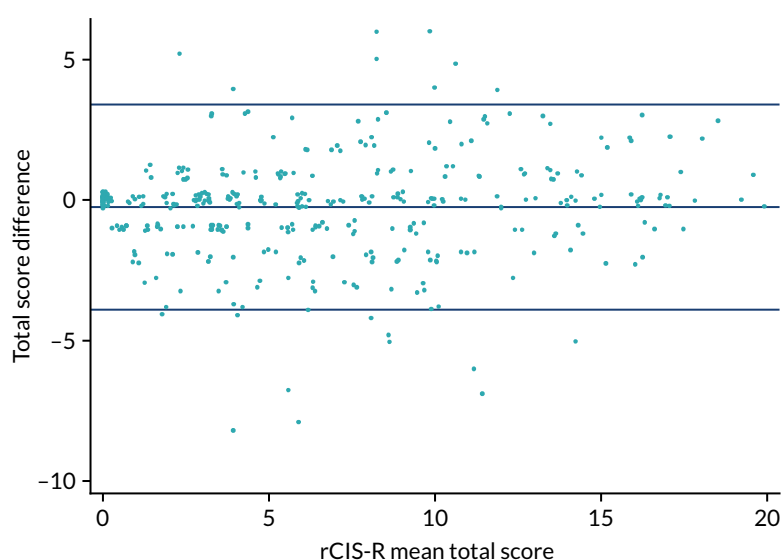


FIGURE 3 Difference in the total rCIS-R score against the mean total score.

will hopefully determine whether such drugs are a help or not. It would be really valuable to know the answer to this through the study' (reproduced with permission from Paul Lanham, personal communication). Paul was a co-applicant on the HTA proposal and a co-author on the final report.

During the trial, Paul Lanham was a member of the Trial Steering Group and the Trial Management Group, but he decided once the trial was established that his contribution was not required. In 2017, we recruited another PPI member, Lucy Carr, who was a former participant in a NIHR-funded trial, PANDA.³³ She was an independent member of the Trial Steering Committee and contributed, along with Paul Lanham, to the design and content of the study documentation, including patient information sheets and self-harm protocols.

In addition, we enlisted the support of the North London Service Users Research Forum (SURF). North London SURF was co-founded in 2007 by service users and clinical academic psychiatrists at University College London to provide meaningful consultation on research. It has 12 members who have mental health problems. Since 2007, it has been consulted on over 100 projects, and North London SURF members have also been invited to join steering/management groups on many of these. As a result, the group is very experienced and confident about the advice and input that it provides; members' comments on the trial paperwork have been invaluable. The letter templates, patient information sheets and questionnaires were amended to reflect the North London SURF feedback. We also consulted on the protocol concerning self-harm or risk of self-harm, which was used if patients reported this in the course of the trial.

We also contacted Luke Montagu, who has experienced withdrawal symptoms from antidepressants and is a well-known activist in this area. Luke Montagu, along with others with lived experience, helped us modify the DESS scale measure of withdrawal symptoms. We shortened the scale (from 42 to 15 items) to improve its acceptability and selected the most commonly found items using published literature and a survey of Luke Montagu's contacts. After consulting with the group, we included a question on electric sensations in the brain (brain zaps), apparently a very common, disabling symptom that has been poorly understood and not included in previous scales. Close involvement of the PPI for the duration of the trial has been invaluable for its success from the set-up stage onwards, shaping the study through to the discussion on the interpretation of results. We plan to carry on using the services users' involvement in the design, documentation and analysis of any future studies.

Chapter 3 Results

Flow of participants in the trial

Recruitment began on 9 March 2017 and the last participant was randomised on 1 March 2019. The flow of participants through the trial is shown in the Consolidated Standards of Reporting Trials (CONSORT) flow diagram (*Figure 4*). The GP record search identified 23,429 potentially eligible patients, who were sent an invitation letter. Another 124 potentially eligible patients were referred during GP consultation, resulting in 1466 patients wanting to take part. Of these patients, 606 were eligible. A total of 478 participants were randomised: 238 to the maintenance group and 240 to the discontinuation group. All of the participants provided data on whether or not they relapsed; however, 10 participants (maintenance group, $n = 6$; discontinuation group, $n = 4$) did not provide data on timing of relapse, so could not be included in the analysis of the primary outcome.

Baseline comparability

The baseline characteristics of the sample overall and by treatment group are shown in *Table 6*. The treatment groups were well balanced at baseline, with just over one-quarter of male participants and a mean age of 54 years (SD 13 years) in the maintenance group and 55 years (SD 12 years) in the discontinuation group. Just over 40% of participants were recruited from London, 20% from each of Bristol and Southampton, and 17% from York. Under half of the participants were taking citalopram, one-third fluoxetine, one-sixth sertraline and < 5% mirtazapine. Almost three-quarters of the participants had taken antidepressants for more than 3 years, with over one-third taking them for ≥ 6 years. The mean age at becoming aware of having depression was 33 years (SD 16 years) in the maintenance group and 32 years (SD 14 years) in the discontinuation group. The median time between randomisation and taking the trial medication was 9 days (interquartile range 6–13 days) in the maintenance group and 8 days (interquartile range 6–13 days) in the discontinuation group.

Primary outcome

The time to relapse was shorter in participants who discontinued antidepressants than in those who stayed on maintenance treatment [hazard ratio (HR) 2.06, 95% CI 1.56 to 2.70; $p < 0.0001$] (*Table 7* and *Figure 5*). This result was unaltered after sensitivity analyses.

Overall, relapse was experienced by 39% ($n = 92/238$) (95% CI 32% to 45%) of participants in the maintenance group and 56% ($n = 135/240$) (95% CI 50% to 63%) of participants in the discontinuation group by the end of the trial (52 weeks). Relapses according to treatment group are shown in the tree diagrams (see *Report Supplementary Material 9*).

Secondary outcomes

The PHQ-9 and GAD-7 scores were higher (worse) in the discontinuation group than in the maintenance group at 12 and 26 weeks, the highest of which for both was at 12 weeks (PHQ-9: coefficient 2.16, 95% CI 1.47 to 2.84; GAD-7: coefficient 2.40, 95% CI 1.81 to 2.99).

RESULTS

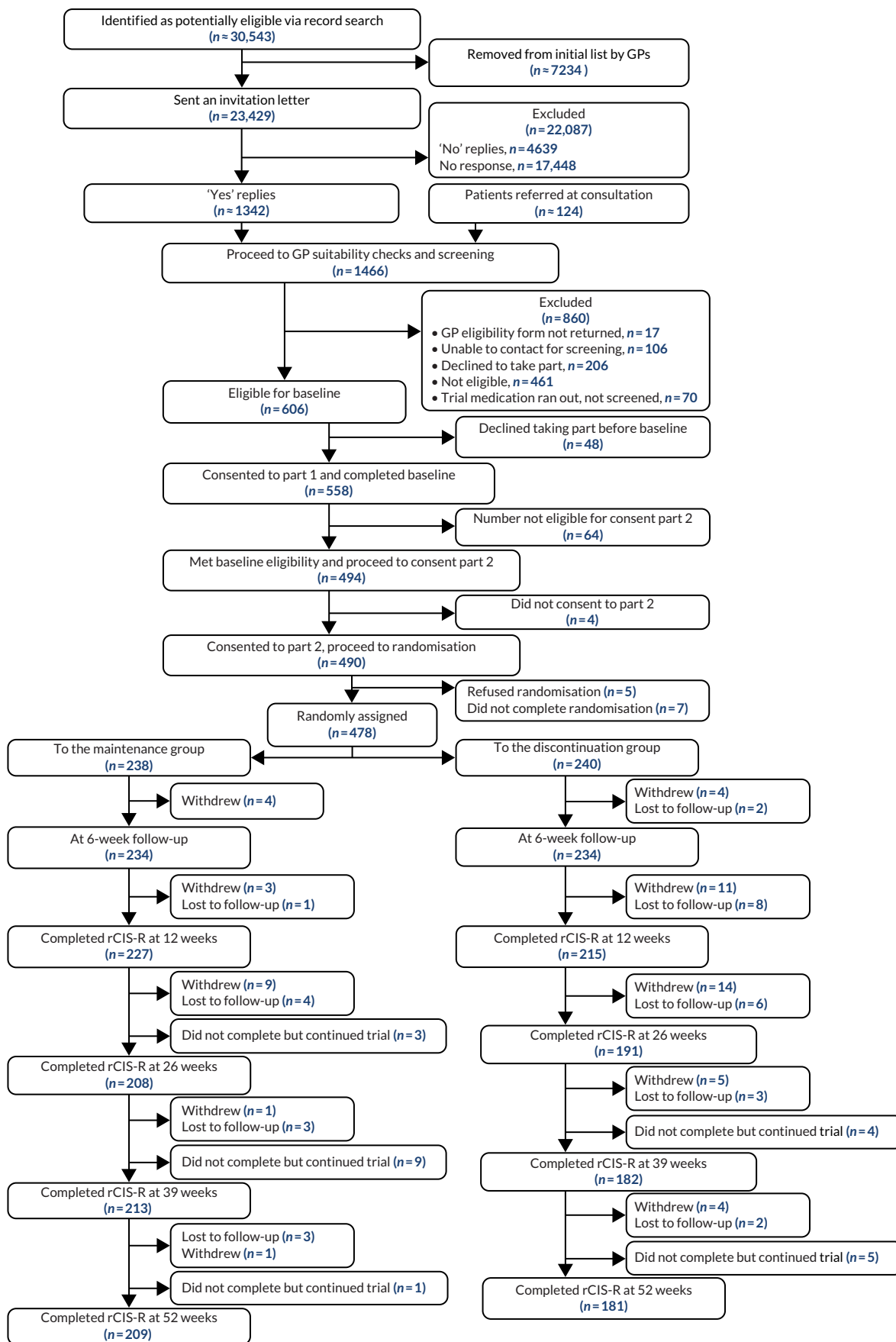


FIGURE 4 The CONSORT flow diagram for the ANTLER trial.

TABLE 6 Baseline demographic and clinical characteristics by treatment group

Characteristic	Treatment group	
	Maintenance	Discontinuation
Age (years), mean (SD)	54 (13)	55 (12)
Male, n/N (%)	70/238 (29)	59/240 (25)
Ethnicity, n/N (%)		
White	221/238 (93)	228/235 (97)
Not white	17/238 (7)	7/235 (3)
Marital status, n/N (%)		
Married	146/238 (61)	161/240 (67)
Single	35/238 (15)	26/240 (11)
Separated or divorced	39/238 (16)	33/240 (14)
Widowed	18/238 (8)	20/240 (8)
Employment status, n/N (%)		
Employed	140/238 (59)	152/240 (63)
Retired	71/238 (30)	68/240 (28)
Other	27/238 (11)	20/240 (8)
Site, n/N (%)		
London	101/238 (42)	98/240 (41)
Bristol	48/238 (20)	54/240 (23)
Southampton	48/238 (20)	48/240 (20)
York	41/238 (17)	40/240 (17)
Antidepressant, n/N (%)		
Sertraline	41/238 (17)	37/240 (15)
Citalopram	111/238 (47)	112/240 (47)
Fluoxetine	77/238 (32)	83/240 (35)
Mirtazapine	9/238 (4)	8/240 (3)
CIS-R above the median, ^a n/N (%)	116/237 (49)	110/240 (46)
Age first became aware of having depression (years), mean (SD)	33 (16)	32 (14)
Three or more previous episodes of depression, n/N (%)	224/238 (94)	219/239 (92)
Time taking antidepressants, n/N (%) ^{ab}		
9 months to < 1 year	12/238 (5)	18/239 (8)
1 to 2 years	56/238 (24)	53/239 (22)
3 to 5 years	81/238 (34)	79/239 (33)
6 to 10 years	48/238 (20)	48/239 (20)
≥ 11 years	41/238 (17)	41/239 (17)

continued

RESULTS

TABLE 6 Baseline demographic and clinical characteristics by treatment group (continued)

Characteristic	Treatment group	
	Maintenance	Discontinuation
Courses of antidepressants in the past, n/N (%)		
0	102/238 (43)	92/239 (38)
1	40/238 (17)	39/239 (16)
≥ 2	96/238 (40)	108/239 (45)
Taking other psychotropic medication, n/N (%)		
Diazepam or lorazepam	3/238 (1)	3/240 (1)
Zopiclone or zolpidem	5/238 (2)	2/240 (0.8)
Using psychotherapy, n/N (%)	24/237 (10)	18/239 (8)
PHQ-9, mean (SD) ^c	3.9 (3.5)	3.8 (3.6)
GAD-7, mean (SD) ^d	3.2 (3.1)	2.8 (3.0)
SF-12 physical, mean (SD) ^e	48 (11)	50 (9)
SF-12 mental, mean (SD) ^e	47 (9)	48 (9)
Modified Toronto Side Effects Scale, mean (SD) ^f	4.2 (2.7)	3.7 (2.7)
At least one symptom on the Toronto Side Effects Scale, n/N (%)	217/235 (92)	218/239 (91)
Number of new or worsening symptoms using modified DESS scale, mean (SD) ^g	1.0 (1.4)	0.6 (1.0)
At least one new or worsening symptom using modified DESS scale, n/N (%) ^g	118/238 (50)	95/240 (40)
Mood worse than 2 weeks ago, n/N (%)	13/237 (5)	9/239 (4)
<p>a Range 0–57. b Without a break of ≥ 2 weeks (including changed antidepressants without taking a break); 26% in the maintenance group and 43% in the discontinuation group. c Range 0–27. d Range 0–21. e Range 0–100. f Range 0–13. g Range 0–15.</p> <p>From <i>N Engl J Med</i>, Lewis G, Marston L, Duffy L, Freemantle N, Gilbody S, Hunter R, et al., Maintenance or discontinuation of antidepressants in primary care, vol. 385, pp. 1275–67.³⁴ Copyright © 2021 Massachusetts Medical Society. Reprinted with permission.</p>		

TABLE 7 Hazard ratios for relapse in the discontinuation group compared with the maintenance group

Outcome	HR (95% CI)
Time to first depression relapse (N = 468)	2.06 (1.56 to 2.70)
Time to first depression relapse, including minimisation variables (N = 468) ^a	2.07 (1.57 to 2.72)
Time to first depression relapse, good outcome in control and bad outcome in intervention (N = 478) ^b	2.12 (1.61 to 2.78)
<p>a Study centre (four groups), antidepressant medication (four groups) and severity of depressive symptoms (above or below the median, two groups). b Maintenance group: censored at the date of last follow-up or withdrawal (good outcome, no relapse). Discontinuation group: relapse on the day before last follow-up or withdrawal (bad outcome, relapse). From <i>N Engl J Med</i>, Lewis G, Marston L, Duffy L, Freemantle N, Gilbody S, Hunter R, et al., Maintenance or discontinuation of antidepressants in primary care, vol. 385, pp. 1275–67.³⁴ Copyright © 2021 Massachusetts Medical Society. Reprinted with permission.</p>	

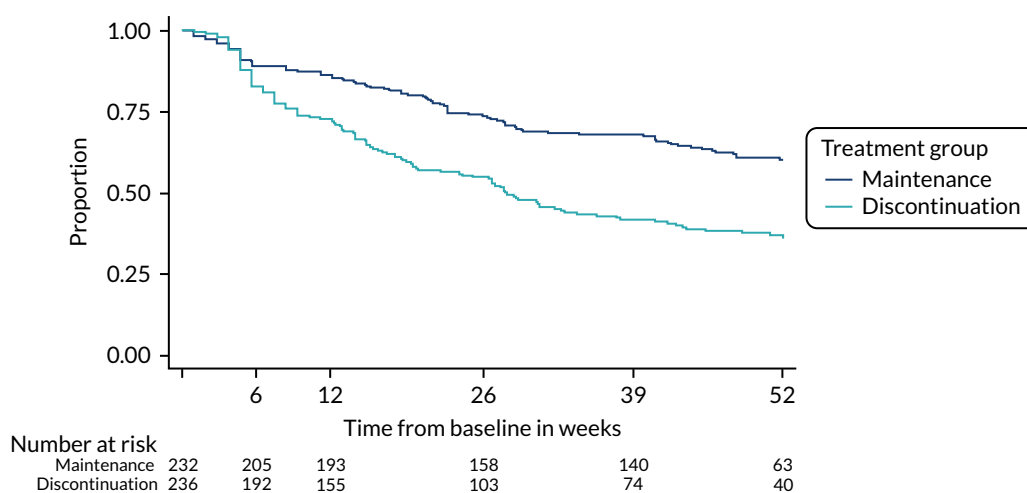


FIGURE 5 Kaplan-Meier plot for the primary outcome. From *N Engl J Med*, Lewis G, Marston L, Duffy L, Freemantle N, Gilbody S, Hunter R, *et al.*, Maintenance or discontinuation of antidepressants in primary care, vol. 385, pp. 1275–67.³⁴ Copyright © 2021 Massachusetts Medical Society. Reprinted with permission.

