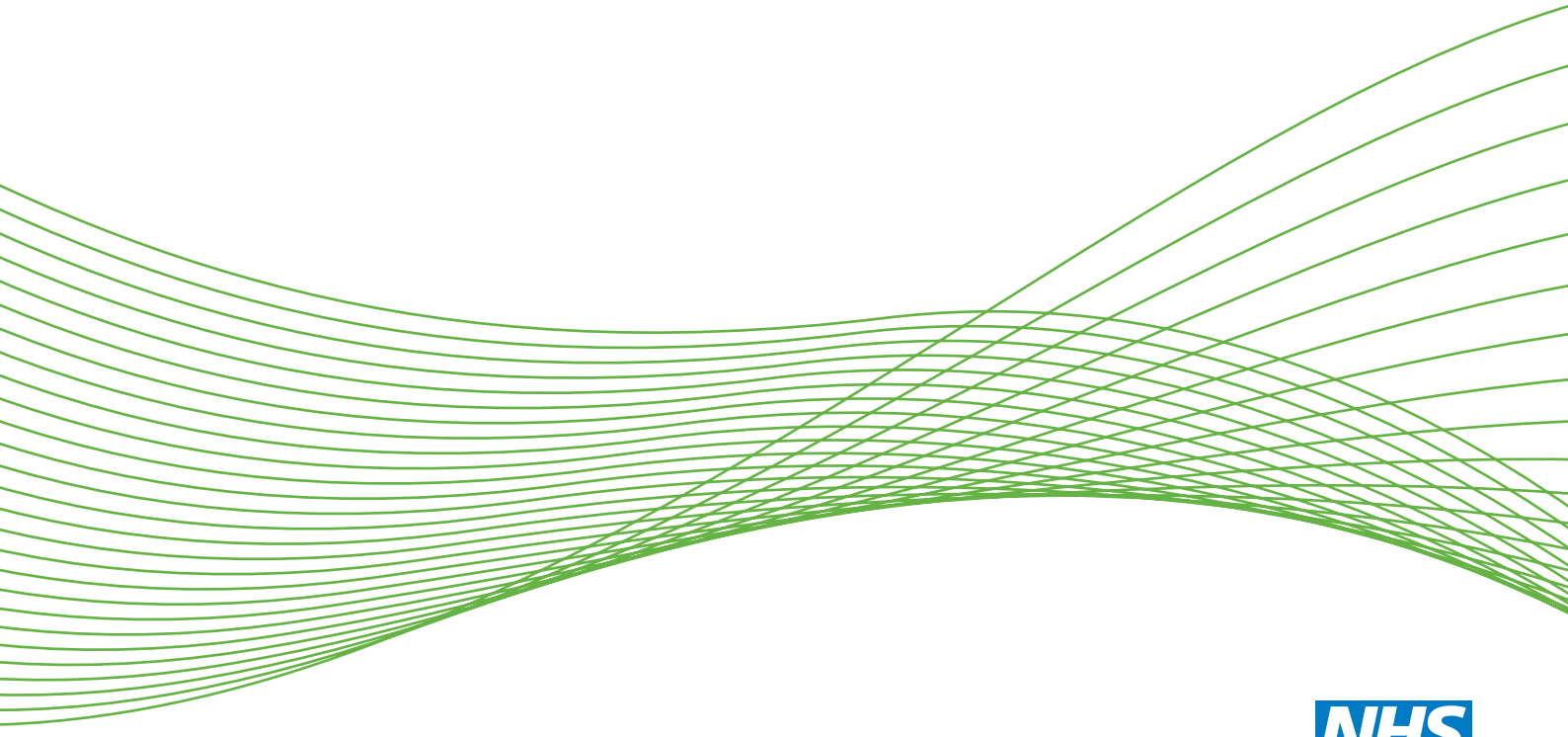
EFFICACY AND MECHANISM EVALUATION

VOLUME 2 ISSUE 4 JUNE 2015
ISSN 2050-4365

Randomised controlled trial of Antiglucocorticoid augmentation (metyrapone) of antiDepressants in Depression (ADD Study)

I Nicol Ferrier, Ian M Anderson, Jane Barnes, Peter Gallagher, Heinz CR Grunze, Peter M Haddad, Allan O House, Tom Hughes, Adrian J Lloyd, Chrysovalanto Mamasoula, Elaine McColl, Simon Pearce, Najma Siddiqi, Baxi Sinha, Chris Speed, Nick Steen, June Wainwright, Stuart Watson, Fiona H Winter, R Hamish McAllister-Williams and the ADD Study Team



NHS
National Institute for Health Research

DOI 10.3310/eme02040

Randomised controlled trial of Antiglucocorticoid augmentation (metyrapone) of antiDepressants in Depression (ADD Study)

I Nicol Ferrier,[1,2] Ian M Anderson,[3] Jane Barnes,[4] Peter Gallagher,[1] Heinz CR Grunze,[1,2] Peter M Haddad,[3] Allan O House,[5] Tom Hughes,[6] Adrian J Lloyd,[1,2] Chrysovalanto Mamasoula,[4] Elaine McColl,[4] Simon Pearce,[7] Najma Siddiqi,[8] Baxi Sinha,[9] Chris Speed,[4] Nick Steen,[4] June Wainwright,[9] Stuart Watson,[1,2] Fiona H Winter,[9] R Hamish McAllister-Williams[1,2]* and the ADD Study Team

[1]Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, UK
[2]Northumberland Tyne and Wear NHS Foundation Trust, Newcastle upon Tyne, UK
[3]Greater Manchester West Mental Health NHS Foundation Trust and Neuroscience and Psychiatry Unit, Manchester University, Manchester, UK
[4]Newcastle Clinical Trials Unit, Institute of Health and Society, Newcastle University, Newcastle upon Tyne, UK
[5]Institute of Health Sciences, University of Leeds, Leeds, UK
[6]Leeds and York Partnership NHS Foundation Trust, Leeds, UK
[7]Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK
[8]Bradford District NHS Care Trust, Saltaire, West Yorkshire, UK
[9]Tees, Esk and Wear Valleys NHS Foundation Trust, Darlington, County Durham, UK

*Corresponding author

Declared competing interests of authors: Dr McA-W reports grants from the Northumberland, Tyne and Wear Comprehensive Local Research Network, and Northumberland, Tyne and Wear NHS Foundation Trust (Research Capacity Funding) during the conduct of the study. INF reports grants from the Northumberland, Tyne and Wear NHS Foundation Trust (Research Capacity Funding) during the conduct of the study. EMcC is a member of the National Institute for Health Research Journal Editorial Group. HG reports personal fees from Bristol-Myers Squibb (BMS), Desitin, Lundbeck, Hoffmann-La Roche, Servier, Sanofi-aventis, and Richter Gedeon, outside the submitted work. IA reports personal fees from Lundbeck and Alberta Psychiatric Association, and royalties from Springer, outside the submitted work. He also reports fees paid to his institution from Servier, Alkermes, AstraZeneca and Medicine Publishing (Elsevier) during the conduct of the study. PMH reports personal fees and non-financial support from Janssen, Eli Lilly and Company, Otsuka and Servier, and personal fees from Sunovion, Lundbeck, BMS and AstraZeneca, outside the submitted work. SW reports grants from the Efficacy and Mechanism Evaluation board of the Medical Research Council during the conduct of the study.

Published June 2015
DOI: 10.3310/eme02040

This report should be referenced as follows:

Ferrier IN, Anderson IM, Barnes J, Gallagher P, Grunze HCR, Haddad PM, *et al*. Randomised controlled trial of Antiglucocorticoid augmentation (metyrapone) of antiDepressants in Depression (ADD Study). *Efficacy Mech Eval* 2015;**2**(4).

Efficacy and Mechanism Evaluation

ISSN 2050-4365 (Print)

ISSN 2050-4373 (Online)

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

Editorial contact: nihredit@southampton.ac.uk

The full EME archive is freely available to view online at www.journalslibrary.nihr.ac.uk/eme. 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 *Efficacy and Mechanism Evaluation* journal
Reports are published in *Efficacy and Mechanism Evaluation* (EME) if (1) they have resulted from work for the EME programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

EME programme

The Efficacy and Mechanism Evaluation (EME) programme was set up in 2008 as part of the National Institute for Health Research (NIHR) and the Medical Research Council (MRC) coordinated strategy for clinical trials. The EME programme is broadly aimed at supporting 'science driven' studies with an expectation of substantial health gain and aims to support excellent clinical science with an ultimate view to improving health or patient care.

Its remit includes evaluations of new treatments, including therapeutics (small molecule and biologic), psychological interventions, public health, diagnostics and medical devices. Treatments or interventions intended to prevent disease are also included.

The EME programme supports laboratory based or similar studies that are embedded within the main study if relevant to the remit of the EME programme. Studies that use validated surrogate markers as indicators of health outcome are also considered.

For more information about the EME programme please visit the website: http://www.nets.nihr.ac.uk/programmes/eme

This report

The research reported in this issue of the journal was funded by the EME programme as project number 08/43/39. The contractual start date was in May 2010. The final report began editorial review in June 2014 and was accepted for publication in January 2015. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The EME editors and production house have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the final report document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research. 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, MRC, NETSCC, the EME programme or the Department of Health. 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 EME programme or the Department of Health.

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

Abstract

Randomised controlled trial of Antiglucocorticoid augmentation (metyrapone) of antiDepressants in Depression (ADD Study)

I Nicol Ferrier,[1,2] Ian M Anderson,[3] Jane Barnes,[4] Peter Gallagher,[1] Heinz CR Grunze,[1,2] Peter M Haddad,[3] Allan O House,[5] Tom Hughes,[6] Adrian J Lloyd,[1,2] Chrysovalanto Mamasoula,[4] Elaine McColl,[4] Simon Pearce,[7] Najma Siddiqi,[8] Baxi Sinha,[9] Chris Speed,[4] Nick Steen,[4] June Wainwright,[9] Stuart Watson,[1,2] Fiona H Winter,[9] R Hamish McAllister-Williams[1,2]* and the ADD Study Team

[1]Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, UK
[2]Northumberland Tyne and Wear NHS Foundation Trust, Newcastle upon Tyne, UK
[3]Greater Manchester West Mental Health NHS Foundation Trust and Neuroscience and Psychiatry Unit, Manchester University, Manchester, UK
[4]Newcastle Clinical Trials Unit, Institute of Health and Society, Newcastle University, Newcastle upon Tyne, UK
[5]Institute of Health Sciences, University of Leeds, Leeds, UK
[6]Leeds and York Partnership NHS Foundation Trust, Leeds, UK
[7]Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK
[8]Bradford District NHS Care Trust, Saltaire, West Yorkshire, UK
[9]Tees, Esk and Wear Valleys NHS Foundation Trust, Darlington, County Durham, UK

*Corresponding author r.h.mcallister-williams@ncl.ac.uk

Background: Depressed patients who do not respond to second-line antidepressant drugs are characterised as suffering from treatment-refractory depression (TRD). Chronic psychosocial stress hypothalamic–pituitary–adrenal (HPA) axis dysfunction is associated with attenuated responses to antidepressants. Corticosteroid co-administration reduces the increase in forebrain 5-hydroxytryptamine with selective serotonin reuptake inhibitors, whereas antiglucocorticoids have the opposite effect. A Cochrane review suggesting that antiglucocorticoid augmentation of antidepressants may be effective in treating TRD included a pilot study of the cortisol synthesis inhibitor, metyrapone. The **A**ntiglucocorticoid augmentation of anti**D**epressants in **D**epression (ADD Study) was a multicentre randomised placebo-controlled trial of metyrapone augmentation of serotonergic antidepressants in patients with TRD.

Objective: To determine the efficacy and safety of augmentation of standard serotonergic antidepressants with metyrapone 500 mg twice a day for 3 weeks in patients with TRD.

Methods: A total of 165 patients with moderate to severe TRD aged 18–65 years were randomised to metyrapone 500 mg twice daily or placebo for 3 weeks, in addition to ongoing serotonergic antidepressants. The primary outcome was improvement in Montgomery–Åsberg Depression Rating Scale (MADRS) score 5 weeks after randomisation estimated using analysis of covariance. Secondary outcomes included the degree of persistence of treatment effect for up to 6 months, and also safety and tolerability of metyrapone. ADD included substudies investigating the potential mechanism of action of metyrapone, and utilised a comparator group of healthy participants.

Results: The estimated mean difference for each of our study outcomes between randomised groups, 5 weeks post randomisation (allowing for variation between centres and whether or not patients originate from primary or secondary care) was MADRS –0.51 [95% confidence interval (CI) –3.48 to 2.46]; Beck Depression Inventory (BDI) –2.65 (95% CI –6.41 to 1.10); Clinical Anxiety Scale 0.46 (95% CI –1.20 to 2.12); State–Trait Anxiety Inventory 1.2 (95% CI –0.6 to 3.0); European Quality of Life-5 Dimensions 0.015 (95% CI –0.069 to 0.099); EuroQol visual analogue scale 5.6 (95% CI –0.7 to 12.0); and Young Mania Rating Scale –0.04 (95% CI –0.52 to 0.45). The differences were not statistically significant and were small in relation to the change from baseline in both groups that was observed immediately after completion of therapy. Endocrinological data required for compliance assessment are not yet available. HPA function, similar in patients and control subjects, was not associated with differing clinical responses. Neuropsychological impairments were found, along with changes in brain structure and function, but no effect of metyrapone was seen on these measures.

Discussion: The inclusion criteria led to the sample being broadly representative of patients with TRD, within the UK NHS, with high anxiety and BDI scores. Metyrapone augmentation of antidepressants is not efficacious for outpatients with TRD who are moderately depressed. There was no obvious benefit associated with the use of metyrapone, either on the primary outcome or over the period of follow-up, and this negative result extended to other secondary outcomes. Metyrapone was well tolerated. There were no serious adverse events attributable to it and adverse events were as common with the placebo. HPA axis function was not associated with differing clinical or neuropsychological outcomes.

Conclusions: The results of the study suggest that although metyrapone augmentation was well tolerated, it is ineffective in the treatment of refractory depression. This finding is contrary to a previous proof of principle study in more acutely unwell patients. Future research should consider whether or not antiglucocorticoid treatments, such as metyrapone, should be targeted at patients with confirmed hypercortisolaemia.

Trial registration: Current Controlled Trials ISRCTN45338259.

Funding details: This study was funded by the National Institute for Health Research Efficacy and Mechanism Evaluation (EME) programme, a MRC and NIHR partnership.