The number of withdrawal symptoms reported was higher in the discontinuation group than in the maintenance group at 6, 12, 26 and 39 weeks, with the difference largest at 12 weeks (coefficient 1.87, 95% CI 1.46 to 2.28).

Participants in the discontinuation group had lower (worse) SF-12 mental health-related quality-of-life scores at 12, 26 and 39 weeks, with the difference largest at 12 weeks (coefficient –4.86, 95% CI –6.44 to –3.29). At 12 weeks, in the discontinuation group, the odds of feeling worse, determined using the Global Rating Question, was more than twice that (odds ratio 2.88, 95% CI 1.90 to 4.38) in the maintenance group (Table 8).

TABLE 8 Means and estimated mean difference for secondary outcomes in the discontinuation group compared with the maintenance group

Outcome	Treatment group		Estimate (95% CI)
	Maintenance	Discontinuation	
PHQ-9 (coefficient),^a mean (SD)			
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 to 0.87)
12 weeks (n = 477)	4.1 (3.8)	6.3 (5.1)	2.16 (1.47 to 2.84)
26 weeks (n = 477)	4.2 (3.7)	5.0 (4.6)	0.72 (0.02 to 1.42)
39 weeks (n = 477)	3.8 (3.9)	4.4 (4.2)	0.55 (–0.14 to 1.24)
52 weeks (n = 477)	3.7 (3.7)	4.0 (4.5)	0.38 (–0.32 to 1.07)
Over all time points (n = 478)			0.84 (0.38 to 1.29)
GAD-7 (coefficient),^b mean (SD)			
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 to 1.03)
12 weeks (n = 477)	3.1 (3.3)	5.3 (4.6)	2.40 (1.81 to 2.99)
26 weeks (n = 477)	3.4 (3.8)	4.1 (4.4)	0.79 (0.13 to 1.45)

continued

RESULTS

TABLE 8 Means and estimated mean difference for secondary outcomes in the discontinuation group compared with the maintenance group (continued)

Outcome	Treatment group		Estimate (95% CI)
	Maintenance	Discontinuation	
39 weeks (n = 477)	2.9 (3.5)	3.8 (4.1)	0.99 (0.36 to 1.62)
52 weeks (n = 477)	3.0 (3.7)	3.1 (3.0)	0.27 (-0.36 to 0.89)
Over all time points (n = 478)			1.00 (0.58 to 1.42)
Modified Toronto Side Effects Scale (coefficient),^c mean (SD)			
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 to 0.92)
12 weeks (n = 477)	4.2 (2.9)	4.6 (3.0)	0.68 (0.25 to 1.11)
26 weeks (n = 477)	4.0 (2.6)	3.9 (2.8)	0.20 (-0.26 to 0.66)
39 weeks (n = 476)	3.8 (2.5)	3.7 (2.6)	0.16 (-0.30 to 0.62)
52 weeks (n = 475)	3.7 (2.6)	3.5 (2.8)	0.04 (-0.41 to 0.49)
Over all time points (n = 478)			0.36 (0.06 to 0.65)
Number of new or worsening symptoms using modified DESS scale (coefficient),^d mean (SD)			
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 to 0.84)
12 weeks (n = 478)	1.3 (2.4)	3.1 (3.5)	1.87 (1.46 to 2.28)
26 weeks (n = 478)	1.4 (2.3)	1.9 (2.9)	0.50 (0.12 to 0.89)
39 weeks (n = 478)	0.8 (1.6)	1.7 (2.7)	0.94 (0.60 to 1.28)
52 weeks (n = 478)	0.8 (1.8)	1.1 (2.5)	0.32 (-0.02 to 0.65)
Over all time points (n = 478)			0.86 (0.62 to 1.11)
SF-12 physical (coefficient),^e mean (SD)			
Baseline	48 (11)	50 (9)	
12 weeks (n = 476)	48 (10)	50 (9)	0.44 (-0.91 to 1.78)
26 weeks (n = 476)	48 (10)	49 (10)	0.15 (-1.33 to 1.62)
39 weeks (n = 476)	48 (11)	51 (10)	1.49 (-0.06 to 3.04)
52 weeks (n = 476)	49 (10)	49 (11)	-0.59 (-2.09 to 0.92)
Over all time points (n = 476)			0.44 (-0.60 to 1.48)
SF-12 mental (coefficient),^e mean (SD)			
Baseline	47 (9)	48 (9)	
12 weeks (n = 476)	46 (10)	41 (11)	-4.86 (-6.44 to -3.29)
26 weeks (n = 476)	46 (11)	44 (11)	-2.56 (-4.35 to -0.77)
39 weeks (n = 476)	48 (10)	45 (11)	-3.07 (-4.84 to -1.31)
52 weeks (n = 476)	47 (10)	46 (11)	-1.59 (-3.43 to 0.25)
Over all time points (n = 476)			-3.02 (-4.23 to -1.81)

TABLE 8 Means and estimated mean difference for secondary outcomes in the discontinuation group compared with the maintenance group (continued)

Outcome	Treatment group		Estimate (95% CI)
	Maintenance	Discontinuation	
Global Rating Question (OR), 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
Feeling worse	41/223 (18)	41/223 (18)	1.00 (0.62 to 1.61)
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 to 4.38)
26 weeks (N = 403)			
Feeling the same or better	164/210 (78)	151/193 (78)	1.00
Feeling worse	46/210 (22)	42/193 (22)	0.99 (0.62 to 1.59)
39 weeks (N = 396)			
Feeling the same or better	185/212 (87)	153/184 (83)	1.00
Feeling worse	27/212 (13)	31/184 (17)	1.39 (0.79 to 2.43)
52 weeks (N = 391)			
Feeling the same or better	181/210 (86)	154/181 (85)	1.00
Feeling worse	29/210 (14)	27/181 (15)	1.09 (0.62 to 1.93)
Time to stopping trial medication (HR) (N = 477)			2.28 (1.68 to 3.08)
Time to stopping trial medication including minimisation variables (HR) (N = 477)			2.39 (1.76 to 3.24)

OR, odds ratio.

a Range 0–27.

b Range 0–21.

c Range 0–13.

d Range 0–15.

e Range 0–100.

Note

N is number of participants included in the models.

From *N Engl J Med*, Lewis G, Marston L, Duffy L, Freemantle N, Gilbody S, Hunter R, *et al.*, Maintenance or discontinuation of antidepressants in primary care, vol. 385, pp. 1275–67.³⁴ Copyright © 2021 Massachusetts Medical Society. Reprinted with permission.

Missing data

The results of the predictors of missingness analysis are in *Report Supplementary Material 12*. We conducted sensitivity analyses that included predictors of missingness for continuous secondary outcomes. Results for all outcomes were similar to the main analyses (see *Table 8*) when including predictors of missingness in the models (*Table 9*).

RESULTS

TABLE 9 Secondary outcomes controlling for factors associated with missingness

Outcome	Estimate (95% CI)
PHQ-9 (coefficient)^a	
6 weeks	0.31 (-0.25 to 0.88)
12 weeks	
26 weeks	0.80 (0.05 to 1.56)
39 weeks	0.64 (-0.05 to 1.33)
52 weeks	0.38 (-0.32 to 1.08)
Over all time points	1.02 (0.65 to 1.39)
GAD-7 (coefficient)^b	
6 weeks	0.50 (-0.03 to 1.03)
12 weeks	
26 weeks	1.13 (0.42 to 1.85)
39 weeks	1.03 (0.41 to 1.67)
52 weeks	0.28 (-0.34 to 0.91)
Over all time points	1.19 (0.74 to 1.64)
Toronto Side Effects Scale (coefficient)^c	
6 weeks	0.54 (0.16 to 0.92)
12 weeks	
26 weeks	0.25 (-0.28 to 0.77)
39 weeks	0.21 (-0.24 to 0.66)
52 weeks	0.09 (-0.34 to 0.53)
Over all time points	0.47 (0.16 to 0.77)
Modified DESS scale new or worsening symptoms (coefficient)^d	
6 weeks	0.51 (0.18 to 0.84)
12 weeks	
26 weeks	0.53 (0.12 to 0.95)
39 weeks	0.95 (0.61 to 1.29)
52 weeks	0.31 (-0.02 to 0.65)
Over all time points	0.96 (0.70 to 1.22)
SF-12 physical (coefficient)^e	
12 weeks	
26 weeks	0.16 (-1.46 to 1.77)
39 weeks	1.65 (0.11 to 3.19)
52 weeks	-0.44 (-1.94 to 1.05)
Over all time points	0.27 (-0.84 to 1.37)
SF-12 mental (coefficient)^e	
12 weeks	
26 weeks	-2.91 (-4.78 to -1.04)

TABLE 9 Secondary outcomes controlling for factors associated with missingness (continued)

Outcome	Estimate (95% CI)
39 weeks	-3.05 (-4.80 to -1.30)
52 weeks	-1.68 (-3.51 to 0.15)
Over all time points	-3.13 (-4.39 to 1.88)
Global Rating Question: feeling worse (OR)	
6 weeks	1.03 (0.63 to 1.66)
12 weeks	
26 weeks	1.08 (0.63 to 1.85)
39 weeks	1.39 (0.79 to 2.43)
52 weeks	1.12 (0.63 to 1.98)
Coefficient, β coefficient; OR, odds ratio.	
a Range 0–27. Predictors of missingness: 6 weeks, site; 12 weeks, none; 26 weeks, age at randomisation, able to replace worn out furniture, able to buy new clothes and stop taking medication if feeling worse; 39 weeks, age at randomisation, SF-12 physical component score at baseline; 52 weeks, age at randomisation and site.	
b Range 0–21. Predictors of missingness: 6 weeks, site; 12 weeks, none; 26 weeks, age at randomisation, able to replace worn-out furniture and stop taking medication if feeling worse; 39 weeks, age at randomisation and SF-12 physical component score at baseline; 52 weeks, age at randomisation and site.	
c Range 0–13. Predictors of missingness: 6 weeks, site and CIS-R score at randomisation; 12 weeks, none; 26 weeks, able to buy new clothes; 39 weeks, age at randomisation and SF-12 physical component score at baseline; 52 weeks, age at randomisation and CIS-R score at randomisation.	
d Range 0 to 15. Predictors of missingness: 6 weeks, site; 12 weeks, none; 26 weeks, age at randomisation and able to buy new clothes; 39 weeks, age at randomisation; 52 weeks, age at randomisation and site.	
e Range 0–100. Predictors of missingness: 12 weeks, none; 26 weeks, age at randomisation, able to buy new clothes, CIS-R score at randomisation and stop taking medication if feeling worse; 39 weeks, age at randomisation; 52 weeks, age at randomisation and site.	
Notes	
For the analysis at all time points, all predictors of missingness from the individual time points for the given outcome are included in the model.	
There were no factors with statistical evidence to suggest that they were associated with missingness for any outcomes at the 12-week point, so no sensitivity analyses were performed.	

Adherence to trial medication

A larger percentage of participants in the placebo than maintenance group stopped taking their trial medication before the end of the trial (48% vs. 30%; HR 2.26, 95% CI 1.67 to 3.07). By 52 weeks, 39% (95% CI 32% to 45%) of participants in the discontinuation group and 20% (95% CI 15% to 25%) in the maintenance group had stopped taking trial medication and had returned to an antidepressant prescribed by their GP.

Subgroup analyses

For the subgroup analyses for the primary and secondary outcomes, there was no evidence of any differences according to the number of previous episodes of depression dichotomised at two compared with three or more or types of antidepressant. The most consistent evidence that we found supported a larger difference between groups in those who were younger at onset. For example, the p -value for interaction between age and PHQ-9 at 12 weeks was 0.0001. Among participants who were older at the onset of depression, those in the discontinuation group were less likely to relapse, although this finding was not statistically significant ($p = 0.0553$) (Tables 10–13).

RESULTS

TABLE 10 Subgroup analysis: time to relapse

Subgroup	HR (95% CI)	p-value for interaction
Sertraline	2.41 (1.20 to 4.82)	0.6010
Citalopram	2.14 (1.44 to 3.18)	
Fluoxetine	1.70 (1.05 to 2.76)	
CIS-R depression score below the median ^a	2.34 (1.56 to 3.52)	0.4252
CIS-R depression score above the median ^a	1.80 (1.24 to 2.61)	
CIS-R anxiety score below the median ^b	2.08 (1.49 to 2.89)	0.9893
CIS-R anxiety score above the median ^b	1.95 (1.20 to 3.18)	
Two previous episodes of depression	0.96 (0.26 to 3.62)	0.2132
Three or more previous episodes of depression	2.15 (1.63 to 2.85)	
Age at onset of depression below the median ^c	2.66 (1.84 to 3.85)	0.0313
Age at onset of depression above the median ^c	1.43 (0.94 to 2.16)	

a CIS-R depression dichotomised at < 3 vs. ≥ 3.
b CIS-R anxiety dichotomised at < 2 vs. ≥ 2.
c Age when became aware of depression dichotomised at < 32 years vs. ≥ 32 years.

TABLE 11 Subgroup analysis: PHQ-9^a

Subgroup	Coefficient (95% CI)	p-value for interaction
6 weeks		
Sertraline	0.74 (-0.76 to 2.25)	0.5279
Citalopram	0.56 (-0.28 to 1.41)	
Fluoxetine	-0.10 (-1.00 to 0.80)	
CIS-R depression score below the median ^b	0.29 (-0.23 to 0.81)	0.0649
CIS-R depression score above the median ^b	0.45 (-0.50 to 1.40)	
CIS-R anxiety score below the median ^c	0.49 (-0.09 to 1.07)	0.0233
CIS-R anxiety score above the median ^c	0.01 (-1.19 to 1.21)	
Two previous episodes of depression	1.23 (-0.21 to 2.67)	0.9420
Three or more previous episodes of depression	0.27 (-0.32 to 0.87)	
Age when became aware of depression below median ^d	0.52 (-0.30 to 1.34)	0.0631
Age when became aware of depression above median ^d	0.06 (-0.71 to 0.83)	
12 weeks		
Sertraline	4.74 (2.93 to 6.55)	0.0351
Citalopram	1.61 (0.67 to 2.55)	
Fluoxetine	1.73 (0.60 to 2.86)	
CIS-R depression score below the median ^b	2.25 (1.45 to 3.04)	0.0204
CIS-R depression score above the median ^b	2.16 (1.14 to 3.17)	
CIS-R anxiety score below the median ^c	2.05 (1.29 to 2.81)	0.1288

TABLE 11 Subgroup analysis: PHQ-9^a (continued)

Subgroup	Coefficient (95% CI)	p-value for interaction
CIS-R anxiety score above the median ^c	2.49 (1.19 to 3.80)	
Two previous episodes of depression	1.69 (0.06 to 3.31)	0.4567
Three or more previous episodes of depression	2.22 (1.50 to 2.95)	
Age when became aware of depression below the median ^d	3.01 (2.02 to 3.99)	0.0001
Age when became aware of depression above the median ^d	1.14 (0.23 to 2.06)	
26 weeks		
Sertraline	3.73 (1.84 to 5.61)	0.0029
Citalopram	0.61 (-0.42 to 1.65)	
Fluoxetine	-0.41 (-1.48 to 0.67)	
CIS-R depression score below the median ^b	0.79 (-0.05 to 1.63)	0.0002
CIS-R depression score above the median ^b	0.66 (-0.33 to 1.64)	
CIS-R anxiety score below the median ^c	1.09 (0.31 to 1.87)	0.0005
CIS-R anxiety score above the median ^c	0.06 (-1.22 to 1.33)	
Two previous episodes of depression	0.57 (-0.84 to 1.98)	0.7258
Three or more previous episodes of depression	0.79 (0.05 to 1.52)	
Age when became aware of depression below the median ^d	1.34 (0.35 to 2.34)	0.0423
Age when became aware of depression above the median ^d	0.07 (-0.88 to 1.03)	
39 weeks		
Sertraline	1.23 (-0.57 to 3.03)	0.4909
Citalopram	0.47 (-0.55 to 1.48)	
Fluoxetine	0.02 (-1.09 to 1.14)	
CIS-R depression score below the median ^b	0.33 (-0.44 to 1.10)	0.0044
CIS-R depression score above the median ^b	0.82 (-0.20 to 1.85)	
CIS-R anxiety score below the median ^c	0.45 (-0.30 to 1.21)	0.1390
CIS-R anxiety score above the median ^c	0.92 (-0.41 to 2.24)	
Two previous episodes of depression	-1.70 (-2.90 to -0.49)	0.1743
Three or more previous episodes of depression	0.77 (0.04 to 1.50)	
Age when became aware of depression below the median ^d	1.06 (0.13 to 2.00)	0.1984
Age when became aware of depression above the median ^d	0.01 (-1.01 to 1.03)	
52 weeks		
Sertraline	1.72 (-0.10 to 3.54)	0.5997
Citalopram	-0.04 (-1.07 to 0.99)	
Fluoxetine	-0.01 (-1.07 to 1.05)	
CIS-R depression score below the median ^b	0.13 (-0.69 to 0.95)	0.0036
CIS-R depression score above the median ^b	0.67 (-0.34 to 1.67)	
CIS-R anxiety score below the median ^c	0.48 (-0.30 to 1.26)	0.0032
CIS-R anxiety score above the median ^c	0.22 (-1.09 to 1.52)	
continued		

RESULTS

TABLE 11 Subgroup analysis: PHQ-9^a (continued)

Subgroup	Coefficient (95% CI)	p-value for interaction
Two previous episodes of depression	-0.89 (-2.53 to 0.75)	0.6074
Three or more previous episodes of depression	0.48 (-0.26 to 1.22)	
Age when became aware of depression below the median ^d	0.28 (-0.65 to 1.21)	0.9822
Age when became aware of depression above the median ^d	0.42 (-0.63 to 1.48)	
All time points		
Sertraline	2.26 (0.97 to 3.55)	0.0319
Citalopram	0.72 (0.07 to 1.37)	
Fluoxetine	0.22 (-0.48 to 0.93)	
CIS-R depression score below the median ^b	0.74 (0.21 to 1.27)	0.0092
CIS-R depression score above the median ^b	1.00 (0.34 to 1.67)	
CIS-R anxiety score below the median ^c	0.88 (0.37 to 1.40)	0.0192
CIS-R anxiety score above the median ^c	0.84 (-0.01 to 1.70)	
Two previous episodes of depression	0.62 (-0.55 to 1.78)	0.5401
Three or more previous episodes of depression	0.90 (0.43 to 1.38)	
Age when became aware of depression below the median ^d	1.27 (0.62 to 1.92)	0.0074
Age when became aware of depression above the median ^d	0.35 (-0.28 to 0.98)	

a Range 0–27.

b CIS-R depression dichotomised at < 3 vs. ≥ 3.

c CIS-R anxiety dichotomised at < 2 vs. ≥ 2.

d Age when became aware of depression dichotomised at < 32 vs. ≥ 32.

From *N Engl J Med*, Lewis G, Marston L, Duffy L, Freemantle N, Gilbody S, Hunter R, *et al.*, Maintenance or discontinuation of antidepressants in primary care, vol. 385, pp. 1275–67.³⁴ Copyright © 2021 Massachusetts Medical Society. Reprinted with permission.

TABLE 12 Subgroup analysis: GAD-7^a

Subgroup	Coefficient (95% CI)	p-value for interaction
6 weeks		
Sertraline	0.11 (-1.32 to 1.54)	0.4866
Citalopram	0.99 (0.24 to 1.74)	
Fluoxetine	-0.24 (-1.15 to 0.67)	
CIS-R depression score below the median ^b	0.54 (0.00 to 1.08)	0.3555
CIS-R depression score above the median ^b	0.48 (-0.42 to 1.39)	
CIS-R anxiety score below the median ^c	1.03 (0.52 to 1.55)	0.0002
CIS-R anxiety score above the median ^c	-0.52 (-1.67 to 0.63)	
Two previous episodes of depression	0.45 (-0.64 to 1.54)	0.3273
Three or more previous episodes of depression	0.56 (-0.01 to 1.13)	
Age when became aware of depression below median ^d	0.83 (0.04 to 1.62)	0.0361
Age when became aware of depression above median ^d	0.13 (-0.58 to 0.83)	

TABLE 12 Subgroup analysis: GAD-7^a (continued)

Subgroup	Coefficient (95% CI)	p-value for interaction
12 weeks		
Sertraline	3.66 (2.18 to 5.15)	0.4127
Citalopram	2.37 (1.59 to 3.15)	
Fluoxetine	1.63 (0.58 to 2.68)	
CIS-R depression score below the median ^b	1.97 (1.29 to 2.66)	0.9273
CIS-R depression score above the median ^b	2.84 (1.92 to 3.76)	
CIS-R anxiety score below the median ^c	2.37 (1.72 to 3.01)	0.0567
CIS-R anxiety score above the median ^c	2.53 (1.42 to 3.65)	
Two previous episodes of depression	0.17 (-1.18 to 1.51)	0.1142
Three or more previous episodes of depression	2.56 (1.93 to 3.19)	
Age when became aware of depression below the median ^d	3.26 (2.39 to 4.14)	0.0001
Age when became aware of depression above the median ^d	1.32 (0.57 to 2.08)	
26 weeks		
Sertraline	1.96 (0.00 to 3.91)	0.2058
Citalopram	0.70 (-0.22 to 1.63)	
Fluoxetine	0.30 (-0.77 to 1.38)	
CIS-R depression score below the median ^b	0.93 (0.14 to 1.73)	0.0045
CIS-R depression score above the median ^b	0.59 (-0.41 to 1.59)	
CIS-R anxiety score below the median ^c	1.02 (0.31 to 1.74)	0.0043
CIS-R anxiety score above the median ^c	0.38 (-0.85 to 1.60)	
Two previous episodes of depression	-0.85 (-2.83 to 1.12)	0.1842
Three or more previous episodes of depression	0.96 (0.27 to 1.65)	
Age when became aware of depression below the median ^d	0.98 (0.04 to 1.92)	0.5669
Age when became aware of depression above the median ^d	0.61 (-0.31 to 1.52)	
39 weeks		
Sertraline	1.37 (-0.29 to 3.04)	0.7993
Citalopram	0.82 (-0.10 to 1.73)	
Fluoxetine	0.75 (-0.24 to 1.75)	
CIS-R depression score below the median ^b	0.30 (-0.46 to 1.05)	0.0809
CIS-R depression score above the median ^b	1.68 (0.73 to 2.63)	
CIS-R anxiety score below the median ^c	0.75 (0.11 to 1.40)	0.0887
CIS-R anxiety score above the median ^c	1.52 (0.25 to 2.80)	
Two previous episodes of depression	-1.19 (-2.96 to 0.57)	0.3409
Three or more previous episodes of depression	1.18 (0.52 to 1.85)	
Age when became aware of depression below the median ^d	1.33 (0.50 to 2.16)	0.9303
Age when became aware of depression above the median ^d	0.66 (-0.30 to 1.62)	

continued

RESULTS

TABLE 12 Subgroup analysis: GAD-7^a (continued)

Subgroup	Coefficient (95% CI)	p-value for interaction
52 weeks		
Sertraline	0.81 (-0.52 to 2.15)	0.3556
Citalopram	0.35 (-0.59 to 1.28)	
Fluoxetine	-0.54 (-1.52 to 0.43)	
CIS-R depression score below the median ^b	-0.21 (-0.97 to 0.55)	0.2593
CIS-R depression score above the median ^b	0.74 (-0.20 to 1.67)	
CIS-R anxiety score below the median ^c	-0.04 (-0.72 to 0.63)	0.0278
CIS-R anxiety score above the median ^c	0.96 (-0.19 to 2.10)	
Two previous episodes of depression	-1.67 (-3.20 to -0.14)	0.2794
Three or more previous episodes of depression	0.41 (-0.25 to 1.07)	
Age when became aware of depression below the median ^d	-0.04 (-0.90 to 0.82)	0.1410
Age when became aware of depression above the median ^d	0.61 (-0.30 to 1.52)	
All time points		
Sertraline	1.56 (0.38 to 2.73)	0.2203
Citalopram	1.12 (0.53 to 1.71)	
Fluoxetine	0.30 (-0.38 to 0.99)	
CIS-R depression score below the median ^b	0.69 (0.20 to 1.19)	0.3928
CIS-R depression score above the median ^b	1.35 (0.70 to 1.99)	
CIS-R anxiety score below the median ^c	1.04 (0.58 to 1.50)	0.0086
CIS-R anxiety score above the median ^c	0.98 (0.18 to 1.78)	
Two previous episodes of depression	-0.38 (-1.49 to 0.73)	0.0947
Three or more previous episodes of depression	1.14 (0.69 to 1.58)	
Age when became aware of depression below the median	1.31 (0.71 to 1.90)	0.0767
Age when became aware of depression above the median	0.64 (0.05 to 1.22)	

a Range 0–21.

b CIS-R depression dichotomised at < 3 vs. ≥ 3.

c CIS-R anxiety dichotomised at < 2 vs. ≥ 2.

d Age when became aware of depression dichotomised at < 32 years vs. ≥ 32 years.

From *N Engl J Med*, Lewis G, Marston L, Duffy L, Freemantle N, Gilbody S, Hunter R, *et al.*, Maintenance or discontinuation of antidepressants in primary care, vol. 385, pp. 1275–67.³⁴ Copyright © 2021 Massachusetts Medical Society. Reprinted with permission.

Post hoc analyses

Of the 134 participants randomised to the discontinuation group who had relapsed by the end of the trial, 49 (37%, 95% CI 28% to 45%) remained on trial medication, 71 (53%, 95% CI 44% to 62%) had returned to a known antidepressant and 14 (10%, 95% CI 6% to 17%) were not taking any antidepressant. Of the 89 participants randomised to the maintenance group who had relapsed, 46 (52%, 95% CI 41% to 62%) remained on trial medication, 32 (36%, 95% CI 26% to 47%) had returned to a known antidepressant and 11 (12%, 95% CI 6% to 21%) were not taking any antidepressant.

TABLE 13 Subgroup analysis: Global Rating Question – feel worse

Subgroup	OR (95% CI)	p-value for interaction
6 weeks		
Sertraline	1.07 (0.36 to 3.17)	0.3302
Citalopram	1.43 (0.66 to 3.10)	
Fluoxetine	0.62 (0.28 to 1.38)	0.5026
CIS-R depression score below the median ^a	1.21 (0.59 to 2.48)	
CIS-R depression score above the median ^a	0.87 (0.45 to 1.67)	
CIS-R anxiety score below the median ^b	1.12 (0.61 to 2.06)	
CIS-R anxiety score above the median ^b	0.86 (0.39 to 1.90)	0.6116
Two previous episodes of depression	^c	
Three or more previous episodes of depression	1.01 (0.62 to 1.64)	
Age when became aware of depression below the median ^d	1.22 (0.67 to 2.22)	0.2465
Age when became aware of depression above the median ^d	0.66 (0.28 to 1.54)	
12 weeks		
Sertraline	3.88 (1.37 to 10.98)	0.4329
Citalopram	3.36 (1.80 to 6.27)	
Fluoxetine	1.96 (0.97 to 3.97)	0.3171
CIS-R depression score below the median ^a	3.53 (1.97 to 6.33)	
CIS-R depression score above the median ^a	2.30 (1.27 to 4.19)	
CIS-R anxiety score below the median ^b	2.65 (1.59 to 4.42)	
CIS-R anxiety score above the median ^b	3.54 (1.71 to 7.31)	0.5232
Two previous episodes of depression	2.12 (0.43 to 10.52)	
Three or more previous episodes of depression	2.93 (1.90 to 4.51)	
Age when became aware of depression below the median ^d	3.01 (1.71 to 5.28)	0.7826
Age when became aware of depression above the median ^d	2.67 (1.43 to 4.98)	
26 weeks		
Sertraline	1.52 (0.51 to 4.53)	0.7241
Citalopram	0.90 (0.46 to 1.76)	
Fluoxetine	1.00 (0.40 to 2.49)	0.0641
CIS-R depression score below the median ^a	1.53 (0.79 to 2.95)	
CIS-R depression score above the median ^a	0.62 (0.31 to 1.24)	
CIS-R anxiety score below the median ^b	1.25 (0.69 to 2.27)	
CIS-R anxiety score above the median ^b	0.68 (0.31 to 1.52)	0.2368
Two previous episodes of depression	0.36 (0.03 to 4.50)	
Three or more previous episodes of depression	1.06 (0.65 to 1.71)	
Age when became aware of depression below the median ^d	1.15 (0.61 to 2.18)	0.5130
Age when became aware of depression above the median ^d	0.84 (0.41 to 1.70)	

continued

RESULTS

TABLE 13 Subgroup analysis: Global Rating Question – feel worse (continued)

Subgroup	OR (95% CI)	p-value for interaction
39 weeks		
Sertraline	1.18 (0.28 to 4.92)	0.8095
Citalopram	1.49 (0.66 to 3.39)	
Fluoxetine	0.98 (0.37 to 2.60)	
CIS-R depression score below the median ^a	1.45 (0.64 to 3.27)	0.9045
CIS-R depression score above the median ^a	1.35 (0.63 to 2.92)	
CIS-R anxiety score below the median ^b	1.96 (0.94 to 4.07)	0.1690
CIS-R anxiety score above the median ^b	0.86 (0.34 to 2.16)	
Two previous episodes of depression	0.46 (0.06 to 3.35)	0.2554
Three or more previous episodes of depression	1.53 (0.85 to 2.75)	
Age when became aware of depression below the median ^d	1.56 (0.70 to 3.47)	0.7327
Age when became aware of depression above the median ^d	1.28 (0.58 to 2.80)	
52 weeks		
Sertraline	1.20 (0.29 to 5.02)	0.5980
Citalopram	1.28 (0.55 to 2.97)	
Fluoxetine	0.66 (0.24 to 1.82)	
CIS-R depression score below the median ^a	1.02 (0.44 to 2.36)	0.8008
CIS-R depression score above the median ^a	1.18 (0.55 to 2.56)	
CIS-R anxiety score below the median ^b	1.04 (0.53 to 2.05)	0.7995
CIS-R anxiety score above the median ^b	1.22 (0.44 to 3.39)	
Two previous episodes of depression	0.85 (0.05 to 15.16)	0.8727
Three or more previous episodes of depression	1.08 (0.60 to 1.93)	
Age when became aware of depression below the median ^d	1.30 (0.59 to 2.90)	0.4592
Age when became aware of depression above the median ^d	0.84 (0.37 to 1.94)	
OR, odds ratio.		
a CIS-R depression dichotomised at < 3 vs. ≥ 3.		
b CIS-R anxiety dichotomised at < 2 vs. ≥ 2.		
c Perfect prediction.		
d Age when became aware of depression dichotomised at < 32 years vs. ≥ 32 years.		
From <i>N Engl J Med</i> , Lewis G, Marston L, Duffy L, Freemantle N, Gilbody S, Hunter R, <i>et al.</i> , Maintenance or discontinuation of antidepressants in primary care, vol. 385, pp. 1275–67. ³⁴ Copyright © 2021 Massachusetts Medical Society. Reprinted with permission.		