Contents

List of tables

List of figures

List of abbreviations

5-HT	5-hydroxytryptophan, serotonin
ACTH	adrenocorticotrophic hormone
AE	adverse event
ANCOVA	analysis of covariance
ANOVA	analysis of variance
ANT	Attentional Network Test
AUCg	area under the curve ground
AUCi	area under the curve increase
BA	Brodmann area
BDI	Beck Depression Inventory
BOLD	blood oxygen level dependent
BSE	between-search error
CANTAB	CAmbridge Neuropsychological Test Automated Battery
CAR	cortisol awakening response
CAS	Clinical Anxiety Scale
CBT	cognitive behavioural therapy
CI	confidence interval
CLRN	Comprehensive Local Research Network
COM	combined memory condition
CTQ	Childhood Trauma Questionnaire
df	degree of freedom
dlPFC	dorsolateral prefrontal cortex
DMEC	Data Monitoring and Ethics Committee
DSM-IV	*Diagnostic and Statistical Manual of Mental Disorders*-Fourth Edition
EEG	electroencephalography
EEM	emotional enhancement of memory
EL	emotional labelling
EM	emotional matching
EME	Efficacy and Mechanism Evaluation
EPI	echo planar imaging
EQ-5D	European Quality of Life-5 Dimensions
EQ-VAS	EuroQol visual analogue scale
ERP	event-related potential
ESMT	emotional source memory task
FEP	facial emotion processing
fMRI	functional magnetic resonance imaging
FWE	family-wise error
GP	general practitioner
GR	glucocorticoid receptor
GRID-HAMD	GRID Hamilton Depression Rating Scale
HDRS17	Hamilton Depression Rating Scale-17 item
HEOG	horizontal electrooculogram
HPA	hypothalamic–pituitary–adrenal
HV	healthy volunteer
IAPS	International Affective Picture Set
LDAEP	loudness dependency of auditory evoked potentials
LTP	long-term potentiation
MADRS	Montgomery–Åsberg Depression Rating Scale
MGH-TRD	Massachusetts General Hospital Treatment Resistant Depression
MHRA	Medicines and Healthcare products Regulatory Agency
MHRN	Mental Health Research Network
MNI	Montreal Neurological Institute
MR	mineralocorticoid receptor
MRC	Medical Research Council
MRI	magnetic resonance imaging
NART	National Adult Reading Test

NIHR	National Institute for Health Research
OLB	object-location binding
phMRI	pharmacological functional magnetic resonance imaging
PI	principal investigator
POM	position-only memory
QoL	quality of life
rACC	rostral anterior cingulate cortex
RCT	randomised controlled trial
ROI	region of interest
RR	relative risk
RT	reaction time
SAE	serious adverse event
SAP	statistical analysis plan
SCID	Structured Clinical Interview for DSM
SD	standard deviation
SEM	standard error of the mean
SM	shape matching
SPM8	Statistical Parametric Mapping
SSRI	selective serotonin reuptake inhibitor
STAI	State–Trait Anxiety Inventory
TE	echo time
TR	repetition time
TRD	treatment-refractory depression
TSES	Toronto Side Effects Scale
U&E	urea and electrolytes
VEOG	vertical electrooculogram
VEP	visual evoked potential
WSE	within-search error
YMRS	Young Mania Rating Scale

Plain English summary

Depression is common and can have a terrible impact on patients and their families. Guidelines recommend talking therapies for patients with milder depression and adding antidepressants for moderate and severe depression. However, antidepressants work in only about 70% of sufferers. Why this happens is not known. One possibility is that increases in the stress hormone cortisol may reduce the effectiveness of antidepressants. Cortisol rises when people are stressed and may stay high in depression. Raised cortisol reduces the effect of antidepressants on the chemicals in the brain, which are thought to be important in how they work. Small studies have suggested that reducing the level of cortisol produces better outcomes for depression. We have tested the drug metyrapone, which blocks the production of cortisol, in a group of 165 people who remained depressed after previous treatment with at least two antidepressants. Metyrapone was given for 3 weeks. We investigated whether or not it led to benefits over the next 6 months compared with a group of similar people who received dummy tablets. Both groups of depressed people improved but there was no difference between them. There was also no difference in anxiety scores or quality of life with metyrapone treatment. Changes in memory, processing emotions and brain function were found in the depressed patients but no effects of metyrapone were found. It appears that this particular drug does not produce benefits for such patients. Further research is needed to find a treatment that reduces the effects of cortisol and improves the outcome for depression.

Scientific summary

Background

Many patients (approximately 30–50%) with depression do not respond to first-line antidepressant drugs. Responses to a second antidepressant are also disappointingly low (approximately 30%). Such non-responding patients are characterised as suffering from treatment-refractory depression (TRD). Chronic psychosocial stress and dysfunction of the hypothalamic–pituitary–adrenal (HPA) axis are both common in depression, and are associated with an attenuated clinical response to antidepressants. In preclinical studies, co-administration of corticosteroids leads to a reduction in the ability of selective serotonin reuptake inhibitors (SSRIs) to increase forebrain 5-hydroxytryptamine, whereas co-administration of antiglucocorticoids has the opposite effect. A Cochrane review suggested efficacy of antiglucocorticoid augmentation of antidepressants in patients with depression, with the largest effect size seen with metyrapone, a cortisol synthesis inhibitor that crosses the blood–brain barrier (Gallagher P, Malik N, Newham J, Young AH, Ferrier IN, Mackin P. Antiglucocorticoid treatments for mood disorders. *Cochrane Database Syst Rev* 2008;**1**:CD005168). A positive double-blind, placebo-controlled randomised trial of metyrapone was conducted in a centre in Germany, with 63 depressed inpatients.

Objectives

The Antiglucocorticoid augmentation of antiDepressants in Depression (ADD Study) was a multicentre, patient-randomised, double-blind placebo-controlled trial of metyrapone augmentation of serotonergic antidepressants in patients with TRD in the UK NHS. The primary objective was to determine whether or not metyrapone (500 mg twice a day) for 21 days is efficacious in augmenting ongoing treatment with conventional serotonergic antidepressants in TRD. The primary outcome by which this objective was assessed was the Montgomery–Åsberg Depression Rating Scale (MADRS) scored at baseline and 2 weeks post treatment (week +5 from randomisation) in a representative sample of depressed patients who had failed to respond to at least two courses of antidepressants, drawn from primary care and psychiatric outpatient clinics in the UK. Treatment with metyrapone was compared with treatment with placebo, using analysis of covariance. Secondary clinical objectives were to (1) determine the clinical effect size at 2 weeks post completion of treatment (5 weeks post randomisation) of a 3-week course of metyrapone (vs. placebo) augmentation of antidepressants; (2) assess whether or not any observed response was sustained for up to 21 weeks post cessation of metyrapone; (3) assess whether or not metyrapone augmentation improves patients' quality of life (QoL) using the self-completed EuroQol EQ-5D instrument (European Quality of Life-5 Dimensions); (4) assess the tolerability and safety of metyrapone augmentation in this study population; and (5) assess the mechanism of action of metyrapone in mechanistic substudies using neuropsychological, neuroendocrine and neuroimaging outcomes.

Assessment of additional secondary outcome measures of symptomatology [Clinical Anxiety Scale (CAS), Beck Depression Inventory (BDI), State–Trait Anxiety Inventory (STAI) and Young Mania Rating Scale (YMRS)] were conducted at the same time points as described for the MADRS.

Methods

A total of 165 patients with moderate to severe TRD, aged 18–65 years, were randomised to metyrapone 500 mg twice daily or placebo for 3 weeks, in addition to ongoing treatment with serotonergic antidepressants. Treatment occurred between weeks 0 (randomisation) and week +3 relative to randomisation. Patients were assessed on the above outcomes at –2, 0, +3, +5, +8, +16 and +24 weeks. Inclusion criteria were that the patient had (1) a major depressive episode assessed using the Structured Clinical Interview for DSM (SCID) (*Diagnostic and Statistical Manual of Mental Disorders*-Fourth Edition, DSM-IV); (2) a Hamilton Depression Rating Scale-17 item (HDRS17) score of ≥ 18 at weeks –2 and 0; (3) a Massachusetts General Hospital Treatment Resistant Depression (MGH-TRD) staging score of 2–10 at week –2 as a measure of treatment refractoriness; and (4) current treatment with a single agent or combination antidepressant treatment [which included a serotonergic drug (a SSRI), a tertiary amine tricyclic, venlafaxine, duloxetine or mirtazapine]. At the point of randomisation, patients were required to have been on their current antidepressant medication, at the current dose, for a minimum of 4 weeks, and this medication needed to be continued unchanged during the trial. Exclusion criteria included any other DSM-IV axis 1 diagnosis other than an anxiety disorder; a physical comorbidity that would make metyrapone inappropriate; pregnancy or breastfeeding; use of medication that would interfere with metyrapone; and dependence on alcohol or other drug(s) in the past 12 months and/or current harmful use of such substances. Metyrapone treatment potentially engenders hypocortisolaemia, with manifestations including a risk of postural hypotension, hyperkalaemia and hyponatraemia. Therefore, safety assessments included serum cortisol measures at week +1, as well as measuring sitting and standing blood pressure and urea and electrolytes at weeks +1 and +5. Serious adverse events (SAEs) and adverse events (AEs) were routinely enquired for and recorded. Tolerability was further assessed using the Toronto Side Effects Scale (TSES). Metyrapone administration has previously been shown to cause an increase in levels of 11-deoxycortisol, and the increase in 11-deoxycortisol between weeks –2 and +1 was to be used, when available, as a measure of adherence to medication, as this has been shown to be highly sensitive to treatment with metyrapone. The study also investigated a number of mechanistic objectives, including whether or not the patients had evidence of baseline hypercortisolaemia and, if present, whether or not this had any impact on clinical and neuropsychological outcomes. A comparator group was also recruited to support these mechanistic investigations.

Results

Overall, 877 patients were referred to the study team: 237 from primary care, 320 from secondary care and 310 as self-referrals following media exposure of the study or seeing posters. The origin of 10 patients was unclear. A total of 284 underwent detailed screening for eligibility. The remainder were either deemed to be ineligible on the basis of a brief telephone screen or did not follow up contact, and 173 were deemed to be eligible. Of the 111 who did not meet inclusion criteria, 10 did not meet the criteria for a major depressive episode using the SCID, 52 had HDRS17 item scores of <18, 17 had axis 1 disorders other than anxiety, nine were on an inappropriate antidepressant, three had MGH-TRD staging scores outside the range of 2–10, 18 had physical disorders that excluded them, and five had other miscellaneous exclusion criteria (three patients were excluded for more than one reason).

Eight patients subsequently dropped out before randomisation (i.e. between weeks –2 and 0) and so 165 patients were randomised (82 to placebo and 83 to metyrapone). Of these, 143 (86.7%) completed the primary outcome at +5 weeks (74 on placebo and 69 on metyrapone). A further 39 dropped out between week +5 and week +24, so that 104 (63%) completed the study (58 on placebo and 46 on metyrapone). The groups were well balanced at randomisation in terms of demographics and key clinical variables. The mean MADRS score for the groups indicated moderate to severe depression, with MGH scores well in the range of treatment resistance. The group showed evidence of high Beck Depression Inventory (BDI) scores. There was evidence of high scores on measures of anxiety, and comorbid anxiety conditions were frequent.

The estimated mean difference for each of our study outcomes between randomised groups 5 weeks post randomisation (adjusting for variation between centres, whether or not patients originate from primary or secondary care and baseline score) was MADRS (the primary outcome measure) –0.51 (95% CI –3.48 to 2.46); BDI –2.65 (95% CI –6.41 to 1.10); Clinical Anxiety Scale (CAS) 0.46 (95% CI –1.20 to 2.12); STAI 1.2 (95% CI –0.6 to 3.0); European Quality of Life-5 Dimensions 0.015 (95% CI –0.069 to 0.099); EuroQol visual analogue scale 5.6 (95% CI –0.7 to 12.0); Young Mania Rating Scale –0.04 (95% CI –0.52 to 0.45). The differences were not statistically significant for any measure and tended to be small in relation to the change in both groups observed at week +5. Response rates were low and almost identical in both groups (21.6% in the placebo group and 20.3% in the metyrapone group). Remission rates were similarly low and almost identical in the two groups at this time point.

Endocrinological data required for compliance assessment are not yet available. HPA axis function, similar in patients and control subjects, was not associated with differing clinical responses. A wide range of neuropsychological impairments were found, along with changes in brain structure and function, but no differential effect of metyrapone was seen on these measures.

Metyrapone was generally well tolerated. There were 14 SAEs reported for 11 of the patients randomised to study medication (five in the group randomised to metyrapone and six in the group randomised to placebo) but none was attributed to metyrapone and most occurred well after the period of active drug treatment. Non-serious AEs were more common and broadly similar in type and frequency between the two groups, but we cannot exclude the possibility that treating patients with metyrapone may increase the risk of AEs. Of particular note, however, is that metyrapone did not increase the risk of suicidal ideation or suicide attempts. Scores on the TSES were broadly similar between the two groups, with no evidence of a difference in the frequency of postural hypotension or dizziness.

Conclusions

The broad inclusion criteria led to the sample being broadly representative of patients with TRD who are treated within the NHS. The sample had high anxiety and BDI scores and frequent comorbid anxiety. No evidence was found that metyrapone augmentation of serotonergic antidepressants is efficacious for patients with moderate to severe depression – managed in NHS secondary care outpatient clinics or by general practitioners in primary care – who have failed to respond to at least two antidepressants. There was no obvious benefit to its use either on the primary outcome or over the period of follow-up, and this negative result extended to other secondary outcomes, such as the CAS, BDI and quality-of-life measures. Metyrapone was well tolerated by this group and there were no serious AEs attributable to it. AEs were as common in patients treated with placebo. Clinical outcomes have not yet been analysed with respect to the measure of adherence utilised.

A wide range of neuropsychological impairments were found along with changes in brain structure and function, but no differential effect of metyrapone was seen on these measures. This population with TRD was characterised by increased exposure to childhood adversity (compared with the control subjects) and normal HPA axis function. These findings accord with the existing literature in chronic populations; the former predicts non-response to treatment. However, baseline HPA axis function, change in cortisol awakening response in response to drug treatment or severity of childhood trauma did not predict clinical response to metyrapone.

There are very few data specifically on the neuropsychology of treatment-resistant depressed groups. Those studies that have been conducted suggest that deficits are restricted to tests of processing speed. In this TRD sample we see broad deficits in verbal and visuospatial working memory and emotion processing compared with healthy control subjects. Deficits in attention were not general and, instead, were restricted to the executive control of attention. These findings are indicative of an impairment in effortful processing in TRD. No differential effects of metyrapone were seen on these measures.

Trial registration

This study is registered as ISRCTN45338259.

Funding

Funding for this study was provided by the Efficacy and Mechanism Evaluation programme of the National Institute for Health Research.

Chapter 1 Background and objectives

Scientific background

Depression is a common disorder, affecting some 10% of the population.[1] It is rated by the Disease Control Priorities Project as one of the leading medical contributors to the global burden of disease.[2] It can become long-lasting and may recur frequently. Depression has a large negative impact on the quality of life (QoL) of service users and their carers. It is associated with high morbidity (depression has been identified as one of the leading causes of work days lost and working-age adults receiving disability payments in the UK)[3] and mortality, through suicide and increased deaths from cardiovascular disease.[4] Clinical guidelines recommend the use of antidepressant medication for the treatment of a moderate to severe depressive episode.[1,5] Antidepressant drugs have established efficacy compared with placebo in clinical trials; however, in naturalistic settings many patients have unsatisfactory outcomes. The large pragmatic STAR*D study conducted in the USA showed that, even with protocol-driven treatment, first-line therapy with a selective serotonin reuptake inhibitor (SSRI: citalopram) in over 2500 patients was associated with remission in only 28% and response (defined as a 50% decrease in symptom scores) in < 50% of patients.[6] Further, second-line treatment in the STAR*D study with an alternative antidepressant was associated with an even lower remission rate of 20–25%.[7] Much of the burden of depression is consequent on this treatment-refractory depression (TRD).

Clinical response to antidepressant treatment may be influenced by a number of factors but an inadequate neurochemical response is one likely mechanism of non-response. The majority of antidepressants, on long-term administration, elevate forebrain levels of 5-hydroxytryptamine (serotonin; 5-HT). The clinical importance of this is supported by the fact that acute depletion of tryptophan (the precursor of 5-HT) can lead to the rapid return of depressive symptoms in patients treated with antidepressants.[8,9] There is a long-held notion that life events and lack of social support predict worse treatment outcomes in patients with depression.[10] The mechanism of this may relate to the hypothalamic–pituitary–adrenal (HPA) axis. It is well established from animal and human studies that glucocorticoids influence multiple aspects of 5-HT neurotransmission (including the sensitivity of 5-HT_{1A} autoreceptors and postsynaptic 5-HT receptor efficacy,[11–14] which are postulated to be central to the antidepressant mechanism of action).[15] Furthermore, dysregulation of the HPA axis can reduce the effects of antidepressants in the frontal cortex. Implantation of corticosteroid-releasing pellets in rodents to induce an HPA axis dysregulation has been shown to attenuate the ability of antidepressants to elevate forebrain 5-HT levels.[16] Interestingly, there is also increasing evidence that antiglucocorticoid treatments have the opposite effect by enhancing the forebrain 5-HT response to antidepressants.[17] This suggests that reducing normal physiological levels of glucocorticoid receptor activation can increase the elevation of 5-HT in response to SSRIs. These data suggest a potential mechanism by which antiglucocorticoid strategies can enhance the clinical effectiveness of antidepressants in clinical practice.