In total, 59% (141/240) of participants in the discontinuation group were unblinded through either withdrawal from the trial or emergency code break. The rate of unblinding was much lower in the maintenance group (29%; 68/236). Participants were asked whether they thought that they were taking the active drug (antidepressant) or the placebo. Over the course of the trial, 71% (162/228) of participants in the discontinuation group and 47% (108/232) of participants in the maintenance group correctly guessed their randomised group at some time before being unblinded.

We also investigated whether or not withdrawal symptoms differed according to antidepressant class. At 6 weeks, there was weak evidence ($p = 0.07$) that the effect of discontinuation (compared with maintenance) on withdrawal symptoms was smaller among those receiving fluoxetine than sertraline and citalopram. At 12 weeks, there was stronger evidence that the effect of discontinuation compared with maintenance on withdrawal symptoms differed by group ($p = 0.002$). Withdrawal symptoms were less common in those taking fluoxetine and citalopram than in those taking sertraline. There was no evidence of an interaction between treatment group and antidepressant class at week 26, 39 or 52. Withdrawal symptoms in those who discontinued and those who remained on maintenance antidepressant are shown in *Report Supplementary Material 10*, according to antidepressant class.

Chapter 4 Economic evaluation

Introduction

The aim of the economic evaluation was to calculate the mean incremental cost per quality-adjusted life-year (QALY) gained by discontinuing antidepressant medication and replacing it with placebo, compared with antidepressant maintenance, from a health-care cost perspective, using trial data collected over 12 months from participants and primary care electronic medical records. SSRIs are the most common and recommended antidepressants, and their mean purchase cost is relatively low, around 4p per day. The majority of analyses evaluating the cost-effectiveness of prescribing antidepressants for depression have conducted head-to-head decision modelling of different antidepressants to determine the most cost-effective antidepressant to treat current symptoms, rather than considering the question of the wider cost-effectiveness of their long-term use,^{20,54,55} with analyses rarely going beyond a 12-month time horizon. The impact of side effects and withdrawal symptoms following long-term use is rarely given consideration; therefore, only limited, poor-quality data are available on which decision modelling could be based to describe long-term use.²¹ In this chapter, we present the results of a trial-based cost-utility analysis (CUA) comparing discontinuation with continued antidepressant maintenance in primary care in England over 12 months, using patient-level data on health-care resource use and a preference-based measure of health-related quality of life [EuroQol-5 Dimensions, five-level version (EQ-5D-5L)].

Methods

Outcome measures

Primary health-care resource use information was collected from GP electronic records for primary care contacts and prescriptions, via the form given in *Report Supplementary Material 6*. This covered from 6 months preceding baseline to 12 months post randomisation.

The participants completed a modified Client Service Receipt Inventory (CSRI)⁵⁶ for other health-care and social care resource use and wider societal impact. The CSRI captured information on community and acute care health service contacts, mental health community and inpatient service use, social care, employment, and welfare payments, gathering information that could not be obtained from primary care electronic medical records. The version of the CSRI used in the trial is given in *Report Supplementary Material 5*.

The participants completed the EQ-5D-5L⁵⁷ and SF-12 at baseline and at 12, 26, 39 and 52 weeks. The EQ-5D-5L is a short, participant-completed, generic health-related quality-of-life questionnaire comprising five questions or domains, asking about mobility, self-care, usual activities, pain or discomfort, and anxiety or depression. Each question carries five possible responses, or levels, and these responses can be used to calculate utility scores.

Resource use and costs

The costs of the four ANTLER trial medications in each group were calculated according to group allocation and protocol doses, and the prescription information collected from participants' primary care electronic medical records. The ANTLER trial medication in the discontinuation group was costed for citalopram, sertraline and mirtazapine as 1 month at half of the original dose, followed by 1 month at one-quarter of the original dose, followed by no cost for the remaining 10 months of the trial (i.e. placebo administered during the trial was priced at zero for this cost-utility analysis). Fluoxetine was costed as 1 month at half of the original dose followed by no cost for the remaining 11 months of the trial, unless participants in the discontinuation group reported stopping their trial tablets before the end of month 2 (or month 1 for those initially on fluoxetine). The ANTLER trial medication in the

maintenance group was the continuation of their medication at the dose prescribed at recruitment for the 12 months of the trial or until the date at which participants reported stopping their medication. The use of other relevant antidepressant medications prescribed in either group at any point during the trial (citalopram, fluoxetine, mirtazapine, sertraline, amitriptyline, diazepam and zopiclone) was captured from participants' electronic medical records and costed according to reported daily doses and other prescription information. Unit costs for medications were obtained from the *British National Formulary*⁵⁸ and were applied using the lowest package cost to the NHS, according to the duration, dose and frequency of each reported prescription.

The unit costs of health-care contacts were obtained from the Personal Social Services Research Unit (PSSRU)⁵⁹ and *NHS Reference Costs 2018/19*⁶⁰ (Table 14). Private health-care resource use was costed

TABLE 14 Unit costs applied to resource use collected from patient electronic medical records at GP surgeries, and collected directly from participants via the modified CSRI questionnaire

Resource category	Unit cost (NHS, unless stated otherwise; £)	Assumptions	Source
GP surgery consultation	28.00	9 minutes	PSSRU 2018–19 ⁵⁹
GP telephone consultation	15.50	5 minutes	PSSRU 2018–19 ⁵⁹
GP home visit	34.72	11.2 minutes (PSSRU 2015 ⁶¹)	PSSRU 2018–19 ⁵⁹
Practice nurse surgery consultation	12.30	20 minutes	PSSRU 2018–19 ⁵⁹
Practice nurse telephone consultation	6.17	10 minutes	PSSRU 2018–19 ⁵⁹
Practice nurse home visit	21.60	35 minutes	PSSRU 2018–19 ⁵⁹
Phlebotomist	4.00	10 minutes	PSSRU 2018–19 ⁵⁹
Cognitive-behavioural therapist	54.50	Band 7	PSSRU 2018–19 ⁵⁹
Cognitive-behavioural therapist: privately funded by the patient	50.35	Mean from trial data	PSSRU 2018–19 ⁵⁹
Clinical psychologist	64.68	Band 8a	PSSRU 2018–19 ⁵⁹
Exercise or physical activity scheme or 'Exercise on prescription': NHS	10.28	Uplifted to 2018–19 prices using HCHS indices from PSSRU 2018–19 ⁵⁹	Isaacs <i>et al.</i> ⁶²
NHS walk-in centres	35.38	Estimated using PSSRU	PSSRU 2018–19 ⁵⁹
Ambulance or hospital transport	257.34		<i>NHS Reference Costs 2018–19</i> ⁶⁰
NHS Direct or 'Call 111'	13.26	Uplifted to 2018–19 prices using HCHS indices from PSSRU 2018–19 ⁵⁹	Pope <i>et al.</i> ⁶³
A&E attendance	155.70	Weighted mean of top two non-admitted categories	<i>NHS Reference Costs 2018–19</i> ⁶⁰
Hospital admission	1909.49	Weighted mean of EL, NEL, NES and RP costs	<i>NHS Reference Costs 2018–19</i> ⁶⁰
Mental health nurse (or 'community psychiatric nurse')	33.83	Band 5, community-based scientific and professional staff	PSSRU 2018–19 ⁵⁹
Occupational therapist	44.16	Band 8a, community-based scientific and professional staff	PSSRU 2018–19 ⁵⁹
Social worker	44.55	Social worker, adult services	PSSRU 2018–19 ⁵⁹
Other medical professional (mostly consultant-level NHS)	100.82	Mean from free-text descriptions from participants	PSSRU 2018–19 ⁵⁹
Other medical professional: privately funded by patient	32.67	Mean from trial data	

A&E, accident and emergency; EL, elective inpatients; HCHS, Hospital and Community Health Service; NEL, non-elective inpatients; NES, non-elective short stay; RP, regular day or night admissions.
Group therapy was costed with a mean of four participants per group.

based on participants' reported out-of-pocket costs. For the very few participants who reported using private health care but did not report actual out-of-pocket costs, we assumed the equivalent PSSRU and NHS reference costs. Productivity was costed using the human capital approach to cost time off work with mean costs of Office for National Statistics employment categories applied according to the occupation described in the free text in the CSRI. All costs are in 2018/19 Great British pounds.

Utilities and quality-adjusted life-years

Quality-adjusted life-years were calculated from participants' responses to the EQ-5D-5L using the Devlin and Krabbe⁶⁴ time trade-off (TTO) tariff for the UK for the primary economic analysis. The van Hout *et al.*⁶⁵ mapping algorithm for generating utilities from EQ-5D-5L via the EQ-5D-3L tariff is currently preferred by NICE and, therefore, was used in a secondary analysis. Participants' responses to the SF-12 were used in another secondary analysis to calculate utilities and QALYs using the Short Form questionnaire-6 Dimensions (SF-6D) utility-scoring tariff⁶⁶ to further test the robustness of the results to choice of utility estimation method. Although NICE currently recommends the van Hout *et al.*⁶⁵ mapping tariff for calculating QALYs, there is concern that the mapping algorithm is not as sensitive to changes in depression as the Devlin and Krabbe tariff;⁶⁴ therefore, we planned to use the Devlin and Krabbe tariff as the primary analysis.⁶⁷

Quality-adjusted life-years were calculated as the area under the curve using the methodology set out in Hunter *et al.*⁶⁸ The costs in the primary analysis were from a health-care and social care cost perspective. Given that the time horizon for the analysis was 12 months, costs and QALYs were not discounted.

Economic evaluation analytical methods

Descriptive statistics for the primary analysis, which used multiple imputation for missing utility scores and CSRI information, are reported for resource use, costs and utilities at each time point. The baseline age and SF-12 Physical Component Summary score were identified as predictors of missingness for the imputations.

The mean per-participant differences in 12-month costs and QALYs by treatment group were jointly estimated via bootstrapped seemingly unrelated regression, with 1000 iterations to account for the correlation between costs and QALYs,⁶⁹ adjusting for baseline values and the minimisation variables of study centre, ANTLER trial medication and binary severity of depressive symptoms at baseline, with imputed data sets combined according to Rubin's rules.⁷⁰ The primary economic analysis was calculated using the multiple imputation data set and the bootstrapped, seemingly unrelated, regression results, as set out by Leurent *et al.*⁷¹

We took a probabilistic approach to aid decision-making for resource allocation and calculated the probability that discontinuing antidepressants to zero dose was cost-effective for a range of thresholds of cost per QALY gained compared with antidepressant maintenance. The 'new' treatment here is discontinuation and the 'old' treatment is maintenance; therefore, the incremental costs and QALYs are calculated as discontinuation pathway values minus maintenance pathway values. In the protocol paper⁷² and statistical analysis plan, we proposed to use the placebo as the 'old' treatment, as might be usual in most placebo-controlled trials. However, on reflection, we realised that the 'new' intervention in our participants was to discontinue antidepressants, so we decided to conduct the analysis from this viewpoint.

The incremental cost-effectiveness ratio (ICER) for each analysis was calculated as the mean estimated difference in costs divided by the mean estimated difference in QALYs. The bootstrapped results were plotted on cost-effectiveness planes (CEPs) and the proportions of estimates that were above the cost-effectiveness threshold were plotted on corresponding cost-effectiveness acceptability curves (CEACs) for a range of thresholds.

Health economics secondary and sensitivity analyses

We report ICERs, CEACs and CEPs for the following secondary analyses:

- health-care and social care cost perspective using the EQ-5D-5L responses and mapping tariff⁷³ for the calculation of utilities and QALYs
- health-care and social care cost perspective using the SF-12 responses and SF-6D tariff for the calculation of utilities and QALYs⁶⁶
- wider cost perspective, including out-of-pocket and productivity costs and using the EQ-5D-5L responses and Devlin and Krabbe⁶⁴ TTO tariff for the calculation of utilities and QALYs
- wider cost perspective, including out-of-pocket and productivity costs and using the EQ-5D-5L responses and mapping tariff⁶⁵ for the calculation of utilities and QALYs
- wider cost perspective, including out-of-pocket and productivity costs and using the SF-12 responses and SF-6D tariff for the calculation of utilities and QALYs.

A sensitivity analysis was conducted based on the primary economic analysis (health-care and social care cost perspective and utilities and QALYs calculated using the Devlin and Krabbe⁶⁴ tariff) for complete cases only.

A post hoc sensitivity analysis included relapse as a covariate at each follow-up point and for total costs and QALYs to investigate the relationship between relapse and costs and utilities, given that it was identified by the study team that this might be a more important factor than the treatment group itself and could potentially be driving the observed results. This involved creating variables for each follow-up time point (3, 6, 9 and 12 months), which indicated whether or not participants had relapsed, as defined by the primary clinical outcome, at any time up to that time point.

We did not prespecify an analysis that would have captured the primary outcome, for example calculating the mean incremental cost per depression-free day, given that there is no cost-effectiveness threshold for a condition-specific outcome related to depression. There is also increasing evidence of the validity and responsiveness of the EQ-5D-5L in depression;⁷⁴ therefore, a cost-per-QALY analysis calculated using the EQ-5D-5L is likely to capture this information. If the mean incremental cost per depression-free day was calculated, discontinuation would be dominated by long-term maintenance given that discontinuation would cost more and result in fewer depression-free days, reflecting the same result as the cost-per-QALY analysis.

Health economics results

Costs

Descriptive statistics for resource use are reported in *Tables 15–18*, including the raw mean (SD) and adjusted differences by group for the total overall costs using linear regression (see *Table 18*). *Table 15* shows the primary care costs, *Table 16* the CSRI costs and *Table 17* the antidepressant medications (including ANTLER medication), with the total cost statistics shown in *Table 18*. There was a difference in the overall total baseline costs between the two randomised groups and the impact of this on the results can be seen by comparing the raw with the adjusted mean differences by randomised group (see *Table 18*). The antidepressant medication costs over the 12 months were lower in the discontinuation group than in the maintenance group (mean per-participant difference of –£6.04, 95% CI –£6.97 to –£5.11). GP (GP consultation) costs over the 12 months were higher in the discontinuation group than the maintenance group (mean per-participant difference of £16.62, 95% CI £0.70 to £32.53), which equates to approximately half of a GP visit. Improving Access to Psychological Therapies costs over the 12 months were also higher in the discontinuation group than in the maintenance group (mean per-participant difference of £16.98, 95% CI £1.11 to £32.86), which equates to approximately 15–20 minutes of a therapist's time. Adjustments were made in each case for baseline values, treatment group and the three minimisation variables.

TABLE 15 Mean raw costs of health-care resource use and estimated adjusted differences between the discontinuation group and the maintenance group: primary care

Time period	Treatment group					
	Maintenance		Discontinuation		Discontinuation vs. maintenance	
	n	Mean (£) (raw) (SD)	n	Mean (£) (raw) (SD)	Adjusted difference (£) (95% CI)	p value
GP cost						
Baseline	233	51 (50)	237	49 (54)		
6 months	233	53 (55)	237	68 (60)		
12 months	233	58 (65)	237	59 (62)		
Total	233	111 (101)	237	127 (99)	16.616 (0.701 to 32.532)	0.041 ^a
Practice nurse						
Baseline	233	6 (13)	237	6 (11)		
6 months	233	10 (17)	237	10 (16)		
12 months	233	10 (17)	237	9 (16)		
Total	233	20 (28)	237	19 (28)	-1.263 (-5.786 to 3.259)	0.584
Phlebotomist						
Baseline	233	0.48 (1.67)	237	0.2 (1.15)		
6 months	233	0.24 (1.09)	237	0.35 (1.41)		
12 months	233	0.33 (1.56)	237	0.35 (1.50)		
Total	233	0.57 (2.07)	237	0.71 (2.12)	0.205 (-0.178 to 0.588)	0.294
Other community contacts						
Baseline	233	21 (90)	237	22 (77)		
6 months	233	28 (108)	237	20 (64)		
12 months	233	21 (69)	237	20 (76)		
Total	233	49 (165)	237	41 (119)	-9.948 (-34.360 to 14.464)	0.424
Total primary care cost						
Baseline	233	79 (105)	237	77 (100)		
6 months	233	91 (136)	237	98 (93)		
12 months	233	90 (115)	237	88 (103)		
Total	233	181 (228)	237	187 (164)	5.899 (-25.392 to 37.189)	0.712
a Standard error when CSRI imputed.						

A post hoc analysis that aimed to assess whether or not relapse was the driver behind any differences between the treatment groups presented results by both treatment group and relapse status, under the NHS and PSS perspective. The results of these analyses are given in *Report Supplementary Material 8* (see *Table A8.1* for costs calculated from primary care records at baseline and each time point; *Table A8.2* for corresponding adjusted overall 12-month costs; *Table A8.3* for costs calculated from the patient-completed CSRI; *Table A8.4* for corresponding adjusted overall 12-month costs; *Table A8.5* for antidepressant medication costs; *Table A8.6* for corresponding adjusted overall 12-month costs; *Table A8.7* for raw mean differences in overall costs; and *Table A8.8* for corresponding adjusted mean differences).

TABLE 16 Health-care resource use: CSRI

Time period	Treatment group				Discontinuation vs. maintenance	
	Maintenance	Discontinuation			Adjusted difference (£) (95% CI)	p-value
	n	Mean (£) (raw) (SD ^a)	n	Mean (£) (raw) (SD ^a)		
Mental health contacts (CSRI)						
Baseline	237	36.65 (335.78)	239	9.88 (48.13)		
6 months	211	14.85 (128.90)	193	10.94 (48.57)		
12 months	210	9.97 (54.71)	181	18.88 (57.31)		
Total	206	23.38 (156.28)	179	30.33 (89.31)	16.984 (1.111 to 32.8558)	0.036
Other community-based contacts (CSRI)						
Baseline	237	3.07 (25.11)	239	0.09 (1.33)		
6 months	211	0 (0)	193	0.66 (6.72)		
12 months	210	1.38 (18.49)	181	0 (0)		
Total	206	1.41 (18.67)	179	0.71 (6.97)	0.963 (-0.233 to 2.160)	0.115
Emergency care (CSRI)						
Baseline	237	9.43 (124.77)	239	0.06 (0.86)		
6 months	211	2.44 (35.43)	193	0.14 (1.35)		
12 months	210	5.58 (56.29)	181	0.20 (2.63)		
Total	206	8.19 (66.99)	179	0.35 (2.98)	-7.188 (-15.390 to 1.013)	0.086
Total CSRI costs, with missing values imputed (base case)						
Baseline	232	50.20 (361.62)	236	10.15 (48.61)		
6 months	232	17.11 (8.96)	236	11.09 (3.31)		
12 months	232	17.06 (6.31)	236	18.49 (4.04)		
Total	232	34.17 (12.25)	236	29.58 (5.95)	8.895 (-19.696 to 37.486)	0.541
Total CSRI costs, complete-case analysis (secondary analysis)						
Baseline	237	49.14 (357.84)	239	10.02 (48.32)		
6 months	211	17.29 (133.41)	193	11.74 (49.20)		
12 months	210	16.93 (95.08)	181	19.08 (57.31)		
Total	206	32.97 (186.79)	179	31.39 (89.65)	9.147 (-11.101 to 29.396)	0.376

a Standard error when CSRI is imputed.

TABLE 17 Health-care resource use: antidepressants medication costs, including the ANTLER trial medication

Time period	Treatment group				Discontinuation vs. maintenance	
	Maintenance	Discontinuation			Adjusted difference (£) (95% CI)	p-value
	n	Mean (£) (raw) (SD)	n	Mean (£) (raw) (SD)		
Antidepressant medications						
Baseline	238	5.42 (3.25)	240	5.37 (3.58)		
6 months	238	6.56 (2.73)	240	3.24 (2.45)		
12 months	238	6.71 (4.07)	240	3.88 (3.16)		
Total	238	13.27 (6.12)	240	7.12 (4.89)	-6.037 (-6.96858 to -5.1058)	<0.001

TABLE 18 Health-care resource use: total costs

Time period	Maintenance		Discontinuation		Discontinuation vs. maintenance		Adjusted difference (£) (95% CI)	p-value
	n	Mean (£) (raw) (SD ^a)	Mean (£) (adjusted) (SD ^a)	n	Mean (£) (raw) (SD ^a)	Mean (£) (adjusted) (SD ^a)		
Total costs, all categories (primary care, CSRI imputed and medications) (base case)								
Baseline	232	134.59 (374.49)		236	92.96 (111.17)			
6 months	232	114.36 (12.80)	99.18 (9.34)	236	112.77 (6.82)	116.70 (9.68)		
12 months	232	114.40 (9.82)	107.08 (8.63)	236	111.09 (7.91)	111.31 (9.21)		
Total	232	228.76 (19.55)	204.44 (14.84)	236	223.86 (12.19)	227.65 (15.78)	23.218 (-19.463 to 65.900)	0.285
Total costs, all categories (primary care, CSRI, medications; complete cases) (secondary analysis)								
Baseline	232	134.59 (374.49)		236	92.96 (111.17)			
6 months	206	101.63 (168.49)	99.18 (10.73)	192	114.06 (108.10)	116.70 (7.41)		
12 months	205	110.55 (143.24)	107.08 (8.98)	180	107.35 (121.06)	111.31 (8.59)		
Total	201	210.93 (265.77)	204.44 (16.19)	178	220.32 (192.91)	227.65 (12.92)	23.218 (-15.656 to 62.093)	0.242
Total costs (sensitivity analysis), all categories (primary care, CSRI with missing = 0, medications) (secondary analysis)								
Baseline	232	134.59 (374.49)		236	92.96 (111.17)			
6 months	232	112.98 (187.37)	111.20 (12.27)	236	111.28 (101.91)	113.03 (7.45)		
12 months	232	112.67 (146.06)	110.41 (9.50)	236	107.22 (115.99)	109.44 (7.95)		
Total	232	223.87 (289.17)	219.86 (18.76)	236	218.08 (181.84)	222.02 (13.18)	2.157 (-39.742 to 44.056)	0.920
a Standard error when CSRI imputed (multiple imputation by chained equations).								
Note								
Presented for the base case (CSRI costs imputed where missing) and for two sensitivity analyses (complete cases only for CSRI, and with missing CSRI totals imputed as zero).								

The mean total imputed unadjusted health-care and social care costs were £224 [standard error (SE) £20] per participant in the discontinuation group and £229 (SE £12) per participant in the maintenance group. Adjusting for baseline differences and other covariates, as specified in *Economic evaluation analytical methods*, the mean adjusted values were £228 (SE £16) per participant in the discontinuation group and £204 (SE £15) per participant in the maintenance group, with a mean adjusted difference of discontinuation costing £23 more (95% CI -£19 to £66) per patient over 12 months. There was a £41 (95% CI -£222 to £303) adjusted difference in costs owing to productivity loss for the discontinuation group compared with the maintenance group, with a total cost difference of £0.14 (95% CI -£230 to £230) when this, along with other private and out-of-pocket costs across the different costing categories, were added to the total health-care and social care costs.

Breakdowns of costs included in the societal perspective analyses, that is NHS and PSS base-case costs plus those paid out of pocket by patients and those calculated as a result of productivity losses from time off work, are given in *Report Supplementary Material 8*:

- *Table A8.9* – primary and community care costs, including costs paid out of pocket by patients
- *Table A8.10* – CSRI costs, including costs paid out of pocket by patients
- *Table A8.11* – numbers of days off work
- *Table A8.12* – costs due to productivity losses
- *Table A8.13* – total costs, including costs paid out of pocket by patients and productivity losses.

The numbers of participants reporting any psychotherapy use in the 6 months preceding baseline and in the first and second halves of the 12-month follow-up are as follows.

In the 6-month period preceding baseline:

- 24 out of 237 (10.1%) participants in the maintenance group reported some use of psychotherapy
- 18 out of 239 (7.5%) participants in the discontinuation group reported some use of psychotherapy.

In the period from baseline to 6 months:

- 10 out of 211 (4.7%) participants in the maintenance group reported some use of psychotherapy
- 16 out of 193 (8.3%) participants in the discontinuation group reported some use of psychotherapy.

In the period from 6 to 12 months post randomisation:

- 15 out of 210 (7.1%) participants in the maintenance group reported some use of psychotherapy
- 30 out of 181 (16.6%) participants in the discontinuation group reported some use of psychotherapy.

Utility scores and quality-adjusted life-years

Table 19 shows the complete-case raw utility scores at each time point and the QALYs calculated as the area under the curve, as well as the adjusted differences according to treatment group. The *p*-values indicate whether or not the difference in the value in that row (e.g. 3-month utility) is statistically significantly different between the two treatment groups, using linear regression. *Table A8.14* in *Report Supplementary Material 8* shows the complete-case raw and adjusted utility scores, and the adjusted differences according to both treatment group and relapse status. *Table A8.15* in *Report Supplementary Material 8* shows the mean adjusted differences in QALYs calculated over the 12-month period, and the mean adjusted difference in utility score at 3 months. These are given by both treatment group and relapse status in a post hoc analysis designed to explore whether or not relapse was driving the differences observed between the treatment groups.

TABLE 19 Mean raw utilities at baseline and all follow-up points and mean 12-month QALYs, and estimated adjusted differences between the discontinuation group and the maintenance group

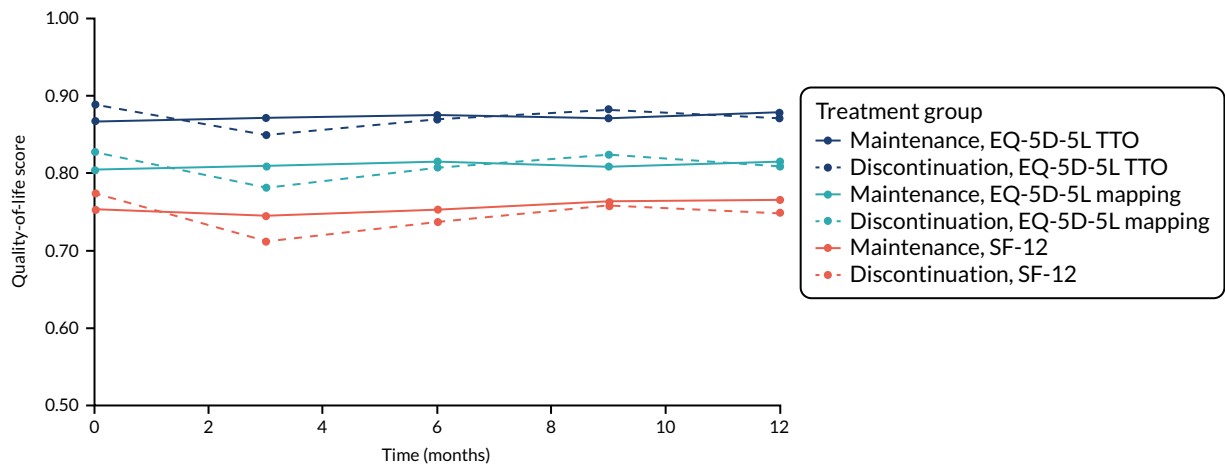
Time point	Treatment group				Discontinuation vs. maintenance	
	Maintenance (0)	Discontinuation (1)			Adjusted difference (95% CI)	p-value
	n	Mean (raw) (SD)	n	Mean (raw) (SD)		
EQ-5D-5L TTO: base case						
Baseline	237	0.868 (0.151)	240	0.889 (0.114)		
3 months	228	0.872 (0.146)	215	0.849 (0.145)	-0.037 (-0.059 to -0.015)	0.001 ^a
6 months	210	0.875 (0.142)	191	0.870 (0.145)	-0.014 (-0.038 to 0.011)	0.268
9 months	212	0.872 (0.148)	183	0.882 (0.122)	-0.001 (-0.023 to 0.022)	0.953
12 months	210	0.879 (0.144)	181	0.871 (0.151)	-0.021 (-0.045 to 0.003)	0.090
QALYs	205	0.876 (0.119)	173	0.869 (0.115)	-0.019 (-0.035 to -0.003)	0.020 ^a
EQ-5D-5L mapping: secondary analysis						
Baseline	237	0.804 (0.186)	240	0.828 (0.146)		
3 months	228	0.810 (0.178)	215	0.782 (0.183)	-0.044 (-0.072 to -0.017)	0.002 ^a
6 months	210	0.815 (0.176)	191	0.808 (0.181)	-0.019 (-0.049 to 0.011)	0.218
9 months	212	0.809 (0.181)	183	0.824 (0.146)	0.003 (-0.024 to 0.031)	0.828
12 months	210	0.815 (0.181)	181	0.810 (0.183)	-0.019 (-0.049 to 0.011)	0.214
QALYs	205	0.814 (0.151)	173	0.806 (0.139)	-0.022 (-0.042 to -0.002)	0.028 ^a
SF-12: secondary analysis						
Baseline	237	0.754 (0.124)	239	0.774 (0.125)		
3 months	227	0.745 (0.132)	216	0.712 (0.128)	-0.045 (-0.065 to -0.024)	< 0.001 ^a
6 months	210	0.754 (0.131)	192	0.738 (0.134)	-0.025 (-0.049 to -0.001)	0.040 ^a
9 months	212	0.764 (0.134)	183	0.759 (0.132)	-0.012 (-0.034 to 0.010)	0.287
12 months	210	0.766 (0.136)	181	0.749 (0.135)	-0.025 (-0.051 to 0.001)	0.055
QALYs	204	0.759 (0.103)	174	0.743 (0.100)	-0.027 (-0.043 to -0.011)	0.001 ^a

a $p < 0.05$.