Hypothalamic–pituitary–adrenal axis abnormalities are often demonstrated in patients with mood disorders. HPA axis dysregulation in depression is often characterised – particularly in those with melancholic symptoms – by a flattened cortisol diurnal rhythm with elevated trough levels of cortisol[18] and by attenuated negative feedback effects of corticosteroids on adrenocorticotrophic hormone (ACTH) and cortisol release.[19] There is increasing evidence that such dysregulation is associated with poor prognosis, including non-response to antidepressants and future relapse.[20–26] Further, a Cochrane review suggested efficacy of antiglucocorticoid augmentation of antidepressants in patients with depression with the largest effect size seen with metyrapone,[27] a cortisol synthesis inhibitor that crosses the blood–brain barrier. There have been several open – including randomised – studies of metyrapone augmentation of antidepressants, showing efficacy in TRD.[28–32] In addition, a successful proof-of-concept double-blind, placebo-controlled randomised study has been conducted by Jahn *et al.*[33] in a centre in Germany, with 63 depressed inpatients. Patients were all

being treated with a serotonergic antidepressant and were randomised to add-on treatment with metyrapone (1 g/day for 3 weeks) or placebo. The primary outcome measure was the percentage of responders [defined by an improvement of ≥ 40% on the Hamilton Depression Rating Scale-17 item (HDRS17)] 5 weeks after randomisation (i.e. 2 weeks after cessation of metyrapone or placebo augmentation). Patients receiving metyrapone were significantly more likely to respond than those receiving placebo, with an effect size of Cohen's $d = 0.63$. Kaplan–Meier analysis revealed a faster response with metyrapone, which was well tolerated and without serious side effects.

Previous research has shown, in patients with depression, hyperactivation of the amygdala and other limbic structures and hypoactivation in prefrontal neocortical regions, specifically in response to negative emotional stimuli,[34] although such findings have been inconsistent and may, for instance, not be present in patients with mild to moderate depression.[35] They have been shown to be altered by levels of serotonin.[36] Neurocognitive deficits and mood-congruent biases have been reported in depression, although there are inconsistencies which may be related to the population studied. We investigated these and the neural basis of working and episodic memory, including the encoding and retrieval of emotional memories in patients with TRD on the basis that this population may be more homogeneous and likely to display neurocognitive deficits than an undifferentiated depressed group.

Previous electroencephalographic studies have entailed two main elements: (1) prediction of treatment response and (2) investigation of the neural correlates of emotional processing and memory.

1. *Treatment response prediction* A number of electroencephalography (EEG) variables recorded prior to the commencement of antidepressant treatment have been shown to predict response in patients with major depressive disorder.[37] For example, studies of EEG frequency spectrum have generally found that responders have greater alpha power prior to treatment,[38–42] although there have been negative findings.[43] In addition to absolute alpha power, there have also been findings suggesting that antidepressant non-responders show alpha hemispheric asymmetry with lower power in the right hemisphere compared with the left,[42,44,45] with a claim that this asymmetry has a sensitivity of 64% and specificity of 71% in identifying responders.[42] An alternative power spectrum band that has been explored extensively is the theta band. This work suggests that lower frontal theta power is associated with greater response to antidepressants,[40,41,46] with a sensitivity and specificity of 64% and 62%, respectively.[46] In line with hypotheses that activity/metabolic rate in the rostral (pregenual) anterior cingulate cortex (rACC) predicts antidepressant response,[47–51] there are findings from three independent research groups that EEG rACC theta activity, which closely correlates with rACC positron emission tomography glucose metabolism,[52] predicts response[53–55] with an effect size of 1.3.[54] Aside from EEG frequency power in various bands, there has also been considerable research into the ability of the loudness dependency of auditory evoked potentials (LDAEPs) to predict response. LDAEPs have been argued to reflect serotonergic function in humans,[56] although the greatest promise for this measure is as a potential biomarker of antidepressant response.[57] The LDAEP's ability to predict antidepressant response has been replicated by more groups than for any other EEG method described above, and includes more patients than any other technique.[37,58–65] Critically, the LDAEP is also reported to be able to predict response to serotonergic and noradrenergic antidepressants differentially.[57,66] In the light of these findings, an obvious question is whether or not the ability to predict antidepressant response is improved by combining different EEG parameters. Bruder *et al.*[42] combined global alpha power (sensitivity 73% and specificity 58%) and alpha asymmetry (sensitivity 64% and specificity 71%) and found an overall sensitivity of 83% and specificity of 68%. An additional study[66] combined theta activity localised to rACC and the LDAEP technique, showing that this had the ability to discriminate between responders and non-responders with an effect size of 1.4.

2. *Neural correlates of emotional processing and memory* An aim of the EEG mechanistic studies is to examine the effects of metyrapone treatment on emotional context on memory retrieval. In studying emotional memory retrieval, presentation of emotional items in the recognition test phase of a memory task is a potential confounder. This may lead to the activation of brain regions engaged in the emotional perception and arousal that occur during the presentation of emotional stimuli, rather than relating the effects of emotion on memory retrieval. A paradigm through which retrieval of a neutral, emotionally non-arousing, object is tested, following encoding of the objects in an emotionally manipulated context, can potentially eliminate this difficulty. The ADD Study utilised such a paradigm, based on the event-related potential (ERP) work of Smith *et al.*:[67] an emotional source memory task (ESMT). In this paradigm, recognition of a neutral, emotionally non-arousing, visual object is tested, following encoding of the objects in an emotional context [the object is presented superimposed on emotional pictures from the International Affective Picture Set (IAPS) or emotional faces]. This technique provides temporal information regarding neural activity, which would not be possible with functional magnetic resonance imaging (fMRI). It demonstrates that retrieval of emotional items, compared with neutral ones, shows increased activity of lateral temporal areas, bilaterally, early in the retrieval process and a later component over left temporofrontal regions.[67,68]

In addition, a putative ERP equivalent of long-term potentiation (LTP)[69] was also investigated. This involves presenting a high-frequency visual stimulus (a black and white chequerboard) to subjects, which has been shown to increase the amplitude of early visual ERPs[69] for 30–60 minutes. This task is of interest given that animal studies have demonstrated that LTP is highly sensitive to corticosteroids.[70] As a result, we hypothesised that changes in the cortisol awakening response (CAR) in patients following metyrapone treatment would correlate with the magnitude and duration of the ERP marker of LTP.

Objectives

The primary aim of the ADD Study was to examine the efficacy and safety of metyrapone augmentation of serotonergic antidepressants in a patient randomised placebo-controlled trial in patients with major depression who had not responded to at least two courses of antidepressants in their current episode. The study extends research in this area by exploring the translatability of the proof-of-concept study described by Jahn *et al.*[33] to an outpatient, primary and secondary care, UK NHS population. To date, all published studies of the use of antiglucocorticoids in patients with TRD have used short treatment periods of 1–3 weeks,[27] which can appear counterintuitive in such a potentially chronic condition. However, evidence suggests that the clinical effects of antiglucocorticoids on HPA axis function persist after their administration has ceased.[71,72] The persistence of effects on depressive symptoms and QoL was, therefore, examined in the ADD Study for 21 weeks after stopping metyrapone treatment compared with the 2-week follow-up period of Jahn *et al.*[33]

The exact mechanism by which metyrapone may enhance antidepressant efficacy is unknown. A secondary aim of the ADD Study was to explore the impact of metyrapone on HPA axis function, and the hypothesis that metyrapone leads to altered neural responsiveness to glucocorticoids with an increase in the frontocortical 5-HT response to antidepressants. In addition, a number of studies in subsamples, drawn from the main randomised controlled trial (RCT) population, were also undertaken using EEG and fMRI techniques, assessment of neuropsychological function, and genetic variability to address this mechanistic aim. The substudies are detailed in a paper[73] describing the protocol in detail, which also outlines the hypotheses tested and the rationale for the tests utilised. The protocol below is consistent with the SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) 2013 recommendations (see www.spirit-statement.org/spirit-statement/).

Primary clinical objective

The primary objective is to determine whether or not metyrapone (500 mg twice a day) for 21 days is efficacious in augmenting conventional serotonergic antidepressants in TRD in a NHS primary and secondary care setting. The primary outcome by which this objective is assessed is the Montgomery–Åsberg Depression Rating Scale (MADRS)[74] scored at baseline and 2 weeks post treatment (week +5 from randomisation), comparing patients randomised to metyrapone with those randomised to placebo.

Secondary objectives

Clinical objectives

- To determine the clinical effect size at 2 weeks post completion of treatment (5 weeks post randomisation) of a 3-week course of metyrapone (vs. placebo) augmentation of antidepressants in a representative sample of depressed patients who have failed to respond to at least two courses of antidepressants, drawn from primary care and psychiatric outpatient clinics in the UK.
- To assess whether or not any observed response is sustained for up to 21 weeks post cessation of metyrapone.
- To repeat the above analyses utilising the addition 'atypical' depression items (rating hypersomnia and increased appetite). In this scoring of the MADRS, the highest score from the conventional sleep and atypical sleep items, and conventional appetite and atypical appetite items, will be used to calculate the total MADRS score.
- To assess the persistence of change in MADRS score, the Clinical Anxiety Scale (CAS), Beck Depression Inventory (BDI) and State–Trait Anxiety Inventory (STAI) scores, using repeated measures analysis of variance (ANOVA), adjusting for randomisation strata of centre and primary compared with secondary care, and using all of the data points available.
- To assess whether or not metyrapone augmentation improves patients' QoL as assessed by the self-completed EuroQol European Quality of Life-5 Dimensions (EQ-5D) instrument (www.euroqol.org/).
- To assess the tolerability and safety of metyrapone augmentation in this study population.

Mechanistic objectives related to the full randomised controlled trial sample

- To assess whether or not metyrapone changes patients' HPA axis function.
- To assess whether or not changes in HPA axis function with metyrapone persist after stopping metyrapone.
- To assess whether or not the change in HPA axis function correlates with clinical response.
- To assess whether or not baseline HPA axis function predicts clinical response.
- To assess if type and severity of childhood trauma [as assessed by the Childhood Trauma Questionnaire (CTQ) scores] predicts clinical response.
- To determine the nature and extent of neuropsychological abnormalities in patients with TRD compared with healthy control subjects.
- To determine whether or not neuropsychological performance improves with metyrapone treatment.

Mechanistic objectives related to the Newcastle and Manchester subgroups of the full randomised controlled trial sample

- To compare visual cortical LTP in patients with TRD and healthy control subjects.
- To examine if visual cortical LTP is altered by treatment with metyrapone.
- To compare emotional source memory performance, and its electrophysiological underpinnings, in patients with TRD and healthy control subjects.
- To examine if emotional source memory performance, and its electrophysiological underpinnings, is altered by treatment with metyrapone.

- To determine whether or not EEG predictors of conventional antidepressant response, and in particular specific predictors of response to serotonergic antidepressants, predict response to metyrapone augmentation.
- To compare the degree of activation of the amygdala in response to emotional faces and words in patients with TRD and healthy control subjects.
- To compare neural activity during episodic and working memory tasks in patients with TRD and healthy control subjects.
- To determine whether or not patients with TRD have altered hippocampal response to hydrocortisone compared with healthy control subjects.
- To determine if the hippocampal response to hydrocortisone is modified in patients by metyrapone treatment.

Chapter 2 Methods

Trial design

Description of trial design

The ADD Study is a multicentre, two-arm (1 : 1 allocation), parallel-group, double-blind, patient-randomised, placebo-controlled superiority trial of augmentation of serotonergic antidepressants with metyrapone in patients with moderate to severe depression who have failed to respond to adequate trials of at least two antidepressants in their current episode. These patients were recruited from primary and secondary care settings.

All assessments were undertaken by trained research personnel under the supervision of the clinically trained principal investigators (PIs: authors INF, RHMW, SW, IMA, AOH, HCRG, PMH, TH, AJL). There was no planned interim analysis and no 'stopping rules' for the study as a whole. Individual patients were withdrawn from the study medication if it appeared that to continue would be deleterious to their mental health or safety. This was determined by the patient, the treating clinician and/or the research team, and was supported by the use of the MADRS, particularly if there was an increase in the score for question 10 'suicidal thoughts'. Withdrawal from medication did not necessitate withdrawal from the study.

There was patient and public involvement in all stages of the study and this was invaluable. Two (FW, JW) of the co-applicants and PIs were service users, one of whom was also a carer. They were very much involved in the design of the study. In particular, they were involved in the design and content of all information sheets and notices for general practitioner (GP) surgeries, etc. Both were involved in promoting recruitment in their respective locales. At a later stage in the study the decision to promote the study to the public through press releases was made and help in doing this well was obtained from the North East Mental Health Research Network (NE MHRN) Hub Service User and Carer Forum. The subsequent articles in the papers and the television appearance of the chief investigator acted as a catalyst, which the NE MHRN Hub Service User and Carer Forum then used to promote the importance of the study further. The response from the public was very marked and positive with frequent calls and referrals.

The study was registered on 21 December 2009 (ISRCTN45338259) under the public title 'Antiglucocorticoid augmentation of antiDepressants in Depression: the ADD Study'. Clinical trial authorisation was given by the Medicines and Healthcare products Regulatory Agency (MHRA: EudraCT: 2009-015165-31). Ethical approval was granted by the Sunderland Local Research Ethics Committee (REC reference number 10/H0904/9) on 22 April 2010.

Visits and assessments

These are outlined in *Table 1* and summarised below.

Screening visit (week –2)

Written informed consent was obtained and study eligibility determined. Height, weight and safety measurements were recorded. Baseline blood tests including urea and electrolytes (U&E), cortisol, thyroid function test, liver function tests, full blood count and β-human chorionic gonadotropin (if indicated) were taken. Family history (in first-degree relative) of mental illness was determined by clinical enquiry by a trained psychiatrist. Background factors, personality, and childhood adversity and life events were assessed between recruitment and randomisation, using the NewMood background questionnaire,[75] given to patients to complete and return at the next visit. This questionnaire includes the Big Five Inventory 44-item personality questionnaire,[76] a negative life events questionnaire adapted from the List of Threatening Experiences,[77] the Social Circumstances Questionnaire, which is an adaptation of the Social Support Questionnaire[78] as used in the NewMood study,[75] the CTQ[79] and the Ruminative Responses Scale.[80]

TABLE 1 Schedule of assessments

	Enrolment	Randomisation	Follow-up							
Time point	Week –2	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 8	Week 16	Week 24
Assessment of eligibility	✓									
Informed consent	✓									
Assessment of baseline characteristics – NewMood questionnaire[a]	✓									
Experimental intervention										
Assessment of depression severity – HDRS17	✓	✓								
Assessment of clinical symptoms – MADRS, CAS, BDI, STAI, YMRS		✓			✓		✓	✓	✓	✓
Assessment of QoL – EQ-5D		✓			✓		✓	✓	✓	✓
Assessment of side effects – TSES		✓			✓		✓	✓		
Assessment of side effects and AEs – self-report		✓	✓	✓	✓	✓	✓	✓	✓	✓
Suicide risk assessment	✓	✓	✓		✓		✓	✓	✓	✓
Pregnancy test if indicated	✓	✓	✓	✓	✓					
Assessment of concomitant medication	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Measurement of HPA axis function (CAR plus 11 p.m. saliva sample)		✓			✓		✓			
Physical observations[b]	✓[c]		✓				✓			
Medical history	✓									
Blood tests – U&E, cortisol	✓[d]		✓				✓			
Neuropsychological assessment[e]		✓					✓			
Neuroimaging[e]		✓					✓			
EEG[e]		✓					✓			

AE, adverse event; U&E, urea and electrolytes; YMRS, Young Mania Rating Scale.
a See *Screening visit* (below) for description of NewMood questionnaire details.
b Physical observations comprised sitting and standing pulse and blood pressure, and respiration rate.
c Screening physical observations also included height and weight.
d Screening blood tests also including thyroid function tests, liver function tests and full blood count.
e See neuropsychological assessment in *Chapter 2 (Trial design)*.

Depression severity was determined using the HDRS17, rated using the GRID Hamilton Depression Rating Scale (GRID-HAMD) for improved reliability,[81] and the MADRS.[74]

Randomisation visit (week 0)

Subjects were excluded if their HDRS17 score dropped to < 18 or if there was any change in their current antidepressant medication (drug or dose) between weeks –2 and 0. Otherwise study medication was supplied to commence the following day.