Figure 6 shows the raw, unadjusted complete-case mean utility scores at each time point by treatment group for each of the three methods: EQ-5D-5L TTO (primary economic analysis), EQ-5D-5L mapping and SF-12/SF-6D. The dashed lines are the discontinuation group and the continuous lines are the maintenance group. The highest values are for EQ-5D-5L TTO (primary analysis) and the lowest values are for SF-12/SF-6D.

Cost-utility analysis

The overall result of the cost-utility analysis is summarised as the ICER, which is the mean incremental cost per QALY gained of discontinuing antidepressant medication compared with maintenance antidepressant medication. In the primary economic analysis, with utilities and CSRI costs multiply imputed, discontinuation was dominated by maintenance in that the former cost more (£2.71, 95% CI -£36.10 to £37.07) and resulted in fewer QALYs (-0.019, 95% CI -0.035 to -0.003) than maintenance,



Months	0	3	6	9	12
Total pts reporting data	477	443	401	395	391

FIGURE 6 Raw, unadjusted complete-case mean utility scores at each time point by treatment group for each of the three methods. Adapted with permission from Clarke *et al.*⁷⁵ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. The figure includes minor additions and formatting changes to the original figure.

although the 95% CI crosses zero in the costs. This means that the bootstrapped differences in costs and QALYs lie predominantly in the north-west quadrant of the CEP (Figure 7), characterised by the ‘new’ intervention under assessment, in this case discontinuation of antidepressants, incurring higher costs and providing fewer QALYs than the ‘old’ treatment, which in this case is maintenance of antidepressants. ICERs for the primary and secondary analyses are given in Tables 20 (primary analysis using health-care and social care perspective) and 21 (wider societal perspective).

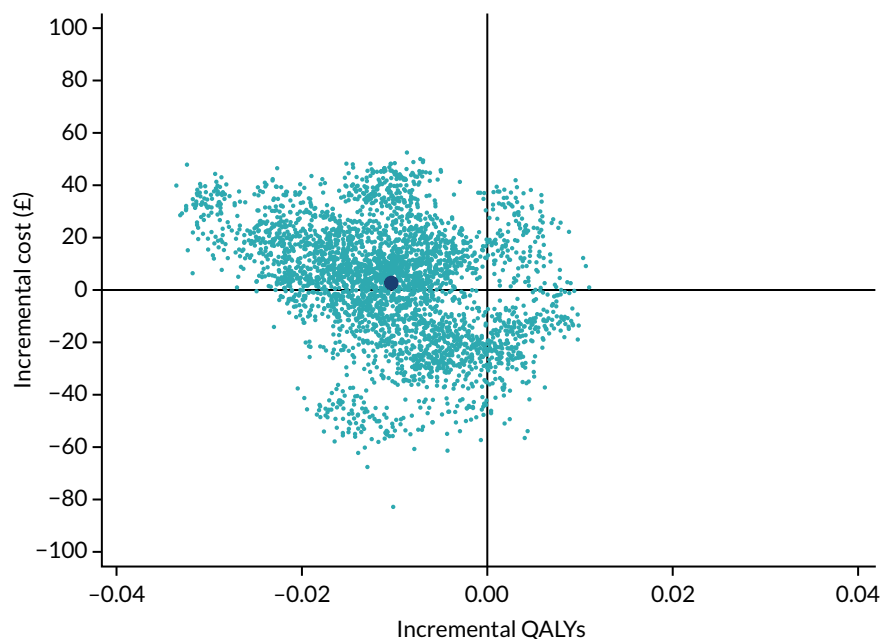


FIGURE 7 Cost-effectiveness plane using multiple imputation for utilities and CSRI costs, showing 1000 bootstrapped results from the seemingly unrelated regression of costs (NHS and PSS perspective) and QALYs (EQ-5D-5L TTO). Dark blue point is the mean costs vs. mean QALYs.

TABLE 20 Summary of the bootstrapped incremental costs and QALYs, and the ICERs, using the health-care and social care cost perspective

Quality-of-life instrument and value set	Difference (95% CI)		
	Costs (£)	QALYs	ICER
<i>Using MICE for missing utilities and CSRI costs</i>			
EQ-5D-5L TTO	2.71 (-36.10 to 37.07)	-0.010 (-0.024 to 0.004)	Maintenance dominates discontinuation
EQ-5D-5L mapping	4.19 (-31.42 to 36.40)	-0.012 (-0.028 to 0.004)	Maintenance dominates discontinuation
SF-12/SF-6D	2.16 (-36.58 to 36.41)	-0.020 (-0.033 to -0.008)	Maintenance dominates discontinuation
<i>Using complete-case analysis only</i>			
EQ-5D-5L TTO	26.29 (-3.32 to 57.29)	-0.019 (-0.032 to -0.005)	Maintenance dominates discontinuation
EQ-5D-5L mapping	26.22 (-3.61 to 57.29)	-0.022 (-0.038 to -0.006)	Maintenance dominates discontinuation
SF-12/SF-6D	23.93 (-6.66 to 56.67)	-0.026 (-0.039 to -0.013)	Maintenance dominates discontinuation
MICE, multiple imputation by chained equations.			

TABLE 21 Overall incremental costs and QALYs using the wider societal perspective

Quality-of-life instrument and value set	Difference (95% CI)		
	Costs (£)	QALYs	ICER
<i>Using MICE for missing utilities and CSRI costs</i>			
EQ-5D-5L TTO	22.30 (-179.25 to 218.76)	-0.011 (-0.024 to 0.003)	Maintenance dominates discontinuation
EQ-5D-5L mapping	21.46 (-168.86 to 210.15)	-0.012 (-0.029 to 0.003)	Maintenance dominates discontinuation
SF-12/SF-6D	21.42 (-179.47 to 216.89)	-0.020 (-0.033 to -0.008)	Maintenance dominates discontinuation
<i>Using complete-case analysis only</i>			
EQ-5D-5L TTO	105.99 (-128.37 to 331.22)	-0.019 (-0.032 to -0.006)	Maintenance dominates discontinuation
EQ-5D-5L mapping	106.02 (-128.38 to 331.39)	-0.022 (-0.039 to -0.006)	Maintenance dominates discontinuation
SF-12/SF-6D	101.48 (-142.45 to 341.27)	-0.026 (-0.040 to -0.013)	Maintenance dominates discontinuation
MICE, multiple imputation by chained equations.			

The information from the CEP was translated onto the CEAC (*Figure 8*), which shows the likelihood of discontinuation being cost-effective at a range of values of the cost-effectiveness threshold. *Figure 8* shows values up to £10,000 per QALY and that at the standard QALY threshold values of £20,000 and £30,000 per QALY gained there was a 12.9% and 12.4% probability that discontinuation was cost-effective compared with maintenance, respectively. The CEAC lies below the 50% region for all thresholds, in agreement with the conclusion that discontinuation is dominated by maintenance.

Secondary and sensitivity analyses

The results remained the same for all secondary analyses, including when productivity and out-of-pocket costs were included; discontinuation consistently results in higher cost and lower QALY gain than maintenance. In addition, when the SF-6D algorithm is used to calculate QALYs from SF-12 responses, the 95% CI for QALYs does not cross zero, which suggests that the QALY gain in the discontinuation group is significantly lower when evaluated in this way (see *Table 21*).

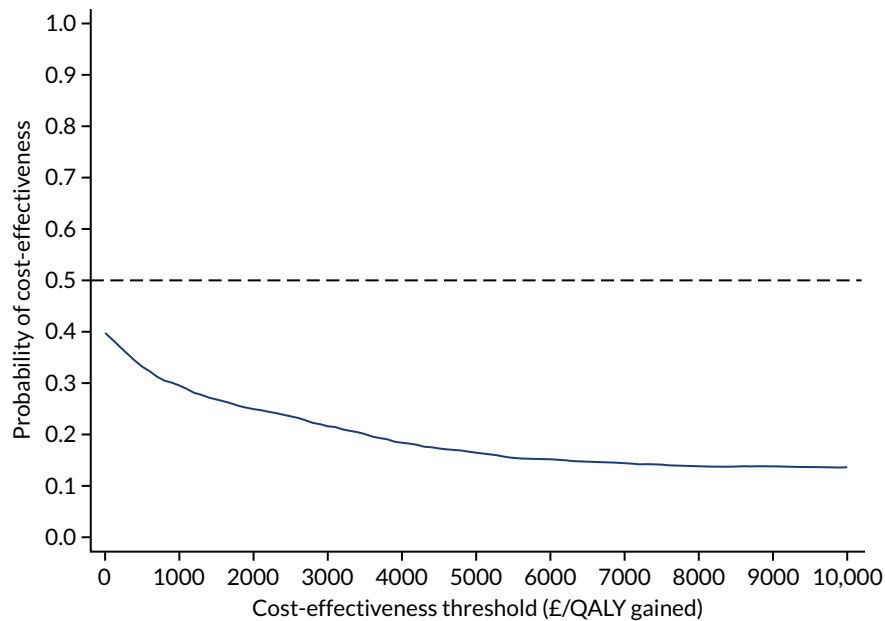


FIGURE 8 Cost-effectiveness acceptability curve generated from the CEP in *Figure 7* showing the likelihood of the discontinuation group being sufficiently cost-effective compared with the maintenance group, at a range of values for the cost-effectiveness threshold from zero to £10,000 per QALY.

When relapse was included in the bootstrapped regression analyses, the difference in utilities at 3 months was significant for both relapse (-0.050 , 95% CI -0.074 to -0.026) and treatment group (-0.030 , 95% CI -0.051 to -0.008). With the difference in QALYs over 12 months, there was a significant difference for relapse (-0.042 , 95% CI -0.058 to -0.026) but not for treatment group (-0.008 , 95% CI -0.025 to 0.008).

Relapse also had a significant impact on the cost of GP appointments (relapse cost an additional £34, 95% CI £18 to £51) and treatment group was no longer significant; this also followed through into a significant difference in total primary care cost between those who relapsed (more expensive by £50, 95% CI £15 to £85, over the year) and those who did not relapse. The difference in antidepressant medication cost was not explained by relapse status and only by treatment group. The total cost (primary care contacts, medications and CSRI-collected costs) was significantly different according to relapse status, with those who relapsed costing £67 (95% CI £23 to £111) more overall over the year than those who did not relapse. CEPs and CEACs for the primary and secondary analyses are presented in *Report Supplementary Material 8*.

Chapter 5 Discussion and conclusions

Summary of findings

Our large pragmatic trial of the effectiveness of maintenance antidepressants in primary care found that the rate of relapse was twice as high in those who received placebo as in those who continued to take antidepressant medication during 12 months of follow-up (hazard ratio 2.04, 95% CI 1.55 to 2.68). Most relapses occurred 12–26 weeks after randomisation and 4–18 weeks after medication was tapered.

Our analyses of secondary outcomes found strong evidence that depressive and anxiety symptoms at 12 weeks were higher in those receiving the placebo than in those receiving antidepressant medication, and the size of this difference was likely to be clinically meaningful.⁷⁶ Weaker evidence of a smaller difference in depressive and anxiety symptoms remained at 6 months, but attenuated thereafter. People in the discontinuation group were more likely than those in the maintenance group to self-report worsening of mood and had lower mental health-related quality of life at 12 weeks, although these differences were not evident at any other time point. At 12 weeks, 44% of those who discontinued their antidepressants reported feeling worse, compared with 21% of those who remained on their medication. Such global changes in mental health are used to calculate minimal clinically important differences and are regarded as patient-centred indicators of clinically meaningful change.⁷⁷ We also found strong evidence that patients who discontinued antidepressants stopped their trial medication sooner than those who remained on maintenance treatment (HR 2.26, 95% CI 1.67 to 3.07). By the end of the trial, 39% (95% CI 32% to 45%) of participants in the discontinuation group had returned to an active antidepressant prescribed by their clinician, which may explain why treatment effects attenuated after the 12-week follow-up.

People who discontinued their antidepressant experienced more withdrawal symptoms than those who remained on maintenance treatment at every time point except 52 weeks, by which time the difference had attenuated substantially (and many people in the discontinuation group had returned to taking an antidepressant). The difference in withdrawal symptoms between the groups was largest at 12 weeks and reduced as follow-up progressed. There was no evidence of a difference in physical symptoms that could be attributed to antidepressant side effects (in contrast to withdrawal effects) until 39 and 52 weeks after randomisation, when those receiving an antidepressant reported slightly more physical symptoms than those receiving placebo. It is possible that any difference in side effects was masked by the increase in depressive and/or withdrawal symptoms in the discontinuation group.

Summary of economic evaluation findings

The cost–utility analysis suggests that there is a low probability that discontinuing antidepressant medications with replacement by placebo is cost-effective compared with continued maintenance of antidepressant treatment in this population. Participants randomised to the discontinuation group experienced significantly fewer QALYs over 12 months than those in the maintenance group, with this difference primarily driven by an increased rate of relapse. Participants who were randomised to the discontinuation group also had higher GP consultation costs and Improving Access to Psychological Therapies costs. The difference in GP costs appeared to be driven by a shorter time to relapse in the discontinuation group, potentially because participants arranged to see their GP following relapse to review their medication, with 53% (95% CI 44% to 62%) of those who relapsed in the discontinuation group having returned to a known antidepressant before the end of the trial.

A limitation of the analysis is that the time horizon is the same as the follow-up period for the trial (i.e. only 12 months). It is possible that there are longer-term impacts on costs and QALYs, such as medication side effects, feelings of stability derived from maintenance medication or feelings of liberation derived from having become 'medication free', that are not captured by this analysis. However, decision analytic models of depression rarely go beyond 12 months, and we are not aware of any other studies with longer time horizons on which a decision model with a longer time horizon could be based; therefore, any modelling would have required many untested assumptions. In addition, given the relatively large difference in QALYs between tapering and maintenance compared with the relatively small difference in costs, it seems unlikely that discontinuation could potentially be cost-effective over the long term. Missing values also represent a limitation to the analysis; this analysis used multiple imputation, which carries the assumption that values are missing at random, but we cannot know whether or not this is correct. A further limitation of the analysis is that the effectiveness measures used were generic health-related quality-of-life measures and, although there is no evidence to suggest that other quality-of-life measures might perform better in this context, it is still possible that information regarding other factors that are important to participants was not captured. Finally, we compared maintenance antidepressants with the replacement of antidepressants with a placebo over a tapering period, although the use of placebo after discontinuation is not an indicated treatment. We chose to include a placebo to ensure that we were studying the pharmacological effects of the antidepressant drug, which is essential to inform policy. However, it is likely that the difference between maintenance and discontinuation without a placebo would be larger.

At a £20,000 to £30,000 threshold for a QALY gain (i.e. the preferred NICE threshold⁷⁸) there would need to be a cost saving of £200–600 per person for tapering over the lifetime horizon following the first 12 months to balance the QALYs lost during the first 12-month period by discontinuing the medication; it is not clear where in the longer-term patient pathway this could potentially occur. This highlights the need for further research into the longer-term implications of discontinuing antidepressant use compared with continuing with long-term maintenance for the prevention of relapse, in terms of both longer-term costs and longer-term benefits to patients.

Overall, there was no substantial cost to the health-care system of discontinuing antidepressants; there were costs only to patients in terms of their well-being. Although the evidence from our analysis would not support recommendations at a national level to discontinue antidepressants based on cost-effectiveness, there is the opportunity to provide participants with the information that they require to make a more informed choice about their continued antidepressant prescription. There was a difference in utility scores at 3 months, although this difference had disappeared by 12 months, meaning that any detriment due to relapse or tapering the original medication, although significant, was, on average, short-lived. Some relapses, however, can be potentially severe and disabling, and our analysis offers no prediction of the impact on any particular person. Patients who wish to discontinue their antidepressant, and are willing to accept the risk of a shorter time to relapse, have the possible future benefit of eventually being medication free. The key costs would be mainly to themselves in terms of an increased likelihood of a relapse that could be potentially severe, for a short-term reduction in health-related quality of life, plus health service costs of an additional GP appointment to manage a potential relapse and other treatment options, such as cognitive-behavioural therapy. In societal terms, we also did not find a significant impact on productivity, partly because of the very wide CI around the result.

Strength and limitations

To our knowledge, the ANTLER trial is the largest individual randomised trial of antidepressant maintenance treatment that is not funded by the pharmaceutical industry. Clinical trials are often criticised for using narrow inclusion criteria, which can reduce external validity.⁷⁹ Our design ensured that we recruited people currently receiving long-term maintenance antidepressants in primary care. Participants had been receiving antidepressant medication for ≥ 9 months (the majority for > 3 years),

which, to the best of our knowledge, was a longer treatment period than any previous trial. Our results are, therefore, more readily generalisable than the results of previous trials to the population currently receiving long-term maintenance antidepressants in primary care. We used the four most commonly prescribed antidepressants in primary care. Given that the vast majority of antidepressant prescriptions are for long-term treatment, including maintenance, we think that our results will apply to most people receiving maintenance treatment. We investigated three SSRIs, which have a similar pharmacological profile and act via similar mechanisms. Our results should, therefore, be generalisable to other SSRIs when used in this population. Although there are similarities in the mechanism of action of all commonly used antidepressants, it is more difficult to generalise our findings to other classes of antidepressant. We were unable to investigate whether or not treatment effects differed for mirtazapine because of the small number of participants receiving this antidepressant. We investigated the usual doses of antidepressants used in the UK. Our results should generalise to higher doses, given that there is no evidence of added efficacy,²² but it is difficult to generalise our findings to maintenance treatment at lower doses.

Attrition is a limitation of most randomised controlled trials (RCTs), and more participants in the discontinuation group than in the maintenance group dropped out. However, attrition was modest and results were unaltered after we adjusted for variables associated with missing data. Participants in the discontinuation group correctly guessed which group they were in more often than those in the maintenance group. This is consistent with prior studies suggesting that patients can distinguish placebo from active treatment. In principle, this could bias the outcome, but it could also be a consequence of relapse or withdrawal symptoms resulting from allocation to discontinuation (and on the causal pathway from exposure to outcome).

Although our design was embedded in clinical practice, it introduces some potential limitations. We had to ask retrospectively about past history of illness and treatment. In particular, we do not have detailed information of the original clinical decision for prescribing the antidepressant or any diagnostic information at that time. Participants who had more recently become well may have been more vulnerable to relapse, although this is unlikely to have biased the treatment effect owing to randomisation. Although our sample is likely to be more representative than prior trials, only a small proportion of those potentially eligible participated, which is common in RCTs. We also recruited people who had experienced at least two prior depressive episodes, so our findings do not necessarily apply to those receiving treatment for their first depressive episode. Participants were also mostly white, married and employed, and were recruited predominantly from moderately sized general practices in urban areas. Those who participated were older than those invited who did not participate. This limits generalisability but is still an improvement on existing literature and provides more realistic estimates of what might happen if someone on long-term maintenance treatment stops their antidepressant, at least for the first 12 months after discontinuation. A more important limitation is the ability to generalise to other health systems, although there are striking similarities in how antidepressants are used in wealthy countries.⁸⁰

In our trial, we used a novel assessment (the rCIS-R) to measure the reappearance of depressive symptoms. The results of our test-retest reliability study nested within the ANTLER trial provide strong evidence that the rCIS-R is a reliable measure of assessing reappearance of depressive symptoms. There was excellent agreement for definitions of relapse, the individual symptoms that were assessed and the sum of the symptoms scores. The main advantage of the rCIS-R is that, to our knowledge, it is the only fully structured interview assessing time to relapse. The pragmatic advantages that deserve considerations are time required for completion, simplicity of scoring and absence of any special training requirements.

Implications for health care

Our findings have several implications for the management of depression in primary care. Our evidence that discontinuing long-term maintenance antidepressants increases the risk of a clinically significant relapse in the following 12 months suggests that, for many patients, long-term treatment is appropriate.

However, the severity of relapse varies. Although 56% of people who discontinued their treatment experienced a relapse, only around half of them (53%, 95% CI 44% to 62%) chose to return to an antidepressant prescribed by their clinician. It is possible that some relapses, or possible withdrawal symptoms, were not severe enough for the individual to decide that they needed to return to their medication. If six (95% CI 3 to 19) people stopped their medication, we estimated that one would experience a relapse who may not have experienced a relapse if they had remained on maintenance treatment. Our results illustrate that remaining on maintenance antidepressants does not guarantee well-being and is offset by any adverse events and the reluctance of many people to stay on medication for many years. If people who want to discontinue their antidepressants are regularly monitored in primary care during the first 6 months, relapse prevention may be possible with alternative treatments, such as psychological therapies. For example, there is good evidence that mindfulness-based cognitive therapy (with support to taper or discontinue antidepressants) is as effective as maintenance antidepressants at preventing relapse.⁵³

Current best practice is to engage with patients' priorities and collaborate in coming to a decision about medication. For the individual patient, it is possible to know only about the average likelihood of relapse and that the severity of potential relapses will be unpredictable. Our findings will give patients and physicians an estimate of the likely benefits and harms of stopping long-term maintenance antidepressants to inform shared decision-making in primary care.

Recommendation for research

The ANTLER trial identified the following research needs:

- Further research into the longer-term implications of tapering antidepressant use compared with continuing with long-term maintenance for the prevention of relapse, in terms of both longer-term costs and longer-term benefits to patients.
- In addition to having longer follow-up duration, it would be beneficial if future studies were larger so that differences according to antidepressant type could be identified.
- Further research into distinguishing between withdrawal symptoms and depressive symptoms in the first few weeks of tapering antidepressant medication. This could involve experimental designs, new measurement instruments or factor analyses.
- The rCIS-R was designed for research purposes, although it may also have application in clinical practice as a simple way of assessing relapse in depression. This could be investigated in future studies.

Acknowledgements

We are grateful to all of the patients who took part in the ANTLER trial. We thank the staff in participating GP surgeries for their help with recruitment. We also thank colleagues who have contributed to the study through recruitment, administrative help and other advice; in particular Carmen Sinclair, Nomsa Chari, Paula Beharry, Jane Hahn, Vivien Jones, Catherine Derrick and Mahsa Rezaei. We have been supported by the following Clinical Research Networks (CRNs): North Thames CRN; CRN North West London; CRN South London; Thame Valley and South Midlands CRN; Luton, Essex and Herts Valley CRN; West of England CRN; and Wessex CRN. We especially thank the following CRN staff: Debbie Kelly, Claire Winch, Zara Prem, Andrei Gabzdyl and Lynsey Wilson. We also thank Allan House, Geoffrey Wong, Paul Lanham, Jonathan Bisson, Lucy Carr, Chris Dowrick, Rafael Perera-Salazar and Mike Crawford for generously agreeing to sit on either the Trial Steering Committee or the Data Monitoring committee.

We acknowledge the support of the National Institute for Health Research University College London Hospitals Biomedical Research Centre. This study was also supported by the NIHR Biomedical Research Centre at University Hospitals Bristol and Weston NHS Foundation Trust and the University of Bristol.

Contributions of authors

Larisa Duffy (<https://orcid.org/0000-0002-0093-3877>) was the trial manager, conducted the analysis of the rCIS-R reliability data and was responsible for drafting and collating the report.

Caroline S Clarke (<https://orcid.org/0000-0002-4676-1257>) conducted the health economic analysis and with **Rachael Hunter** (<https://orcid.org/0000-0002-7447-8934>) drafted the health economic chapter.

Louise Marston (<https://orcid.org/0000-0002-9973-1131>) with input from **Nick Freemantle** (<https://orcid.org/0000-0001-5807-5740>) conducted both primary and secondary analysis and together with **Gemma Lewis** (<https://orcid.org/0000-0001-6666-3681>) carried out additional analyses and drafted the results section.

Gemma Lewis drafted the discussion section.

Simon Gilbody (<https://orcid.org/0000-0002-8236-6983>) had responsibility for the York site.

Tony Kendrick (<https://orcid.org/0000-0003-1618-9381>) and **Michael Moore** (<https://orcid.org/0000-0002-5127-4509>) had responsibility for the Southampton site.

David Kessler (<https://orcid.org/0000-0001-5333-132X>) and **Nicola Wiles** (<https://orcid.org/0000-0002-5250-3553>) had responsibility for the Bristol site.

Nick Freemantle, Simon Gilbody, Rachael Hunter, Tony Kendrick, David Kessler, **Michael King** (<https://orcid.org/0000-0003-4715-7171>), **Paul Lanham** (<https://orcid.org/0000-0001-6864-3318>), Michael Moore, Nicola Wiles, **Irwin Nazareth** (<https://orcid.org/0000-0003-2146-9628>) and **Glyn Lewis** (<https://orcid.org/0000-0001-5205-8245>) were responsible for the original proposal securing funding.

Glyn Lewis was chief investigator of the trial and had clinical responsibility for the RCT.

ACKNOWLEDGEMENTS

All co-applicants with input from Larisa Duffy, Dee Mangin (<https://orcid.org/0000-0003-2149-9376>), Faye Bacon (<https://orcid.org/0000-0002-7566-7585>), Molly Bird (<https://orcid.org/0000-0001-5570-1419>), Sally Brabyn (<https://orcid.org/0000-0001-5381-003X>), Alison Burns (<https://orcid.org/0000-0002-2242-3499>), Yvonne Donkor (<https://orcid.org/0000-0003-2716-7743>), Anna Hunt (<https://orcid.org/0000-0002-0864-4113>) and Jodi Pervin (<https://orcid.org/0000-0003-2452-2391>) designed the trial and developed detailed protocol.

All authors have provided substantial contribution to the conception of the ANTLER trial, interpretation of data and had input into drafting the report and/or revising it critically for important intellectual content. All authors have given final approval of the version to be published.

Publications

Lewis G, Marston L, Duffy L, Freemantle N, Gilbody S, Hunter R, *et al.* Maintenance or discontinuation of antidepressants in primary care. *N Engl J Med* 2021;**385**:1275–67.

Clarke CS, Duffy L, Lewis G, Freemantle N, Gilbody S, Kendrick T, *et al.* Cost-Utility Analysis of Discontinuing Antidepressants in England Primary Care Patients Compared with Long-Term Maintenance: The ANTLER Study [published online ahead of print November 8 2021]. *Appl Health Econ Health Policy* 2021.

Data-sharing statement

We are open to collaborations with other scientists as our ethics permission includes only use of the data by the research team. Please address your enquires to the corresponding author.

References

1. World Health Organization. *Depression: Key Facts*. 2018. URL: www.who.int/news-room/fact-sheets/detail/depression (accessed 3 September 2021).
2. Pilling S, Anderson I, Goldberg D, Meader N, Taylor C, Two Guideline Development Groups. Depression in adults, including those with a chronic physical health problem: summary of NICE guidance. *BMJ* 2009;**339**:b4108. <https://doi.org/10.1136/bmj.b4108>
3. Lewinsohn PM, Solomon A, Seeley JR, Zeiss A. Clinical implications of 'subthreshold' depressive symptoms. *J Abnorm Psychol* 2000;**109**:345–51. <https://doi.org/10.1037/0021-843X.109.2.345>
4. McCrea RL, Sammon CJ, Nazareth I, Petersen I. Initiation and duration of selective serotonin reuptake inhibitor prescribing over time: UK cohort study. *Br J Psychiatry* 2016;**209**:421–6. <https://doi.org/10.1192/bjp.bp.115.166975>
5. Moore M, Yuen HM, Dunn N, Mullee MA, Maskell J, Kendrick T. Explaining the rise in antidepressant prescribing: a descriptive study using the general practice research database. *BMJ* 2009;**339**:b3999. <https://doi.org/10.1136/bmj.b3999>
6. Wiles N, Thomas L, Abel A, Ridgway N, Turner N, Campbell J, et al. Cognitive behavioural therapy as an adjunct to pharmacotherapy for primary care based patients with treatment resistant depression: results of the CoBaT randomised controlled trial. *Lancet* 2013;**381**:375–84. [https://doi.org/10.1016/S0140-6736\(12\)61552-9](https://doi.org/10.1016/S0140-6736(12)61552-9)
7. Viola S, Moncrieff J. Claims for sickness and disability benefits owing to mental disorders in the UK: trends from 1995 to 2014. *BJPsych Open* 2016;**2**:18–24. <https://doi.org/10.1192/bjpo.bp.115.002246>
8. McManus S, Bebbington P, Jenkins R, Brugha T. Mental Health and Wellbeing in England. *Adult Psychiatric Morbidity Survey 2014*. 2016. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/556596/apms-2014-full-rpt.pdf (accessed 3 September 2021).
9. Hewlett E, Moran V. *Making Mental Health Count: The Social and Economic Costs of Neglecting Mental Health Care*. Paris: OECD Health Policy Studies, OECD Publishing, Paris; 2014. <https://doi.org/10.1787/9789264208445-en>
10. Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet* 2018;**391**:1357–66. [https://doi.org/10.1016/S0140-6736\(17\)32802-7](https://doi.org/10.1016/S0140-6736(17)32802-7)
11. Geddes JR, Carney SM, Davies C, Furukawa TA, Kupfer DJ, Frank E, Goodwin GM. Relapse prevention with antidepressant drug treatment in depressive disorders: a systematic review. *Lancet* 2003;**361**:653–61. [https://doi.org/10.1016/S0140-6736\(03\)12599-8](https://doi.org/10.1016/S0140-6736(03)12599-8)
12. Glue P, Donovan MR, Kolluri S, Emir B. Meta-analysis of relapse prevention antidepressant trials in depressive disorders. *Aust N Z J Psychiatry* 2010;**44**:697–705. <https://doi.org/10.3109/00048671003705441>
13. Kaymaz N, van Os J, Loonen AJ, Nolen WA. Evidence that patients with single versus recurrent depressive episodes are differentially sensitive to treatment discontinuation: a meta-analysis of placebo-controlled randomized trials. *J Clin Psychiatry* 2008;**69**:1423–36. <https://doi.org/10.4088/JCP.v69n0910>