Follow-up

Data were collected at weeks +1, +2, +3 (end of active treatment period), +4, +5 (primary outcome time point), +8, +16 and +24 from the date medication was started (± 2 days). The week +2 and +4 visits could be completed by a telephone interview with the trained research assistant. Details of the assessments at each time point are described in *Chapter 3* (see *Table 2*). Depression severity was assessed using the MADRS, administered by trained members of the research team at randomisation (week 0) and weeks +3, +5, + 8, +16 and +24 relative to randomisation. The primary outcome measure was the MADRS score, recorded at baseline and 5 weeks post randomisation. The MADRS has preferable psychometric properties and higher sensitivity to change than other depression rating scales,[82] and was associated with the largest effect size in the Jahn study.[33] Additional secondary outcome measures of symptomatology [CAS,[83] BDI,[84] STAI[85] and Young Mania Rating Scale (YMRS)[86]] were conducted at the same time points as described for the MADRS. QoL was assessed using the self-completed EuroQol EQ-5D instrument (www.euroqol.org/)[87] and tolerability using the Toronto Side Effects Scale (TSES).[88]

Metyrapone treatment potentially engenders hypocortisolaemia with manifestations including a risk of postural hypotension, hyperkalaemia and hyponatraemia. Therefore, safety assessments included serum cortisol measures at week +1 and measurement of sitting and standing blood pressure and U&Es at weeks +1 and +5.

Hypothalamic–pituitary–adrenal axis assessment

Cortisol levels to determine the CAR[89] were obtained at the start of treatment (week 0) and then again at weeks +3 and +5 for all patients. This entailed participants collecting 5 ml of saliva, by passive drool,[90] into a plastic collecting tube on wakening and at 15-minute intervals for a further hour. A total of five samples were collected on each occasion, either the day before or day after the planned study visit. Participants were asked to collect a saliva sample by the same method for cortisol assay at 11 p.m. the night before each of the three CAR assessments (weeks 0, +3 and +5). In addition to collection of saliva samples, participants completed a brief questionnaire relating to the nature and quality of sleep the night before the CAR assessment.

In addition to the saliva samples, serum samples were taken at weeks –2, +1 and +5 for analysis of cortisol precursors and metabolites. Metyrapone administration has previously been shown to cause an increase in levels of ACTH and 11-deoxycortisol, together with an increase in the cortisone–cortisol ratio.[33,91] The increase in 11-deoxycortisol between weeks –2 and +1 was used as a measure of adherence to medication, as this has been shown to be highly sensitive to treatment with metyrapone.[33,91]

Neuropsychological assessment

Spatial working memory

This test from the CAmbridge Neuropsychological Test Automated Battery (CANTAB) is a self-ordered search task that places demands on spatial working memory and executive function. After the practice trials (two to three boxes), subjects must search through an increasing number of boxes (four, six or eight) for a hidden token. Once a token is found, it will not appear in the same box again. Subjects must continue the search without returning to a box that has already contained a token. Accuracy is measured as the number of *between-search errors* (BSEs – the number of times boxes that have already contained tokens on previous trials are searched) and *within-search errors* (WSEs – the number of times boxes that have already been examined on the current trial are searched).

Patients and control subjects [age- and sex-matched healthy volunteers (HVs) (see *Healthy volunteers*, below)] were compared using baseline data by an analysis of covariance (ANCOVA) with Huynh–Feldt correction: difficulty ('level' 4, 6 or 8) was a within-subject factor (for BSE and WSE), group (patient or control) was a between-subject factor; and age and sex were covariates. Uncorrected degrees of freedom (df) are presented in *Chapter 3* for clarity. The effect of metyrapone treatment in patients was assessed using multiple ANCOVA for each outcome with the post-treatment value as the dependent variable and pretreatment value as the covariate.

Attentional Network Test

The Attentional Network Test (ANT) was developed by Fan *et al.*[92] to fractionate and quantify the efficiency of the functional attentional networks of alerting, orienting and executive control proposed by Posner and Peterson.[93] The ANT is a combination of cued reaction time (RT) and a flanker task, and has been described in detail.[92] Briefly, this is presented in a number of conditions, differing by the preceding cue and the target flankers. The cue conditions are (1) *no cue*; (2) *centre cue* (a cue appears directly over the central fixation point); (3) *double cue* (a cue appears simultaneously above and below the fixation); or (4) *spatial cue* (a cue appears in either the upper or lower field, congruent with the location of the subsequent target stimulus). The arrow stimuli are (1) *neutral* (flanked by directionless lines); (2) *congruent* (flanked by arrows pointing the same direction as the target); or (3) *incongruent* (flanked by arrows pointing the opposite direction to the target). Participants must respond as quickly as possible indicating in which direction the arrow is pointing. The alerting effect is calculated by subtracting the mean RT of the double-cue conditions from the mean RT of the no-cue conditions. These conditions provide information that an event is about to occur, but not specifically where, therefore attention is diffused across the two potential target locations. The orienting effect is calculated by subtracting the mean RT of the spatial cue conditions from the mean RT of the centre cue. Both cues alert the participant that an event is about to occur, although only the spatial cue provides precise spatial information, allowing the participant to attend to the exact location. Finally, executive control of attention (conflict) is calculated by subtracting the mean RT of all congruent flanking conditions, summed across cue types, from the mean RT of incongruent flanking conditions.

Patients and control subjects were compared using baseline data by an ANCOVA with Huynh–Feldt correction: group was a between-subject factor; age and sex were covariates. Uncorrected degrees of freedom are presented for clarity. The effect of metyrapone treatment in patients was assessed using multiple ANCOVA for each outcome with the post-treatment value as the dependent variable and pretreatment value as the covariate.

Object-location memory

To assess memory for the locations of objects, the Object Relocation programme was used.[94] The programme presents stimulus displays on a PC fitted with a touchscreen monitor. A number of variations of the task parameters are possible within the programme. Here the programme was run using the immediate memory conditions from Kessels *et al.*[95]

First, subjects completed two control tasks that assessed *object identity memory* and *visuospatial construction and perception*. In the object identity task, subjects viewed 10 different objects for 30 seconds; these objects had to be remembered and subsequently recognised from a set of 20 objects, containing 10 of the ones shown previously and 10 distractors. In the visuospatial construction task, subjects had to copy a frame containing 10 different objects at different locations without a memory component. Each task condition consisted of an example containing only four objects/positions, followed by two different test displays.

Following these control tasks, subjects completed three experimental task conditions:

(a) *Position-only memory (POM)* Subjects viewed an array containing 10 identical objects and were required to remember their precise locations. After 30 seconds the array disappeared and the objects appeared along the top of the screen. Subjects were then required to move the objects down into the empty frame and recreate the exact positions of the array as accurately as possible.
(b) *Object-location binding (OLB)* Subjects viewed an array of 10 different objects and were required to remember where they were located within the frame. After 30 seconds the array disappeared and the objects appeared along the top of the screen. Subjects were then required to move the objects down into the frame and recreate the array, although the precise positions that had been occupied were indicated by premarked black dots.
(c) *Combined memory condition (COM)* This was identical to the OLB condition except for the relocation stage, at which there were no premarked black dots, that is subjects were required to remember and relocate the 10 different objects as precisely as possible to their exact previous locations.

Each task condition consisted of an example containing only four objects/positions, followed by two different test displays. Performance measures were percentage incorrect items in the object identity control condition and OLB conditions, and deviation error [millimetres (mm)] in the visuospatial construction and perception control condition, and POM and COM tasks. In the case of the POM task, as all objects are identical, it is impossible to specify to which location any given object is relocated and, consequently, the best-fit error is used.[95] All other tasks use the absolute error score.

Patients and control subjects were compared using baseline data by an ANCOVA with Huynh–Feldt correction; group was a between-subject factor; age and sex were covariates. Uncorrected degrees of freedom are presented for clarity. The effect of metyrapone treatment in patients was assessed using multiple ANCOVA for each outcome, with the post-treatment value as the dependent variable and pretreatment value as the covariate.

Digit span

To assess memory for verbal immediate/working memory, the forward and backward digit span measures were completed. The test was administered according to standardised instructions.[96] For the 'experimental' outcome measures, the total number of sequences correctly completed for forward, backward and total was calculated (i.e. 1 point for each correct sequence; maximum forward and reverse = 14). For the clinical span, the maximum sequence attained was recorded.

Patients and control subjects were compared using baseline data by an ANCOVA with Huynh–Feldt correction; group was a between-subject factor; age and sex were covariates. Uncorrected degrees of freedom are presented for clarity. The effect of metyrapone treatment in patients was assessed using multiple ANCOVA for each outcome with the post-treatment value as the dependent variable and pretreatment value as the covariate.

Face emotion recognition task

Background

This task assesses the ability of a participant to identify each of the six basic facial emotions, and previous research has shown an impaired ability to identify emotions in depression, an abnormality that resolves with clinical improvement.[97] Early improvement in recognition of happy faces has been reported to predict response to antidepressant treatment.[98]

Each emotion (happiness, sadness, anger, fear, surprise, disgust) was presented for 1.0 second at three intensities (30%, 50% and 70%) by four actors, together with 12 presentations of neutral expressions. The inter-trial interval was 4.5 seconds, participants pressed a labelled key to indicate their choice of emotion.

The data were analysed as accuracy (hit rate measured as the proportion of correct identifications of each emotion to the total number of responses to that emotion) and misattributions (false alarm rate measured as the proportion of misattributions of a specific emotion to the total number of responses to other emotions). RTs are not reported as there was no instruction to react rapidly, and the need to choose between seven response keys makes interpretation of the results unclear.

Patients and control subjects were compared using baseline data by an ANCOVA with Huynh–Feldt correction: emotion type was a within-subject factor, group was a between-subject factor, and age and sex were covariates, given evidence that they can influence face emotion recognition.[97] Illustrative post hoc analysis was by ANCOVA for each emotion separately. Uncorrected degrees of freedom are presented for clarity.

The effect of metyrapone treatment in patients was assessed using multiple ANCOVA for each emotion with the post-treatment value as the dependent variable and pretreatment value as the covariate.

Emotional memory task

This task measures immediate recall as a measure of encoding and recognition memory after 30 minutes for emotional and neutral words. It tests both overall memory and whether or not there is an emotional bias. Depressed patients have been reported to have both impaired memory and a negative memory bias.[99]

The task involved asking participants to learn a target list of 30 words (10 positive, 10 negative and 10 neutral), with words presented one at a time for 5 seconds on a computer screen. The words were presented once and then the participant was asked to recall as many words as possible straight after the presentation. Delayed recognition was tested after 30 minutes. This involved the respondent identifying whether or not words presented, one by one, on a computer screen (at a rate of one word every 1.2 seconds) were in the target list (responding ‘yes’ or ‘no’). The recognition list included the 30 target words and 30 distractors. An alternative version was available for repeat testing.

Group comparisons of recall and recognition were analysed using repeated measures ANCOVA, with valence the within-subject factor (positive, negative, neutral) and group as the between-subject factor and Huynh–Feldt correction applied. Zero-centred IQ was used as a covariate. Post hoc analysis of emotional bias was tested by univariate analysis, with separate analysis of each valence as the dependent variable; group as a fixed factor; and baseline values for total words recalled/recognised as the covariate. Uncorrected degrees of freedom are presented for clarity.

Effect of treatment was analysed using univariate analysis, with the post-treatment value the dependent variable, metyrapone or placebo as a fixed factor, and baseline value a covariate.

Immediate recall was measured as number of words of each valance remembered. It had been planned to analyse intrusions but unfortunately these had not been sufficiently systematically recorded to give valid results.

Delayed recognition was analysed as number of hits and numbers of false alarms for each emotion and overall. RT results are not reported as there was no instruction to respond as rapidly as possible, making interpretation unclear.

Affective GoNoGoTask

This task is measure of attentional emotional bias by measuring interference caused by words of conflicting valence. Non-depressed participants have been reported to respond more quickly to positive than negative words and the reverse has been reported for depressed patients.[99]

The computerised task consisted being asked to respond as quickly as possible with a key press to target words presented on the screen, and to ignore distractor words. Six types of blocks were presented made up of positive, negative and neutral targets, and of positive, negative and neutral distractors, for example positive targets and negative distractors (GoPos+NoGoNeg), positive targets and neutral distractors (GoPos+NoGoNeut), negative targets and positive distractors (GoNeg+NoGoPos), etc. For each type of block, participants were instructed as to which type of target to respond. Each block contained nine targets and nine distractors, with each word presented for 1.75 seconds. Each type of block was presented twice, making 12 blocks in all. Some early participants received a version of the task with three blocks of each type but it was decided that this made the task too long. These results are included in the analysis correcting for the number of presentations by using hit rates (hits/targets presented) and false-alarm rates (false alarms/distractors presented).

Given the large number of ways the data from the task could be analysed, and the number of blocks, we collapse the blocks by either target or distractor to give three Go conditions (GoPos, GoNeg, GoNeut) and three NoGo conditions (NoGoPos, NoGoNeg and NoGoNeut) and we present only the RT data. Participants were excluded if the number of correct hits in any block fell outside three standard deviations (SDs) from the mean (to exclude those who did not understand the instructions or were unable to identify the correct valence).

Patients and control subjects were compared by an ANCOVA on pretreatment data with Huynh–Feldt correction: valence was a within-subject factor; group was a between-subject factor; and zero-centred age was a covariate. Post hoc analysis was by ANCOVA for each emotion separately and also covaried for each individual's mean RT to reveal whether or not valence affected response times. Uncorrected degrees of freedom are presented for clarity.

The effect of metyrapone treatment in patients was assessed using multiple ANCOVA for each emotion with the post-treatment value as the dependent variable and the pretreatment value as the covariate.

Functional magnetic resonance imaging

Newcastle

We investigated levels of blood-oxygen-level-dependent (BOLD) signal changes in response to negative emotional stimuli in a subgroup of patients in the Newcastle centre using fMRI. Two emotional task paradigms were utilised, which are known to recruit both limbic and prefrontal brain regions. The facial emotion processing (FEP) task was based largely on work by Hariri *et al.*,[100] and included two emotional conditions ['emotional matching' (EM) and 'emotional labelling' (EL)] and one control condition ['shape matching' (SM)]. The emotional enhancement of memory (EEM) task was based on work by Kensinger and Corkin[101] and included encoding conditions for negative arousing, negative non-arousing and emotionally neutral words.

The FEP task presented triplets of stimuli to participants, with two stimuli at the bottom left and right of the screen, and one stimulus in the centre top. In all conditions participants had to indicate via button press which of the two stimuli at the bottom matched the stimulus at the top. In the EM condition, all three stimuli were greyscale photographs of faces (from the Ekman set),[102] showing either anger or fear, but from two different actors, with one of the bottom faces being repeated at the top. Thus theoretically no explicit processing of the emotion in the faces was required. In the EL condition, the bottom images showed one face displaying anger and one face displaying fear; the centre-top stimulus was an emotional label ('anger' or 'fear'). Thus participants had to explicitly identify which face showed the indicated emotion. The SM condition showed simple elliptical and circular shapes, filled with random pixels of a similar distribution of greyscale values as the faces. Similar to the EM condition, the top shape was exactly identical to one of the bottom two shapes. Each triplet was presented for 3.5 seconds in blocks of five trials. Within each condition, the order of trials was randomised. Blocks of trials were separated by periods of baseline that were between 15.25 and 15.75 seconds long. Three blocks per condition were presented in one of the two predefined orders of blocks.

In the EEM task, participants were presented with individual words displayed in the centre of the screen, and had to decide whether or not the word was 'concrete' or 'abstract'. Unbeknown to participants, all words fell into one of three categories according to their emotional valence and arousal, based on published word norms: negative arousing words, negative non-arousing words and neutral words. Individual words were presented for 2400 milliseconds, followed by a fixation cross of 600 milliseconds. They were presented in short blocks of three or four words from the same category. Two blocks of trials were always presented directly after each other. They were then followed by a short block of three trials of a distractor task. In the distractor task, participants were presented with days of the week and had to make a decision whether or not the day was in the first or second half of the week. The day directly in the middle of the week, here defined as Thursday, was never presented. A baseline period of 9 seconds followed each distractor block. In total, 28 words of each category were presented. There were two fixed orders of trials that were counterbalanced across participants. Owing to the repeated nature of the task for the patient group, all participants were informed that they would be asked to remember the presented words after the scan. Data from this recognition test will not be presented here.