14. Sim K, Lau WK, Sim J, Sum MY, Baldessarini RJ. Prevention of relapse and recurrence in adults with major depressive disorder: systematic review and meta-analyses of controlled trials. *Int J Neuropsychopharmacol* 2015;**19**:pyv076. <https://doi.org/10.1093/ijnp/pyv076>
15. Johnson CF, Macdonald HJ, Atkinson P, Buchanan AI, Downes N, Dougall N. Reviewing long-term antidepressants can reduce drug burden: a prospective observational cohort study. *Br J Gen Pract* 2012;**62**:e773–9. <https://doi.org/10.3399/bjgp12X658304>
16. Mojtabai R, Olfson M. National trends in long-term use of antidepressant medications: results from the US National Health and Nutrition Examination Survey. *J Clin Psychiatry* 2014;**75**:169–77. <https://doi.org/10.4088/JCP.13m08443>
17. Cook BL, Helms PM, Smith RE, Tsai M. Unipolar depression in the elderly. Reoccurrence on discontinuation of tricyclic antidepressants. *J Affect Disord* 1986;**10**:91–4. [https://doi.org/10.1016/0165-0327\(86\)90031-5](https://doi.org/10.1016/0165-0327(86)90031-5)
18. Kupfer DJ, Frank E, Perel JM, Cornes C, Mallinger AG, Thase ME, et al. Five-year outcome for maintenance therapies in recurrent depression. *Arch Gen Psychiatry* 1992;**49**:769–73. <https://doi.org/10.1001/archpsyc.1992.01820100013002>
19. Bialos D, Giller E, Jatlow P, Docherty J, Harkness L. Recurrence of depression after discontinuation of long-term amitriptyline treatment. *Am J Psychiatry* 1982;**139**:325–9. <https://doi.org/10.1176/ajp.139.3.325>
20. National Institute for Health and Care Excellence. *Depression: Management of Depression in Primary and Secondary Care*. Clinical Guidelines 23. London: NICE; 2010.
21. Marsden J, White M, Annand F, Burkinshaw P, Carville S, Eastwood B, et al. Medicines associated with dependence or withdrawal: a mixed-methods public health review and national database study in England. *Lancet Psychiatry* 2019;**6**:935–50. [https://doi.org/10.1016/S2215-0366\(19\)30331-1](https://doi.org/10.1016/S2215-0366(19)30331-1)
22. Furukawa TA, Cipriani A, Cowen PJ, Leucht S, Egger M, Salanti G. Optimal dose of selective serotonin reuptake inhibitors, venlafaxine, and mirtazapine in major depression: a systematic review and dose-response meta-analysis. *Lancet Psychiatry* 2019;**6**:601–9. [https://doi.org/10.1016/S2215-0366\(19\)30217-2](https://doi.org/10.1016/S2215-0366(19)30217-2)
23. Tallon D, Wiles N, Campbell J, Chew-Graham C, Dickens C, Macleod U, et al. Mirtazapine added to selective serotonin reuptake inhibitors for treatment-resistant depression in primary care (MIR trial): study protocol for a randomised controlled trial. *Trials* 2016;**17**:66. <https://doi.org/10.1186/s13063-016-1199-2>
24. OpenPrescribing. *Search GP Prescribing Data*. 2020. URL: <https://openprescribing.net/analyse> (accessed 3 October 2021).
25. 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**:465–86. <https://doi.org/10.1017/s0033291700030415>
26. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;**16**:606–13. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>
27. Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. *J Gen Intern Med* 2007;**22**:1596–602. <https://doi.org/10.1007/s11606-007-0333-y>
28. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* 2006;**166**:1092–7. <https://doi.org/10.1001/archinte.166.10.1092>

29. Vanderkooy JD, Kennedy SH, Bagby RM. Antidepressant side effects in depression patients treated in a naturalistic setting: a study of bupropion, moclobemide, paroxetine, sertraline, and venlafaxine. *Can J Psychiatry* 2002;**47**:174–80. <https://doi.org/10.1177/070674370204700208>
30. Crawford AA, Lewis S, Nutt D, Peters TJ, Cowen P, O'Donovan MC, *et al.* Adverse effects from antidepressant treatment: randomised controlled trial of 601 depressed individuals. *Psychopharmacology* 2014;**231**:2921–31. <https://doi.org/10.1007/s00213-014-3467-8>
31. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996;**34**:220–33. <https://doi.org/10.1097/00005650-199603000-00003>
32. 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. [https://doi.org/10.1016/S0006-3223\(98\)00126-7](https://doi.org/10.1016/S0006-3223(98)00126-7)
33. Duffy L, Lewis G, Ades A, Araya R, Bone, JK, Brabyn, S *et al.* Antidepressant treatment with sertraline for adults with depressive symptoms in primary care: the PANDA research programme including RCT. *Programme Grants Appl Res* 2019;**7**(10). <https://doi.org/10.3310/pgfar07100>
34. Lewis G, Marston L, Duffy L, Freemantle N, Gilbody S, Hunter R, *et al.* Maintenance or discontinuation of antidepressants in primary care. *N Engl J Med* 2021;**385**:1275–67. <https://doi.org/10.1056/NEJMoa2106356>
35. Tiddeman B, Burt M, Perrett D. Prototyping and transforming facial textures for perception research. *IEEE Comput Graph Appl* 2001;**21**:42–50. <https://doi.org/10.1109/38.946630>
36. Harmer CJ, Goodwin GM, Cowen PJ. Why do antidepressants take so long to work? A cognitive neuropsychological model of antidepressant drug action. *Br J Psychiatry* 2009;**195**:102–8. <https://doi.org/10.1192/bjp.bp.108.051193>
37. Guitart-Masip M, Economides M, Huys QJM, *et al.* Differential, but not opponent, effects of l-DOPA and citalopram on action learning with reward and punishment. *Psychopharmacology* 2014;**231**:955–66. <https://doi.org/10.1007/s00213-013-3313-4>
38. Shih JH. Sample size calculation for complex clinical trials with survival endpoints. *Control Clin Trials* 1995;**16**:395–407. [https://doi.org/10.1016/S0197-2456\(95\)00132-8](https://doi.org/10.1016/S0197-2456(95)00132-8)
39. Twisk 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. <https://doi.org/10.1016/j.conctc.2018.03.008>
40. Frank E, Prien RF, Jarrett RB, Keller MB, Kupfer DJ, Lavori PW, *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. <https://doi.org/10.1001/archpsyc.1991.01810330075011>
41. Rush AJ, Kraemer HC, Sackeim HA, Fava M, Trivedi MH, Frank E, *et al.* Report by the ACNP Task Force on response and remission in major depressive disorder. *Neuropsychopharmacology* 2006;**31**:1841–53. <https://doi.org/10.1038/sj.npp.1301131>
42. Burcusa SL, Iacono WG. Risk for recurrence in depression. *Clin Psychol Rev* 2007;**27**:959–85. <https://doi.org/10.1016/j.cpr.2007.02.005>
43. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960;**23**:56–62. <https://doi.org/10.1136/jnnp.23.1.56>
44. Montgomery A, Asbery M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979;**134**:382–9. <https://doi.org/10.1192/bjp.134.4.382>

REFERENCES

45. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry* 1961;**4**:561–71. <https://doi.org/10.1001/archpsyc.1961.01710120031004>
46. Malpass A, Dowrick C, Gilbody S, Robinson J, Wiles N, Duffy L, Lewis G. Usefulness of PHQ-9 in primary care to determine meaningful symptoms of low mood: a qualitative study. *Br J Gen Pract* 2016;**66**:e78–84. <https://doi.org/10.3399/bjgp16X683473>
47. World Health Organization (WHO). *Composite Diagnostic Interview (CIDI), Ver. 1.0*. Geneva: WHO; 1990.
48. Wittchen HU, Robins LN, Cottler LB, Sartorius N, Burke JD, Regier D. Cross-cultural feasibility, reliability and sources of variance of the Composite International Diagnostic Interview (CIDI). The Multicentre WHO/ADAMHA Field Trials. *Br J Psychiatry* 1991;**159**:645–53, 658. <https://doi.org/10.1192/bjp.159.5.645>
49. Michael B First Robert L. Spitzer, Miriam Gibbon and JBWW. The Structured Clinical Interview for DSM-III-R Personality Disorders (SCID-II). Part I: Description. *J Pers Disord* 1995;**9**:83–91. <https://doi.org/10.1521/pedi.1995.9.2.83>
50. Lobbstaël J, Leurgans M, Arntz A. Inter-rater reliability of the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID I) and Axis II Disorders (SCID II). *Clin Psychol Psychother* 2011;**18**:75–9. <https://doi.org/10.1002/cpp.693>
51. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;**1**:307–10. [https://doi.org/10.1016/S0140-6736\(86\)90837-8](https://doi.org/10.1016/S0140-6736(86)90837-8)
52. Chalder M, Wiles NJ, Campbell J, Hollinghurst SP, Haase AM, Taylor AH, et al. Facilitated physical activity as a treatment for depressed adults: randomised controlled trial. *BMJ* 2012;**344**:e2758. <https://doi.org/10.1136/bmj.e2758>
53. Kuyken W, Hayes R, Barrett B, Byng R, Dalgleish T, Kessler D, et al. Effectiveness and cost-effectiveness of mindfulness-based cognitive therapy compared with maintenance antidepressant treatment in the prevention of depressive relapse or recurrence (PREVENT): a randomised controlled trial. *Lancet* 2015;**386**:63–73. [https://doi.org/10.1016/S0140-6736\(14\)62222-4](https://doi.org/10.1016/S0140-6736(14)62222-4)
54. Ramsberg J, Asseburg C, Henriksson M. Effectiveness and cost-effectiveness of antidepressants in primary care: a multiple treatment comparison meta-analysis and cost-effectiveness model. *PLOS ONE* 2012;**7**:e42003. <https://doi.org/10.1371/journal.pone.0042003>
55. Annemans L, Brignone M, Druais S, De Pauw A, Gauthier A, Demyttenaere K. Cost-effectiveness analysis of pharmaceutical treatment options in the first-line management of major depressive disorder in Belgium. *Pharmacoeconomics* 2014;**32**:479–93. <https://doi.org/10.1007/s40273-014-0138-x>
56. Beecham J, Knapp M. Costing psychiatric interventions. *Meas Mental Heal Needs* 2001;**2**:200–24.
57. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res* 2011;**20**:1727–36. <https://doi.org/10.1007/s11136-011-9903-x>
58. Joint Formulary Committee. *British National Formulary* (online). London: BMJ Group and Pharmaceutical Press. URL: www.medicinescomplete.com (accessed 10 April 2019).
59. Curtis L, Burns A. *Unit Costs of Health and Social Care 2019*. Canterbury: PSSRU, University of Kent; 2019.
60. Department of Health and Social Care (DHSC). *NHS Reference Costs 2018/19*. London: DHSC; 2019.
61. Curtis L, Burns A. *Unit Costs of Health and Social Care 2015*. Canterbury: PSSRU, University of Kent; 2015.

62. Isaacs AJ, Critchley JA, Tai SS, Buckingham K, Westley D, Harridge SD, *et al.* Exercise Evaluation Randomised Trial (EXERT): a randomised trial comparing GP referral for leisure centre-based exercise, community-based walking and advice only. *Health Technol Assess* 2007;**11**(10). <https://doi.org/10.3310/hta11100>
63. Pope I, Burn H, Ismail SA, Harris T, McCoy D. A qualitative study exploring the factors influencing admission to hospital from the emergency department. *BMJ Open* 2017;**7**:e011543. <https://doi.org/10.1136/bmjopen-2016-011543>
64. Devlin NJ, Krabbe PF. The development of new research methods for the valuation of EQ-5D-5L. *Eur J Health Econ* 2013;**14**(Suppl. 1):1–3. <https://doi.org/10.1007/s10198-013-0502-3>
65. van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, *et al.* Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health* 2012;**15**:708–15. <https://doi.org/10.1016/j.jval.2012.02.008>
66. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care* 2004;**42**:851–9. <https://doi.org/10.1097/01.mlr.0000135827.18610.0d>
67. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: an EQ-5D-5L value set for England. *Health Econ* 2018;**27**:7–22. <https://doi.org/10.1002/hec.3564>
68. Hunter RM, Baio G, Butt T, Morris S, Round J, Freemantle N. An educational review of the statistical issues in analysing utility data for cost-utility analysis. *Pharmacoeconomics* 2015;**33**:355–66. <https://doi.org/10.1007/s40273-014-0247-6>
69. Zellner A. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *J Am Stat Assoc* 1962;**57**:348–68. <https://doi.org/10.1080/01621459.1962.10480664>
70. Rubin D. *Multiple Imputation for Nonresponse in Surveys*. New York, NY: John Wiley & Sons, Inc.; 1987. <https://doi.org/10.1002/9780470316696>
71. Leurent B, Gomes M, Faria R, Morris S, Grieve R, Carpenter JR. Sensitivity analysis for not-at-random missing data in trial-based cost-effectiveness analysis: a tutorial. *Pharmacoeconomics* 2018;**36**:889–901. <https://doi.org/10.1007/s40273-018-0650-5>
72. Duffy L, Bacon F, Clarke CS, Donkor Y, Freemantle N, Gilbody S, *et al.* A randomised controlled trial assessing the use of citalopram, sertraline, fluoxetine and mirtazapine in preventing relapse in primary care patients who are taking long-term maintenance antidepressants (ANTLER: ANTidepressants to prevent relapse in dEpReSSION): study protocol for a randomised controlled trial. *Trials* 2019;**20**:1–13. <https://doi.org/10.1186/s13063-019-3390-8>
73. Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997;**35**:1095–108. <https://doi.org/10.1097/00005650-199711000-00002>
74. Franklin M, Enrique A, Palacios J, Richards D. Psychometric assessment of EQ-5D-5L and ReQoL measures in patients with anxiety and depression: construct validity and responsiveness. *Qual Life Res* 2021;**30**:2633–47. <https://doi.org/10.1007/s11136-021-02833-1>
75. Clarke CS, Duffy L, Lewis G, Freemantle N, Gilbody S, Kendrick T, *et al.* Cost-Utility Analysis of Discontinuing Antidepressants in England Primary Care Patients Compared with Long-Term Maintenance: The ANTLER Study [published online ahead of print November 8 2021]. *Appl Health Econ Health Policy* 2021. <https://doi.org/10.1007/s40258-021-00693-x>
76. Lewis G, Duffy L, Ades A, Amos R, Araya R, Brabyn S, *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. *Lancet Psychiatry* 2019;**6**:903–14. [https://doi.org/10.1016/S2215-0366\(19\)30366-9](https://doi.org/10.1016/S2215-0366(19)30366-9)

REFERENCES

77. Button KS, Kounali D, Thomas L, Wiles NJ, Peters TJ, Welton NJ, *et al.* Minimal clinically important difference on the Beck Depression Inventory – II according to the patient’s perspective. *Psychol Med* 2015;**45**:3269–79. <https://doi.org/10.1017/S0033291715001270>
78. National Institute for Health and Care Excellence (NICE). *Guide to the Methods of Technology Appraisal* 2013. London: NICE; 2013.
79. Hotopf M. The pragmatic randomised controlled trial. *Adv Psychiatr Treat* 2002;**8**:326–33. <https://doi.org/10.1192/apt.8.5.326>
80. Gusmão R, Quintão S, McDaid D, Arensman E, Van Audenhove C, Coffey C, *et al.* Antidepressant utilization and suicide in Europe: an ecological multi-national study. *PLOS ONE* 2013;**8**:e66455. <https://doi.org/10.1371/journal.pone.0066455>

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

*This report presents independent research funded by the National Institute for Health Research (NIHR).
The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the
Department of Health and Social Care*

Published by the NIHR Journals Library