Scanning took place on an Achieva® 3.0T MR scanner (Philips Healthcare, Best, the Netherlands). The functional tasks used standard echo planar imaging (EPI) sequences [FEP task: voxel size $2.5 \times 2.5 \times 4\ mm^3$, 35 slices, repetition time (TR) = 2500 milliseconds, echo time (TE) = 30 milliseconds; EEM task: $3 \times 3 \times 4\ mm^3$, 33 slices, TR = 2250 milliseconds, TE = 30 milliseconds]. To reduce motion-related distortions and artefacts, slices were orientated sagittally to keep the most common type of motion (head tilts) within slices. REST (REgional Saturation Technique) slabs were positioned above and below the head. In addition to these functional sequences, a standard T1-weighted, 1-mm isotropic anatomical scan was performed. The anatomical scans were utilised in the preprocessing of the functional data.

Functional image processing was undertaken using Statistical Parametric Mapping (SPM8) (Wellcome Trust Centre for Neuroimaging, University College London, London, UK; www.fil.ion.ucl.ac.uk/spm) and followed standard procedures, including slice timing correction, realignment and co-registration to the anatomical scan. First-level models were computed in native space and included regressors for the different stimulus types, the estimated realignment parameters and their first derivative. Contrast images from these models were normalised to standard space using parameters estimated from the anatomical scan and spatially smoothed with a 8-mm full width at half maximum Gaussian kernel. Second-level models were created using the results from all included participants. Whole-brain analyses were performed to identify if the different task conditions resulted in differential BOLD signal across both groups of participants in the amygdala. If so, BOLD signals were extracted from the significant clusters. In addition, BOLD signals were extracted from anatomically defined regions of interest of the amygdala using probabilistic cytoarchitectonic maps.[103] Extracted BOLD signals were analysed using repeated measures ANOVA, with condition as within-subject factor, group as between-subject factor, and age, sex and premorbid IQ as covariates.

Manchester

Subjects were scanned using a parallel, between-subjects design on a 1.5-tesla (1.5-T) scanner. The scanning session, which took place between 12 noon and 3 p.m., consisted of a brief localisation magnetic resonance imaging (MRI) scan, followed by two fMRI protocols in which participants undertook the n-back task followed by the encoding emotional pictures memory task. After a 6-minute structural MRI scan, subjects completed the retrieval emotional pictures memory task.

All computerised tasks were run on a PC in E-Prime 2.0 (Psychology Software Tools Inc., Sharpsburg, PA, USA) and back-projected onto a screen visible to the participant via two mirrors attached to the head coil. The task responses were acquired from the participant using a fibre-optic button box held in the right hand. All participants received standardised training on how to perform the tasks prior to the scan.

The task used by Whalley *et al.*[104] was adapted to create an in-scanner recognition memory task. Additional images from the IAPS battery were used to lengthen the task.

For the encoding task, a total of 60 images were shown in five blocks of six positive and six neutral images (each block lasting 48 seconds). These were followed by a block of rest, consisting of a fixation cross, which also lasted 24 seconds each. There were five blocks of each condition shown, making the task length 360 seconds in total. Participants were asked to indicate, using the button box, whether or not they felt that the picture was 'emotional', and instructed to remember the images.

During the recognition segment of the task, a total of 96 images were shown in six blocks of eight positive images and six blocks of eight neutral images (each block lasting 48 seconds). These were followed by a block of rest, consisting of a fixation cross, which also lasted 24 seconds each. There were six blocks of each condition shown, making the task length 432 seconds in total. All images shown in the encoding section were re-shown. Participants were asked to indicate, using the button box, whether or not they recognised the image from the encoding task.

A blocked version of the n-back task was adapted from the task of Koychev *et al.*[105] It consisted of four blocks of zero-back and four blocks of two-back, each lasting 32 seconds. These required three correct responses from 13 stimuli per block using the button box. These blocks were interspersed with four blocks of rest consisting of a fixation cross, lasting 20 seconds.

Scanning was carried out on an Intera® 1.5-T MRI scanner (Philips Healthcare, Best, the Netherlands) with prospective motion correction. Data were acquired with T2*-weighted, gradient echo EPI. Full brain coverage was used with TR = 2 seconds, TE = 40 milliseconds, 3.5 mm in-plane resolution, 4.5-mm slice thickness with 0.5-mm slice gap and 29 slices.

Data analysis

Behavioural and demographic data were analysed using Statistical Package for Social Science version 20.0 (SPSS Inc., Chicago, IL, USA). Behavioural data from both tasks were analysed using repeated measure ANOVA with one within-subject factor condition (difficulty: e.g. emotional valency – positive, neutral or zero-back, two-back) and one between-subject factor (participant group), as well as independent *t*-tests. The outcome measures used for the tasks were speed of response, number omitted, number of false positives and number correct. Results are presented as mean and SD, except when adjusted means from ANCOVA analyses are reported with standard error of the mean (SEM). Error bars in figures are also plotted as SEM for clarity, unless otherwise stated.

Data from fMRI were processed and analysed using SPM8. Scans were spatially preprocessed using standard protocols. They were realigned, using the first scan as a reference, normalised into the Talairach and Tournoux[106] stereotactic space using Montreal Neurological Institute (MNI) templates then spatially smoothed using a 8-mm Gaussian kernel. A mask was created, which was a sum of the regions of interest (ROIs) for each task and used as single comparison in the analysis.

For both of the emotional picture tasks (encoding and retrieval), positive and neutral conditions were contrasted with rest. Second-level processing was performed using one-sample *t*-tests to explore the main effect of task on positive pictures minus rest, neutral pictures minus rest, all pictures minus rest and vice versa, and a two-sample *t-test* to explore the effect of depression on each level of the task (HVs minus patients and vice versa). For the encoding task, our ROIs were the amygdala, hippocampus and parahippocampal gyrus, dorsolateral prefrontal cortex (dlPFC) and inferior temporal area [Brodmann area 9 (BA9), BA20 and BA45]. For the retrieval task, the a priori regions of interest were amygdala, hippocampus, parahippocampal gyrus, anterior cingulate cortex (BA32), insula, dlPFC and frontopolar areas (BA9 and BA10).

For the n-back task, both levels of n-back were contrasted with rest. Second-level processing was performed using one-sample *t*-tests to explore (1) the main effect of task at the highest working memory load (two-back minus zero-back) and (2) the effect of depression on each level of the task (HVs minus patients and vice versa). Our ROIs were lateral premotor cortex (BA6), dorsal anterior cingulate cortex (BA32), dlPFC (BA9, BA46) ventrolateral prefrontal cortex (BA44, BA45, BA47), medial posterior parietal and inferior parietal lobule (BA7, BA40) and medial cerebellum, as per Symonds *et al.*[107]

Within these areas we report as significant areas family-wise error (FWE) corrected peak level $p(\text{FWE}) < 0.05$ for the region of interest. Other areas surviving $p(\text{FWE}) < 0.05$ at a whole-brain level are also reported for interest.

Acute effects of hydrocortisone in treatment-resistant depression using pharmacological functional magnetic resonance imaging

We aimed to examine the effect of an acute bolus of hydrocortisone on patients with TRD in comparison with age- and sex-matched HVs. Using pharmacological functional magnetic resonance imaging (phMRI),[107] we investigated whether or not BOLD signal is altered after acute administration of 100 mg of intravenous hydrocortisone in a between-subject, placebo-controlled parallel design study of 30 patients and 30 control subjects.

Electroencephalography

To examine the predictive ability of these identified EEG variables in response to metyrapone augmentation, patients consenting to EEG recordings had recordings made prior to commencement of treatment with metyrapone. Of particular interest was whether or not a high slope in the LDAEP analysis, predictive of response to SSRIs, would predict response to metyrapone, suggestive of an enhancement of serotonergic function. To examine both the effect of emotion on episodic memory retrieval and the putative ERP marker of LTP, EEG was recorded prior to treatment with metyrapone in patients and a cohort of healthy control subjects, with patients tested again at week +5.

Electroencephalography was recorded from 29 active silver/silver chloride electrodes attached to the scalp using an EEG cap (EasyCaps, Germany) with electrodes placed according to the international 10–20 system. Two further electrodes were placed on each mastoid process (right mastoid as a reference; left mastoid as an active channel), two on the outer canthi of the eyes to record horizontal eye movements [horizontal electrooculogram (HEOG) channel] and one on the nasion with linked electrodes under the midpoint of each eye to record vertical eye movements [vertical electrooculogram (VEOG) channel]. All electrode impedances were kept at < 5 kiloohms (kΩ). Recordings were made using a Neuroscan® SynAmps2® amplifier (Compumedics Germany GmbH, Singen, Germany) with an analogue–digital conversion rate of 500 Hz, and high- and low-pass filters of 0.5 Hz and 30 Hz, respectively.

Resting electroencephalography

Resting EEG was recorded in four 5-minute blocks, with participants alternating between an eyes open and eyes closed state (OCOC). Participants were instructed to sit still, avoiding any unnecessary movements, and to look straight ahead during the 'eyes open' state.

Loudness dependency of auditory evoked potentials recordings

The LDAEPs were recorded in relation to 200 individual 1000-Hz tones (50-milliseconds total duration, 10-milliseconds rise and fall times) presented binaurally through headphones at a random repetition rate of 0.56–0.45 Hz. The tones were delivered at intensities of 54, 64, 74, 84 and 94 dB (40 of each) in a randomised order.

Emotional Source Memory Task recordings

The Emotional Source Memory Task is divided into two elements: a study phase and a test phase. In the study phase, participants were presented with an emotionally valenced background picture of an event or scene drawn from the IAPS and were asked to rate its valence (7-point Likert scale, ranging from strongly negative to strongly positive). Following this, the background picture was superimposed with a neutral object. Participants were asked to make an association between the object and the event/scene in the background picture. Following completion of the study phase and a 5-minute delay, the test phase entailed a surprise memory test throughout which EEG was recorded. During the test phase, old and new neutral objects were presented, and participants had to make a recognition judgement responding either 'old' or 'new' via a keypad. If the participant recognised the object from the study phase, and responded as such, a follow-up recollection response was required. For this, the participant was asked to recall the valence of the associated background presented at study with the neutral object and respond via a keypad (positive, neutral, negative, don't know). Analysis focused on behavioural data and ERPs of correctly identified old objects that were also associated with a correct identification of background valences (Hit–Hits), with relation to the emotional valence rating of the associated background given during the study phase.

Long-term potentiation recordings

During each testing session visual evoked potentials (VEPs) were recorded. The visual stimuli and the experimental procedure were based on those used by Teyler *et al.*[69] The stimulus (a series of chequerboards) was presented on a computer monitor (circular chequerboard stimuli subtending 8 degrees of visual angle, check size 0.3 degrees of visual angle). Participants were instructed to fixate on a red square in the centre of the screen. During baseline testing, the stimuli were presented at a frequency of 0.8 Hz (duration 33 milliseconds), and there were two baseline blocks of 100 flickers each. Baseline testing was followed by a block of photic tetanus when the identical chequerboard stimulus is presented at 8.6 Hz for 1000 flickers. This was followed by 2 minutes' rest with eyes closed, followed by two post-tetanus test blocks (with stimulus presentation as per baseline testing) at 4-minute intervals. EEG was recorded during the two baseline and two post-tetanus test blocks.

Electroencephalogram analysis

All EEG analysis was conducted using Neuroscan 4.5 software. EEG recordings were arithmetically adjusted to a linked mastoid reference. Blink correction was performed on the basis of VEOG recordings using the Neuroscan's ocular artefact reduction function. Residual artefacts were removed, whereby any epoch, for which any channel (including HEOG but not VEOG) had a voltage deflection of $> +75$ μV or < -75 μV, was rejected.

It did not prove worthwhile analysing the EEG predictors of response, nor the ESMT ERPs for reasons described below. However, the LTP EEG data were analysed. This is described in *Chapter 3*.

Continuous EEG recorded during the VEP test phases was epoched from –100 milliseconds to + 500 milliseconds around the presentation of the visual chequerboard stimulus. Epochs were artefact corrected and rejected as described above. Retained epochs were averaged to generate VEPs. To determine time windows within which N1, P1 and N1b amplitudes would be determined, peak detection using Neuroscan algorithms was run on the grand-averaged VEP data (all time points, all subjects) for each individual channel. The focus of analysis was on the Oz electrode, as this has been shown to demonstrate the largest and most robust VEP.[69] The average amplitude of N1, P1 and N1b was calculated as an area under the curve for a 20-milliseconds window centred at 160 milliseconds for N1, 248 milliseconds for P2 and 204 milliseconds for N1b, being halfway between N1 and P2. The magnitude of VEP potentiation was calculated as the difference between averaged N1, P2 and N1b components pre-tetanic stimulation and the first post-tetanic VEP measure.

Important changes to methods

The initial study aim was to recruit 190 patients. However, as a result of initial slow recruitment rates and after discussion with the funder, sponsor and the independent Data Monitoring and Ethics Committee (DMEC), a revised target of 140 was agreed by accepting a reduction in power of the study from 90% to 80% on the primary outcome measure.

Participants

Patients

Eligibility criteria

- *Diagnostic and Statistical Manual of Mental Disorders*-Fourth Edition (DSM-IV)[108]-defined major depressive episode, assessed using the Structured Clinical Interview for DSM (SCID) research version.[109]
- HDRS17[110] score of ≥ 18 at weeks –2 and 0 (see below).
- Massachusetts General Hospital Treatment Resistant Depression (MGH-TRD) staging score of 2–10 as a measure of treatment refractoriness.[111] This cut-off is based on MGH-TRD scores seen in primary and secondary care, but short of tertiary care, in the NHS.[112]
- Single-agent or combination antidepressant treatment that includes a serotonergic drug (SSRI, tertiary amine tricyclic, venlafaxine, duloxetine or mirtazapine). At the point of randomisation, patients must have been on their current antidepressant medication, at the current dose, for a minimum of 4 weeks.
- Aged between 18 and 65 years. An upper age limit was included to reduce rates of physical health comorbidities that could complicate decisions around the safety of prescription of a drug (metyrapone) that may lower cortisol levels.

Exclusion criteria

- Other DSM-IV axis 1 diagnosis, other than an anxiety disorder considered to be secondary to a primary diagnosis of depression, confirmed using SCID.
- Physical comorbidity that would make metyrapone inappropriate, including untreated hypothyroidism, disorders of steroid production, cardiac failure, angina, myocardial infarction in the last 3 years and renal failure.
- Pregnancy or breastfeeding.
- Use of medication that would interfere with metyrapone.
- Dependence on alcohol or other drug(s) in the past 12 months, and/or current harmful use of such substances (defined as meeting SCID criteria for harmful use or dependence).
- Current or recent participation in a research study that could interfere with results.

Healthy volunteers

Eligibility criteria

- No evidence of axis 1 disorder on the DSM-IV,[108] assessed using the SCID research version.[109]
- Aged between 18 and 65 years.

Exclusion criteria

- Physical comorbidity as per above (see *Patients*).
- Pregnancy or breastfeeding.
- Use of medication as per above (see *Patients*).
- Dependence on alcohol or other drug(s) in the past 12 months, and/or current harmful use of such substances (defined as meeting SCID criteria for harmful use or dependence).
- Current or recent participation in a research study that could interfere with results.

Settings and locations

Patient identification took place across two hubs of the UK National Institute for Health Research MHRN – the North East, with centres in Newcastle and Durham/Teesside, and the North West, with Manchester as the centre – and in the West Yorkshire Comprehensive Local Research Network (CLRN), with centres in Leeds and Bradford. Potential patient participants were identified through routine clinic appointments at study sites and Participant Identification Centres that included Primary Care and Community Mental Health Teams. HVs were recruited by advertisements in the University of Manchester and by e-mails sent to the Volunteer Database of the Institute of Neuroscience, Newcastle University.

Intervention

Participants continued their existing antidepressant regime. Randomisation was in a ratio of metyrapone to matched placebo using permuted block randomisation and a web-based system, to ensure concealment of allocation. To ensure blinding the metyrapone was overencapsulated and the placebo was visually identical. Participants received study drug (metyrapone 500 mg or placebo) twice daily, prescribed in the morning and at noon, for 21 days. Adherence to medication was assessed using measures of 11-deoxycortisol as described above (see *HPA axis assessment*).

Apart from treatment with the experimental intervention, all other treatments remained under the control of the patient's normal treating psychiatrist and/or GP. However, the patients' clinicians were encouraged to avoid changes to any medications, particularly antidepressants, between enrolment (week –2) and the primary outcome time point (week +5) unless there was a compelling clinical reason to alter treatment. All current medication was recorded at all follow-up time points.

Outcomes

A full detailed Statistical Analysis Protocol was drawn up (and reviewed and agreed by the DMEC) prior to completion of the study and breaking of the study blind. The Statistical Analysis Protocol is appended as *Appendix 1*.

The analysis for the primary outcome was an intention-to-treat ANCOVA of the MADRS scores at +5 weeks, with baseline MADRS included as a covariate, differences between strata (centres and whether or not the patient originated in primary or secondary care) and the differences between groups (patients randomised to metyrapone vs. patients randomised to placebo) were included as fixed effects. The persistence of change in the MADRS score was assessed using repeated measures ANOVA utilising data from all time points. Missing values were imputed as described in the Statistical Analysis Plan (SAP: see *Appendix 1*). Change and persistence of the change in other clinical and QoL measures (secondary outcomes) were examined using the same methods. Additional secondary outcomes included rates of response (defined as a $\geq 50\%$ reduction in MADRS score) and remission (defined as MADRS ≤ 10) with metyrapone compared with placebo at week +5. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples.

With regard to the mechanistic objectives, the study examined whether or not treatment with metyrapone led to a change in HPA axis function (assessed by examining the CAR and 11 p.m. cortisol measures) and whether or not changes in HPA axis function, compared with baseline, were seen 2 weeks post treatment. Both baseline HPA axis function and change in function with treatment were assessed to see if they predicted clinical response to metyrapone defined by the change in MADRS between week 0 and week +5, in an exploratory analysis.

Sample size

The sample size was based on detecting a difference between groups in the change in MADRS scores between baseline and 5 weeks post randomisation. The effect size for this measure in the Jahn study[33] was 0.63. The study was powered around the more conservative moderate effect size of 0.5 which, assuming a post-intervention SD of 12 points,[33] corresponds to a 6-point difference on the MADRS. An achieved sample size of 85 per group was required to detect effect size of 0.5 with 90% power, assuming $\alpha = 0.05$. Allowing for 10% attrition during the trial, the original aim was to randomise 95 per group (190 in total). However, as described above, the recruitment target was modified to accept a power of 80%, requiring a sample size of 63 per group. Again allowing for 10% attrition during the trial, we aimed to randomise 70 per group; 140 in total.

Randomisation and blinding

Patients were randomised in a 1 : 1 ratio to metyrapone or placebo using permuted block randomisation, stratified by centre (North East, North West or Leeds/Bradford), level of care setting (primary or secondary care) and, for the North East and North West centres, by whether or not the patient agreed to participate in the mechanistic studies. The randomisation code to produce random permuted blocks was generated by an independent statistician in the Newcastle Clinical Trials Unit. The length of each block was randomly set at either 2 or 4 (with equal probability), with these limits being concealed from study personnel to ensure concealment of allocation.

Coded (numbered) packs of study drug and matched placebo were produced according to the randomisation schedule, by Catalent Clinical Trials Supply Company, Corby, UK. Metyrapone capsules were overencapsulated (using Coni-Snap® capsules, Capsugel, Morristown, NJ, USA) and appeared identical to the placebo capsules. These were distributed to the clinical trials pharmacist at Northumberland, Tyne and Wear NHS Foundation Trust, located in Newcastle.

Randomisation was achieved through the use of a centralised web-based system set up by the Newcastle Clinical Trials Unit. Within each site, patients were randomised when they returned for their second visit at week 0. This was carried out by a trained member of the staff at that site. After logging on to the system, he/she entered patient details including initials, date of birth and status, with respect to each of the stratification variables. The system then supplied a study identification number to be used for that patient. This was entered on to a prescription for study medication, which was then signed by an authorised clinician at the site. A copy of the prescription was faxed and the original mailed to the pharmacy in Newcastle. Patients received their medication through the post at their home address.

In case of serious untoward incidents, the code for a particular subject was available, without compromising the blind for the other subjects, via the research pharmacist. Existing on-call arrangements with pharmacy were utilised so that emergency code breaks could be accessed out of hours. The randomisation list was available only to the study pharmacist at Northumberland, Tyne and Wear NHS Foundation Trust, the independent statistician in the Newcastle Clinical Trials Unit who produced the list and the database manager who programmed the system. In actuality, no requests were made during the course of the study for unblinding information.

Blinding was maintained for all patients, the clinicians who cared for them, outcome assessors and data analysis team until the protocol for the final 24-week visit had been followed. There were no requests from DMEC to unblind patients, although they did examine the blinded group data to look for any relationship between those taking metyrapone and suicidal thoughts/attempts.

Statistical methods

Statistical methods used to compare groups for primary and secondary outcomes

The primary outcome was the MADRS score at 5 weeks post randomisation adjusted for the MADRS score at baseline. The mean difference between treatment groups was estimated using a mixed model, assuming a normal error structure. The dependent variable was the MADRS score at 5 weeks; the baseline score was included as a covariate. Difference between centres was included as a random effect; origin of care (primary or secondary) was included as a fixed effect. Unadjusted estimates of the difference between groups are also presented (variation between centres and differences between origin of care were removed from the model). The analysis was undertaken on an intention-to-treat basis (i.e. groups were defined by randomised allocation rather than the actual treatment received).

Further secondary analysis of the primary outcome was prespecified. The changes in MADRS scores in the two randomised groups 5 weeks post randomisation were compared using an independent sample *t*-test. Two further binary outcomes were defined. Response was defined as a reduction in MADRS score that was less than or equal to half of the score at baseline; remission was defined as a MADRS score that was ≤ 10. These variables were analysed using the mixed-model approach described above, but with the assumption of a binomial error structure and logit link function.

Secondary outcomes (including the CAS, BDI and STAI) and QoL (EQ-5D health tariff and EuroQol health scale) were analysed using the mixed-model approach described above.

A secondary clinical objective was to consider the persistence of the effect of metyrapone. The prespecified analysis to assess persistence was to extend the mixed-model approach described above to include repeated measures on individuals. Additional explanatory variables included in the model were change over time (differences between scores observed at different visits) and an interaction between these changes and treatment group.

A further clinical objective was to assess how well metyrapone was tolerated. Side effects were assessed using the TSES and YMRS. The YMRS was analysed using the mixed-model approach described above. For the TSES, the relative risk (RR) of individual symptoms in the two groups was calculated.

The number of non-serious adverse events (AEs) reported by patients in each of the randomised groups was compared using a negative binomial regression model.

Methods for additional analyses, such as subgroup analyses and adjusted analyses

In the SAP, further per-protocol analyses were specified, in which groups were defined on the basis of compliance with medication, based on measurements of 11-deoxycortisol. (However, see *Chapter 4*, *Discussion*, *Limitations*.) Additional survival analysis of time to response and time to remission, based on a Cox proportional hazards model, was also specified.

Important changes to statistical methods as proposed

The primary analysis in the original protocol was an independent *t*-test of change in MADRS scores (baseline to week +5). The main reason for specifying this as the primary analysis was that it corresponds to the stated sample size calculation. The main disadvantage for specifying it as the primary analysis are the following recommendations:

(a) ANCOVA is usually to be preferred to an analysis of change scores (see, for example, Van Breukelen,[113] Senn[114]).
(b) The primary analysis should take into account the stratification factors used in the randomisation of patients (Parzen *et al.*,[115] Kahan and Morris[116]).

Consequently, when formulating the SAP, the primary analysis was defined as ANCOVA, and the analysis of change scores was defined as a secondary analysis of the primary outcome. The final version of the SAP was approved prior to the breaking of the blind and is given in *Appendix 1*.

In retrospect, the proposed secondary analyses of persistence and times to response and remission should have been specified as being conditional on metyrapone being proven to be clinically efficacious at the primary end point. With no observed reduction in depression 5 weeks after randomisation, there was no rationale for testing hypotheses based on differences between groups at further time points. For outcomes for which there was no evidence of a difference between groups 5 weeks post randomisation, summary statistics are presented for the subsequent visits.

The TSES was not analysed as a single score (as had originally been specified in the SAP), after a review of the literature describing its development and use indicated that this was not appropriate.[88]

Chapter 3 Results

Participant flow

The numbers of participants who were referred, screened, randomised and followed up to primary end point (5 weeks) and to the end of the study (24 weeks) are shown below in *Figure 1*. A total of 877 patients were referred to the study team: 237 from primary care, 320 from secondary care and 310 as self-referrals following media exposure of the study or seeing posters in GP surgeries. In 10 cases the source of the referral was not clear. Of these 877 patients, 284 were screened for eligibility. Of the 593 who were not

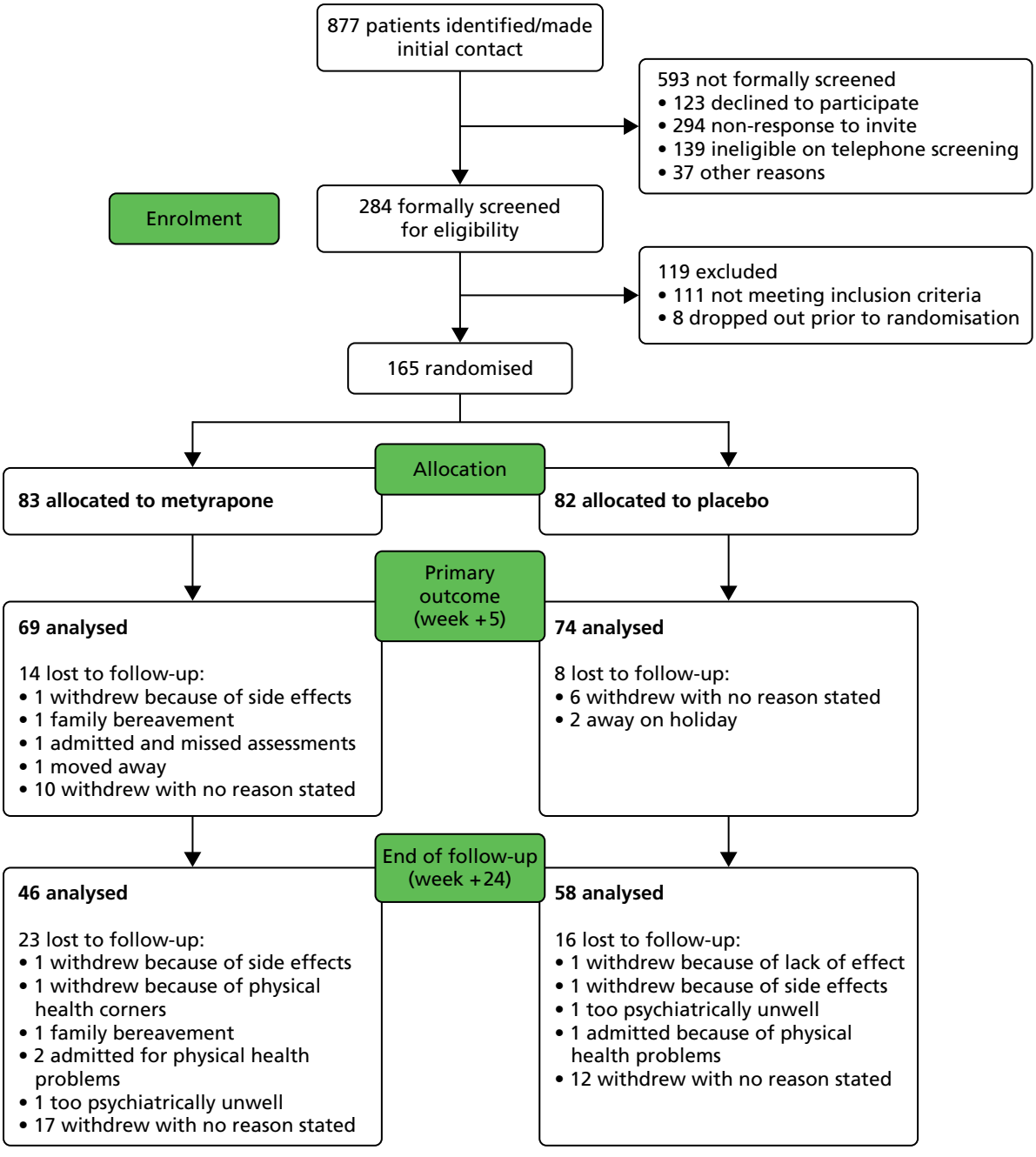


FIGURE 1 Consolidated Standards of Reporting Trials (CONSORT) diagram.

screened, 123 declined when contacted, 294 did not respond to an invitation for screening or did not attend and 139 were deemed ineligible on a brief telephone screening. The reasons for non-screening of 37 patients are not known. Of the 284 who were screened, 173 were deemed to be eligible. Of the 111 who did not meet inclusion criteria, 10 did not meet the criteria for a major depressive episode using the SCID, 52 had HDRS17 item scores of < 18, 17 had axis 1 disorders other than anxiety, nine were on an inappropriate antidepressant and three had MGH-TRD staging scores outside the range of 2–10,[58] 18 had physical disorders that excluded them and five had other miscellaneous exclusion criteria (three patients were excluded for more than one reason). Eight patients subsequently dropped out before randomisation (i.e. between weeks –2 and 0), so 165 patients were randomised (82 to placebo and 83 to metyrapone). For the purposes of randomisation and analysis, patients were divided into two groups: those recruited from secondary care and those recruited from primary care (defined here as all those not clearly identified as currently attending secondary care). Of the 165 randomised, 143 (86.7%) completed the primary outcome at +5 weeks (74 on placebo and 69 on metyrapone) but 22 dropped out of the study (i.e. withdrew from both treatment and follow-up data collection). A further 39 dropped out between weeks +5 and +24, so that 104 (63% of those randomised) completed the study (58 on placebo and 46 on metyrapone).

Overall, 42% of patients were recruited from primary care (*Table 2*). This figure differed between sites. In Newcastle and Durham, 69% of patients were recruited from primary care; for Leeds/Bradford this figure was 40%, and in Manchester only 14% of patients were recruited from primary care.

The allocation of the patients into randomised groups by site and origin of care is shown in *Table 3*.

The losses and exclusions after randomisation for both groups are shown in *Table 4*. The only reason that patients were excluded from the intention-to-treat analyses was that they declined to attend follow-up

TABLE 2 Origin of patient by site

Primary or secondary care?		Site			Total
		Leeds/Bradford	Manchester	Newcastle/Durham	
Primary	Number	15	8	47	70
	% within site	39.5	13.6	69.1	42.4
Secondary	Number	23	51	21	95
	% within site	60.5	86.4	30.9	57.6
Total	Number	38	59	68	165
	% within site	100.0	100.0	100.0	100.0

TABLE 3 Allocation of patients into groups by site and origin

Source of patient	Placebo		Metyrapone	
	n	%	*n*	%
Site				
Leeds/Bradford	19	23.2	19	22.9
Manchester	29	35.4	30	36.1
Newcastle/Durham	34	41.5	34	48.2
Origin				
Primary care	35	42.7	35	42.2
Secondary care	47	57.3	48	57.8

TABLE 4 Attendance at follow-up visits by group to which randomised

	Group to which randomised						
	Placebo ($n = 82$)		Metyrapone ($n = 83$)		RR of attending		
Visit	*n*	%	*n*	%	RR	95% confidence limits	
						Lower	*Upper*
Week 3	77	93.9	72	86.7	0.92	0.83	1.02
Week 5	74	90.2	69	83.1	0.92	0.82	1.04
Week 8	66	80.5	62	74.7	0.93	0.79	1.09
Week 16	61	74.4	54	65.1	0.88	0.71	1.07
Week 24	58	70.7	46	55.4	0.78	0.61	0.99

visits despite prompts and their data were therefore not available. There were no exclusions for other reasons.

Dropout was broadly similar in each of the two arms of the trial, although a slightly greater proportion of those patients randomised to metyrapone failed to attend follow-up visits.

Recruitment

Dates defining recruitment and follow-up

The first patient was randomised at the beginning of March 2011 and the last patient in mid-December 2012. The last 24-week follow-up was in late June 2013.

Recruitment was initially slow largely due to the complexity of the protocol which some clinicians and patients found too daunting. This required the development of a recovery plan, agreed by the funder at month 10, which included some changes to the protocol to reduce of the number of exclusions related to medical comorbidities (criteria shown above are final ones). The presence of any axis 1 disorder in addition to a major depressive episode was initially an exclusion criterion. This was relaxed to allow inclusion of patients with an axis 1 diagnosis of an anxiety disorder. There was much greater emphasis by the MHRN and others on promoting the study within clinical teams. The CLRN and Northumberland, Tyne and Wear Trust supported the study by funding junior clinicians through the Research Capability Funding process and these were an invaluable help, in particular with patients' screenings. Several media events were also very helpful in raising awareness of the study and enhancing recruitment. Adverts were also placed in some local publications in Newcastle and Leeds, and this was a major reason for the higher proportion of primary care patients in Newcastle than, for example, Manchester.

Why did the trial end or stop?

The trial was stopped at the end of the study, as agreed with the Efficacy and Mechanism Evaluation (EME) Board after the last patient's last visit. Owing to enhanced recruitment ($n = 165$ randomised), we were able to exceed the revised target of 140, giving us 84% power to detect the effects hypothesised in the sample size determinations. The trial period included a 7-month extension.

Baseline data

Baseline data of demographic and clinical characteristics are shown in *Table 5*. The placebo and the metyrapone groups were well balanced on key demographic variables and clinical characteristics.

TABLE 5 Baseline demographic data and clinical characteristics

Characteristics	Group			
	Placebo ($n = 82$)		Metyrapone ($n = 83$)	
Gender	**n**	**%**	**n**	**%**
Male	30	36.6	36	43.4
Female	52	63.4	47	56.6
Race	**n**	**%**	**n**	**%**
White	77	93.9	80	96.4
Other	5	6.1	3	3.6
Age	***Mean***	***SD***	***Mean***	***SD***
Age (years)	45.2	10.4	47.6	9.9
Suicide risk	**n**	**%**	**n**	**%**
Zero	16	19.5	11	13.3
Low	32	39.0	38	45.8
Medium	13	15.9	10	12.0
High	21	25.6	24	28.9
'Yes' responses to suicide questions	**n**	**%**	**n**	**%**
Wished that they were dead	56	68.3	61	73.5
Self-harm	14	17.1	18	21.7
Suicidal thoughts	34	41.5	30	36.1
Suicide plan	7	8.5	2	2.4
Suicide attempt in last month	1	1.2	0	0.0
Suicide attempt ever	44	53.7	49	59.0
Yes to at least one question	66	80.5	71	85.5
Smoking status	**n**	**%**	**n**	**%**
Non-smoker	36	43.9	23	27.7
Ex-smoker	16	19.5	25	30.1
Current smoker	30	36.6	35	42.2
***Alcohol consumption* (n = *82 and 83, respectively*)**	***Mean***	***SD***	***Mean***	***SD***
Units per week	6.9	12.8	7.0	11.2
Medical history	n	%	n	%
Medical condition	47	57.3	48	57.8
Visited hospital	67	81.7	64	77.1
Waiting to attend hospital	28	34.1	29	34.9
Allergies	33	40.2	32	38.6
Chest pain/wheeziness	28	34.1	34	41.0
Skin condition	24	29.3	29	34.9
Faints/fits/seizures	20	24.4	23	27.7
Weight change	37	45.1	38	45.8
Liver/kidney disease	10	12.2	6	7.2
Heart disease	14	17.1	12	14.5

TABLE 5 Baseline demographic data and clinical characteristics (*continued*)

Characteristics	Group			
	Placebo ($n = 82$)		Metyrapone ($n = 83$)	
Vital signs	***Mean***	***SD***	***Mean***	***SD***
Height (cm)	169.6	10.4	169	10.4
Weight (kg)	90.0	22.2	89.4	19.8
BMI (kg/m^2)	31.2	6.6	31.5	6.9
Respiratory rate (breaths per minute)	15.9	4.3	15.5	3.5
Systolic blood pressure, standing (mmHg)	132.9	15.8	133.7	19.7
Diastolic blood pressure, standing (mmHg)	86.1	11.7	86	12.2
Heart rate, standing (beats per minute)	79.1	14.4	79.4	15.2
Systolic blood pressure, sitting (mmHg)	133.7	15.9	132.2	18.4
Diastolic blood pressure, sitting (mmHg)	83.2	12.4	83.9	12.0
Heart rate, sitting (beats per minute)	71.4	11.9	79.4	15.2
Psychological health and QoL	***Mean***	***SD***	***Mean***	***SD***
STAI: state anxiety	41.0	5.8	42.8	6.5
STAI: trait anxiety	48.3	5.2	48.5	5.7
MGH	4.6	1.8	4.9	2.0
MADRS	28.1	5.4	27.7	6.7
CAS	10.0	4.6	9.5	4.5
BDI	34.8	10.3	35.6	10.9
HDRS17 at screening	23.0	3.9	23.3	3.9
HDRS17 at randomisation	22.3	3.2	22.2	3.5
EQ-5D	0.37	0.3	0.37	0.3
EQ-VAS	40.9	16.6	42.3	19.4
Big Five Inventory (n = 73 and 72, respectively)	***Mean***	***SD***	***Mean***	***SD***
Openness	3.14	0.85	3.11	0.85
Conscientiousness	3.17	0.73	3.08	0.79
Extroversion	2.23	0.79	2.32	0.85
Agreeableness	3.59	0.70	3.61	0.59
Neuroticism	4.31	0.54	4.28	0.56
Family history of psychiatric disorders (n = 82 and 82, respectively)	**n**	***%***	**n**	***%***
Any disorder	56	68.3	48	58.5
Depression	43	52.4	40	48.8
Suicide	12	14.6	8	9.8
Bipolar disorder	4	4.9	5	6.1
Anxiety	12	14.6	15	18.3
Obsessive–compulsive	5	6.1	6	7.3
Schizophrenia	6	7.3	7	8.5
Eating disorder	3	3.7	2	2.4
Drug or alcohol problem	9	11.0	10	12.2
Other	6	7.3	4	4.9

EQ-VAS, EuroQol visual analogue scale.

In general, for data arising from interview-administered questionnaires, there were very few missing data items (*Table 6*). There was a greater number of missing items in the data arising from self-completion questionnaires, but, overall, the level of missing data was very low and only a small number of missing values were imputed.

In general, for data arising from interview-administered questionnaires, there were very few missing data items. There was a greater number of missing items in the data arising from self-completion questionnaires, but, overall, the level of missing data was very low.

Numbers analysed

Analysis was on the basis of intention-to-treat by original assigned groups (placebo $n = 82$, metyrapone $n = 83$). Details of the numbers analysed for each outcome measure, at each time point, for both groups, can be found in *Appendix 2*.

TABLE 6A Mood, anxiety and QoL scores: level of missing data and data imputation at baseline

		No. of participants at baseline for whom:				
			Score has been calculated			
Measure	*n*	Score is missing	All	Without imputation	With imputation	Mean no. of imputed items[a]
MADRS	165	0	165	165	0	–
CAS	165	0	165	165	0	–
State anxiety	165	3	162	154	8	1.13
Trait anxiety	165	3	162	153	9	1.00
BDI	165	3	162	160	2	6.00
YMRS	165	1	164	164	0	–
EQ-5D tariff	165	3	162	162	0	–
EQ-5D health	165	6	159	159	0	–

a The mean number of missing items for the subset of participants from whom a total score was calculated using imputation.

TABLE 6B Mood, anxiety and QoL scores: level of missing data and data imputation at 5 weeks post randomisation

		No. of participants at primary end point for whom:				
			Score has been calculated			
Measure	*n*	Score is missing	All	Without imputation	With imputation	Mean no. of imputed items[a]
MADRS	143	0	143	143	0	–
CAS	143	0	143	143	0	–
State anxiety	143	4	139	136	3	2.00
Trait anxiety	143	4	139	137	2	1.00
BDI	143	4	139	135	4	3.75
YMRS	143	0	143	143	0	–
EQ-5D tariff	143	4	139	139	0	–
EQ-5D health	143	9	134	134	0	–

a The mean number of missing items for the subset of participants from whom a total score was calculated using imputation.

Outcomes and estimation

Results for the primary and secondary clinical outcomes and secondary analysis of it, as outlined above and described in detail in the prespecified SAP (see *Appendix 1*), are described below. There is one exception to this: the prespecified analysis excluding patients who discontinued treatment during the 3-week treatment period and another utilising the 11-deoxycortisol measure of apparent concordance with treatment are not included, as these analyses have not been completed owing to lack of availability of the data (see *Analyses not yet complete*, below).

Primary outcome measure: Montgomery–Åsberg Depression Research Scale

The primary outcome measure was depression as measured by the MADRS instrument. The key objective of the trial was to ascertain whether or not the strategy of treating patients with metyrapone resulted in a clinically significant reduction in depression at 2 weeks post treatment (5 weeks post randomisation) when comparing treatment with placebo.

Baseline

The mean MADRS score at baseline was 26.0 (SD 5.8) for patients recruited from primary care and significantly higher at 29.4 (SD 6.0) for patients recruited from secondary care; the mean difference at this time point was 3.4 [95% confidence interval (CI) 1.5 to 5.3]. Mean scores were non-significantly higher for patients recruited in Manchester than with the other two sites but this is primarily due to the higher proportion of patients recruited from secondary care in that site. The distribution of baseline scores by site and origin is shown in the box plots in *Figure 2*.

The horizontal line in the centre of each box corresponds with the median score; the lower and upper edges of the box correspond to the lower and upper quartiles of the distribution, respectively. The 'whiskers' show the distribution of the remaining values.

The distribution of scores in the two trial arms is shown in *Figure 3*. Details of the box and whisker plot are as described for *Figure 2*.

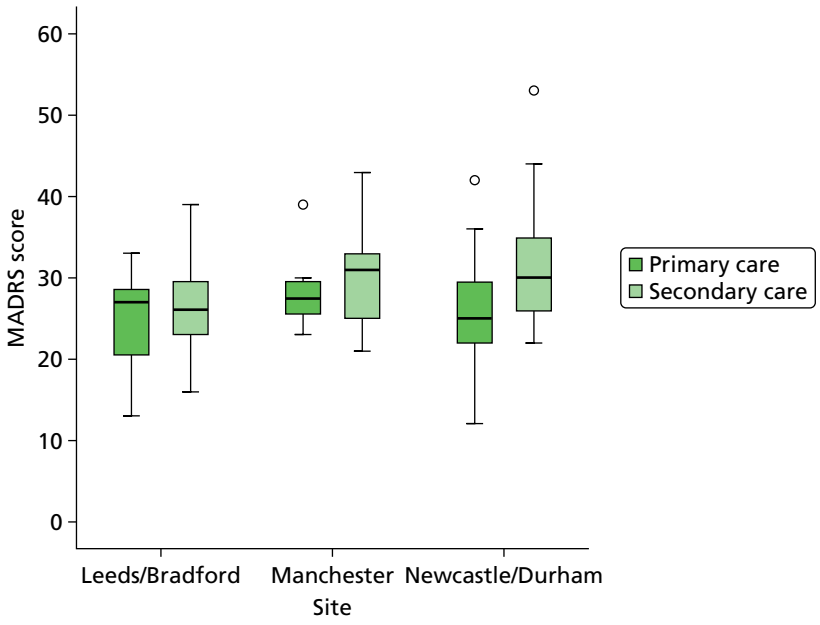


FIGURE 2 Box plots of MADRS scores by site and origin of patient.

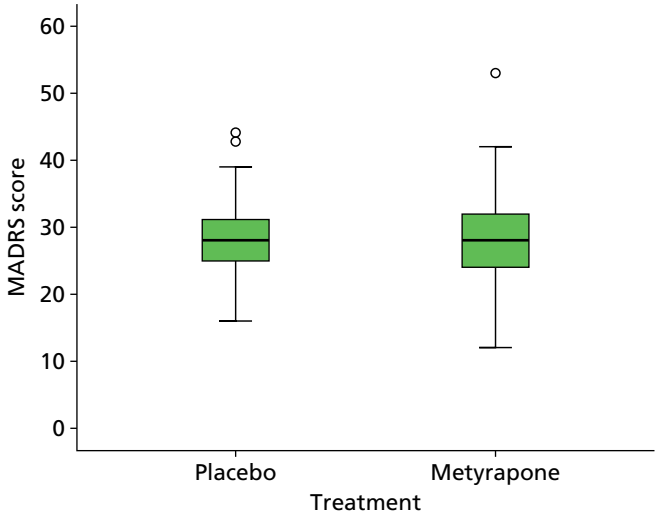


FIGURE 3 Box plots of MADRS scores by group to which randomised.

Effect of treatment group on Montgomery–Åsberg Depression Research Scale score

Summary statistics for the MADRS at all visits are given in *Table 7*, and the distributions of scores are illustrated in *Figure 4*.

TABLE 7 Montgomery–Åsberg Depression Research Scale scores by visit and group to which randomised

	Placebo			Metyrapone			Total		
Visit	Mean	*n*	SD	Mean	*n*	SD	Mean	*n*	SD
Baseline	28.1	82	5.4	27.7	83	6.7	27.9	165	6.1
+3 weeks	20.8	77	9.9	22.6	72	10.9	21.7	149	10.4
+5 weeks	22.4	74	10.6	21.7	69	10.9	22.1	143	10.7
+8 weeks	22.6	66	10.8	21.2	62	10.4	21.9	128	10.6
+16 weeks	20.5	61	11.5	21.4	54	11.0	20.9	115	11.2
+24 weeks	20.0	58	11.6	21.0	46	11.1	20.4	104	11.3

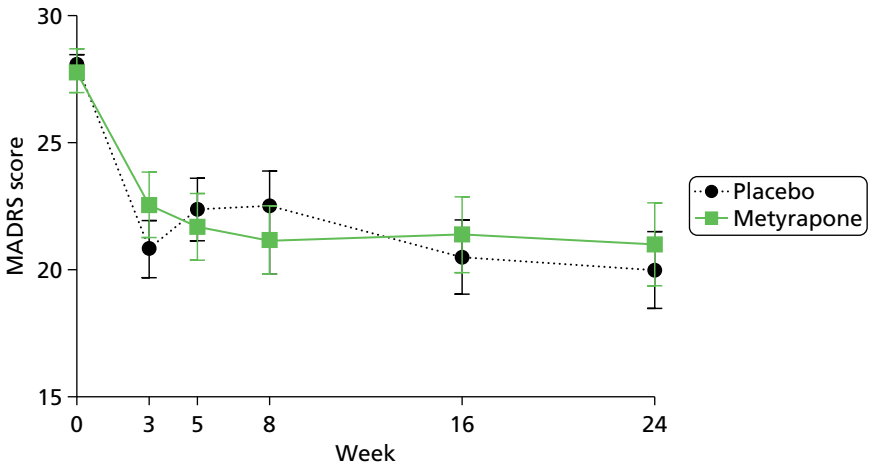


FIGURE 4 Montgomery–Åsberg Depression Research Scale scores over time for patients treated with either metyrapone or placebo. Data plotted as mean, with error bars representing the SEM.

Examination of the summary statistics and *Figure 4* suggests that there is a reduction in the mean scores at all of the follow-up visits compared with baseline, but that there is little difference at any time point in the groups randomised to metyrapone or placebo. The issue of whether or not this result may be effected by differential dropout between the groups is discussed below (see *Ancillary analyses*) and later in this chapter (see *Figures 11* and *12*).

Primary outcome analysis The prespecified primary analysis was an ANCOVA of week +5 MADRS scores, with baseline scores included as a covariate, adjusted for differences between randomisation strata (study site and primary or secondary care origin of the patients). The difference between groups was estimated on an intention-to-treat basis, with patients being included in the group to which they were randomised.

Fitting a simple regression model, including only baseline MADRS score and a binary indicator of group to which randomised, yielded the following regression coefficients:

- baseline MADRS: 0.88 (95% CI 0.63 to 1.12)
- difference between groups (metyrapone–placebo): –0.43 (95% CI –3.48 to 2.63).

The MADRS score at 5 weeks was highly correlated with the score at baseline; the difference between groups at 5 weeks did not differ significantly from zero.

Variation between sites was then included as a random effect, and the difference between patients recruited from primary care and those recruited from secondary care was added as a fixed effect. There was little change to the estimated effect of metyrapone; the estimated difference between groups, based on the adjusted model, was –0.51 (95% CI –3.48 to 2.46).

Further analysis of Montgomery–Åsberg Depression Research Scale data

Change in MADRS scores between baseline and week +5: a prespecified secondary analysis was to calculate the change in MADRS scores from baseline to week +5 for each patient and to compare group means using an independent sample *t*-test. Summary statistics for baseline, week +5 and change scores are given in *Table 8*.

There was a significant reduction in MADRS scores of approximately 6 points between baseline and week +5. The difference between the changes in the metyrapone group and the change in the placebo group was = –0.39 (95% CI –3.37 to 2.67) and did not differ significantly from zero ($t_{141} = 0.252$; $p = 0.80$). The estimated effect of metyrapone is very similar to that obtained from the ANCOVA.

TABLE 8 Mean MADRS scores at baseline and week +5 by group to which randomised

	Baseline		Week 5		Reduction	
Group[a]	Mean	95% CI	Mean	95% CI	Mean	95% CI
Placebo	28.2	26.9 to 29.4	22.4	20.1 to 24.8	5.8	3.7 to 7.9
Metyrapone	27.9	26.2 to 29.6	21.7	19.2 to 24.4	6.2	4.0 to 8.3
All	28.0	27.02 to 29.0	22.1	20.3 to 23.8	6.0	4.5 to 7.5

a Statistics are based on complete data on $n = 74$ (placebo group) and $n = 69$ (metyrapone group).

Analysis using atypical Montgomery–Åsberg Depression Research Scale scores

These were calculated using the highest score from items 4a and 4b and items 5a and 5b on the MADRS. The mean atypical MADRS score at baseline in the metyrapone group was 28.5 (SD 6.7) and 28.9 (SD 5.4) in the placebo group. For both groups there was a significant reduction in atypical MADRS scores of approximately '7' between baseline and week +5. The difference between the changes in atypical MADRS scores in the metyrapone group and the change in the placebo group was –0.42 (95% CI –3.36 to 2.63) and did not differ significantly from zero ($t_{141} = 0.275$; $p = 0.79$).

Persistence of change in Montgomery–Åsberg Depression Research Scale scores

For each individual, the MADRS score was recorded on up to five occasions following randomisation. These repeated measures were analysed using a mixed model in which we assume that for each patient the MADRS scores varied randomly about an individual mean score (with SD σ_e), but that this mean varied randomly across patients (with SD σ_u). A normal distribution was assumed for each of the random effects. Fitting an initial model with only a constant term (corresponding to the assumption that the mean MADRS score is the same at all five follow-up visits) in addition to the random effects, the estimated mean MADRS score across the five follow-up visits was 21.6 (95% CI 20.1 to 23.0). There was significant variation of scores both within patients [$\sigma_e = 6.57$ (95% CI 6.17 to 6.99)] and between patients [$\sigma_u = 8.58$ (95% CI 7.54 to 9.78)].

As we are interested primarily in changes in depression, the second step was to include the baseline MADRS score as a covariate. This model indicated a significant reduction in depression from baseline at each of the five follow-up visits. Baseline depression explained some of the variation between individual patients (the estimate of σ_u fell to 6.69) but there were still significant differences between patients.

Adding in differences between randomisation strata (sites and origin of patient – either primary or secondary care) did not explain very much additional variability. With the baseline score as a covariate there were no significant differences between these strata. The estimated difference in depression across all five follow-up visits between patients on metyrapone and patients on placebo was 0.75 (95% CI –1.59 to 3.10).

There was no evidence of any trend in depression over the follow-up visits. Adding a linear trend over visits to the model, the estimated mean change in MADRS score between consecutive visits was –0.27 (95% CI –0.65 to 0.10) and did not constitute a clinically important change in depression.

One of the objectives of the ADD Study was to ascertain whether or not the effect of metyrapone persists over time. Given that there is no obvious impact of the drug at 2 weeks post treatment, this objective is moot. However, fitting an interaction between the effect of metyrapone and time demonstrates that the interaction is not significant; there is no evidence that changes in depression over time differ between treatment groups.

Response rates

Response was defined as a reduction of $\geq 50\%$ in MADRS score. The proportion of responders at each time point is shown in *Table 9*.

At each visit the proportion of patients who reported a MADRS that was equal to or less than half their baseline score was between 19% and 26%. At the primary end point of 5 weeks post randomisation, 21% of patients were deemed to be responders. Responder rates for each trial arm at 5 weeks are given in *Table 10*.

TABLE 9 Number and percentage of patients who have responded to treatment by visit

Responder?		Visit				
		+3 weeks	+5 weeks	+8 weeks	+16 weeks	+24 weeks
No	*n*	111	113	104	88	77
	%[a]	74.5	79.0	81.3	76.5	74.0
Yes	*n*	38	30	24	27	27
	%[a]	25.5	21.0	18.8	23.5	26.0
Total	*n*	149	143	128	115	104
	%[a]	100.0	100.0	100.0	100.0	100.0

a Column percentages.

TABLE 10 Proportion of patients responding to therapy by group to which randomised

Therapy		Responder?		
		No	Yes	Total
Placebo	*n*	58	16	74
	%[a]	78.4	21.6	100.0
Metyrapone	*n*	55	14	69
	%[a]	79.7	20.3	100.0
Total	*n*	113	30	143
	%[a]	79.0	21.0	100.0

a Row percentages.

Response rates were almost identical in the two groups at this time point. The odds ratio (based on a logistic regression model adjusted for site and origin of patient: metyrapone/placebo) = 0.95 (95% CI 0.41 to 2.20). The difference between groups was not clinically or statistically significant but the wide CI (the odds could be less than half or more than double) suggests that the study was not adequately powered to detect differences in this binary measure of outcome.

Remission rates

A patient was defined as being in remission if their MADRS score was ≤ 10. Remission rates broken down by visit are shown in *Table 11*.

The proportions of patients in remission appear to be similar at each visit. The proportion of patients in remission 5 weeks post randomisation by study group are shown in *Table 12*.

Remission rates at this time point were similar in the two groups. The odds ratio (based on a logistic regression model adjusted for site and origin of patient: metyrapone/placebo = 0.97 (95% CI 0.40 to 2.55). Again, the width of the CI is large (the odds of 'success' in one group could be less than half or more than double that in the other).

TABLE 11 Proportion of patients with a MADRS score of ≤ 10 by visit

Patient in remission?		Visit				
		+3 weeks	+5 weeks	+8 weeks	+16 weeks	+24 weeks
No	*n*	127	120	110	93	80
	%	85.2	83.9	85.9	80.9	76.9
Yes	*n*	22	23	18	22	24
	%	14.8	16.1	14.1	19.1	23.1
Total	*n*	149	143	128	115	104
	%	100.0	100.0	100.0	100.0	100.0

TABLE 12 Proportion of patients in remission 5 weeks post randomisation by group to which randomised

Treatment		Patient in remission?		
		No	Yes	Total
Placebo	*n*	62	12	74
	%	83.8	16.2	100.0
Metyrapone	*n*	58	11	69
	%	84.1	15.9	100.0
Total	*n*	120	23	143
	%	83.9	16.1	100.0

Suicidal thoughts

During the trial, the AE of increased suicidality was reported (as was to be expected in such a population). The DMEC looked at this AE with the patients divided into two – still blinded – groups. DMEC recommended no change to the conduct of the trial and that item 10 on the MADRS ('Suicidal thoughts') was examined in the two groups during analysis, and this was incorporated into the SAP.

The prespecified analysis was an ANCOVA of week +5 suicidality score, with baseline score included as a covariate, adjusted for differences between randomisation strata. The difference between groups was estimated on an intention-to-treat basis, with patients being included in the group to which they were randomised.

Fitting a simple regression model including only baseline MADRS score and a binary indicator of group to which randomised yielded the following regression coefficients:

- baseline MADRS item 10 score: 0.73 (95% CI 0.59 to 0.87)
- difference between groups (metyrapone – placebo): –0.13 (95% CI –0.52 to 0.26).

Suicidality score at 5 weeks was highly correlated with the score at baseline; the effect of metyrapone did not differ significantly from zero.

Variation between sites and the difference between patients recruited from primary care and patients recruited from secondary care were then added as fixed effects. There was little change to the estimated effect of metyrapone; the estimated difference between groups, based on the adjusted model, was –0.15 (95% CI –0.53 to 0.24).

The estimated impact of metyrapone on suicidality score (as measured by item 10 on the MADRS) was a change of –0.15 (95% CI –0.53 and 0.24).

Secondary outcome measures

Beck Depression Inventory

Baseline data

Similar to the MADRS data, baseline BDI scores were (non-significantly) higher for patients from Manchester (mean 37.6, SD 9.6) than patients from Leeds/Bradford (mean 33.8, SD 10.3) and Newcastle/Durham (mean 33.8, SD 12.1). This is consistent with the observation that baseline scores were higher for patients recruited from secondary care (mean 37.5, SD 10.5) than for those recruited from primary care (mean 32.1, SD 9.9), as a greater proportion of patients were recruited from secondary care in Manchester than at the other two centres. The baseline difference between patients originating from primary and secondary care was statistically significant (–5.4, 95% CI –8.5 to –2.3). Mean and SD of BDI scores for each randomised group are reported at baseline in *Table 5*.

Effect of metyrapone on Beck Depression Inventory scores 5 weeks post randomisation

As defined in the SAP, scores obtained 5 weeks post randomisation were analysed using ANCOVA, with baseline scores included as a covariate. This suggested a difference of –2.65 (95% CI –6.53 to 1.23) between the effects of metyrapone and placebo. Adjusting for random variation between centres and a difference between patients originating from primary and secondary care gave an adjusted estimate of –2.65 (95% CI –6.41 to 1.10), which was not significantly different from zero. The adjusted estimate differs very little from the unadjusted one. Although not significant, the direction of effect is consistent with that hypothesised. However, the estimated magnitude of the effect (–2.65) was small in comparison with the average change between baseline and 5 weeks post randomisation of 6.14 (95% CI 4.19 to 8.09) in the two groups combined.

Persistence of effect

Given that the estimated impact of metyrapone at 5 weeks was not statistically significant, it was not felt appropriate to undertake an analysis to assess the persistence of the effect over time. BDI scores over time are plotted in *Figure 5*.

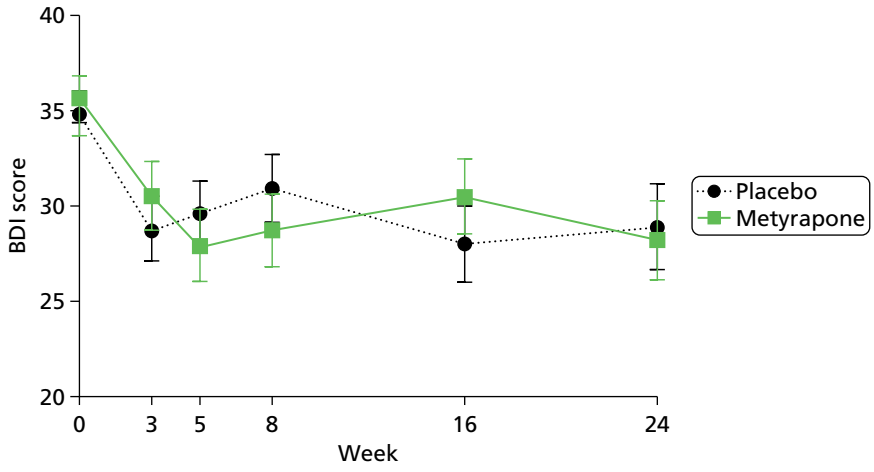


FIGURE 5 Beck Depression Inventory scores over time for patients treated with either metyrapone or placebo. Data plotted as mean, with error bars representing the SEM.

Clinical Anxiety Scale

Baseline data

The mean baseline CAS score for patients recruited at each centre were as follows: Manchester – mean 10.4, SD 4.0; Leeds/Bradford – mean 9.4, SD 6.2; and Newcastle/Durham – mean 9.5, SD 3.8. These differences were not statistically significant. The anxiety scores for patients originating from primary care (mean 9.5, SD 4.6) were similar to those originating from secondary care (mean 10.0, SD 4.5); the difference in means was 0.5 (95% CI –1.0 to 2.0), which did not differ significantly from zero.

Effect of metyrapone on Clinical Anxiety Scale scores 5 weeks post randomisation

Based on ANCOVA of CAS scores recorded 5 weeks post randomisation (with baseline scores included as a covariate) the unadjusted estimate of the effect of metyrapone was 0.55 (95% CI –1.21 to 2.30). Adjusting for differences between centres and patient origin of care the estimated effect of metyrapone on CAS scores was 0.46 (95% CI –1.20 to 2.12). These estimates correspond to differences between groups that are neither statistically nor clinically significant.

Persistence of effect

Lower anxiety levels were recorded at all follow-up visits compared with those observed at baseline. However, there was very little difference between the groups at any time point. Given the lack of a difference between metyrapone and placebo at week +5, no further analysis was conducted. CAS scores over time are plotted in *Figure 6*.

State–Trait Anxiety Inventory

The STAI comprises two scales: state anxiety and trait anxiety. Because the study investigated the effect of treatment with metyrapone over the short term, only state anxiety is described here. This scale evaluates the current state of anxiety, asking how respondents feel 'right now,' using items that measure subjective feelings of apprehension, tension, nervousness, worry, and activation/arousal of the autonomic nervous system.

Baseline data

The mean baseline state anxiety score for patients recruited at each centre were as follows: Manchester – mean 43.1, SD 6.9; Leeds/Bradford – mean 42.5, SD 6.6; and Newcastle Durham – mean 40.7, SD 5.1.

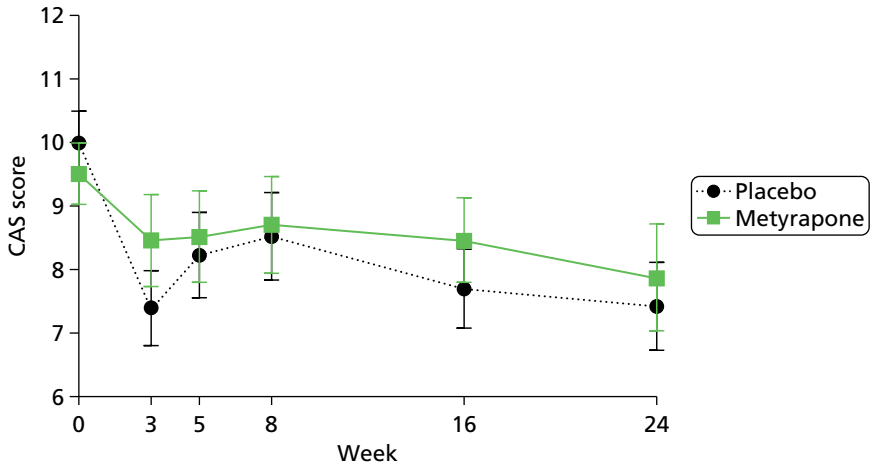


FIGURE 6 Clinical Anxiety Scale scores over time for patients treated with either metyrapone or placebo. Data plotted as mean with error bars representing the SEM.

These differences were not statistically significant. The state anxiety scores for patients originating from primary care (mean 41.4, SD 5.4) were similar to those originating from secondary care (mean 42.4, SD 6.8); the difference in means was 1.0 (95% CI –0.8 to 3.0), which did not differ significantly from zero.

Effect of metyrapone on state anxiety scores 5 weeks post randomisation

Based on an ANCOVA, the estimated effect of metyrapone relative to placebo at week +5 was an increase in state anxiety of 1.2 (95% CI –0.7 to 3.1). This is consistent with the larger fall in anxiety between baseline and week +5 in the group randomised to metyrapone than in the other group that can be seen in *Figure 7*. However, the difference between groups is not statistically significant ($p = 0.21$). Adjusting for origin of patient care and differences between centres resulted in very little change in the estimated impact; adjusted estimate was an increase in anxiety of 1.2 (95% CI –0.6 to 3.0).

Persistence of effect of metyrapone on State–Trait Anxiety Inventory scores

Given that there is no obvious effect on state anxiety 5 weeks post randomisation, persistence was not formally assessed. However, it is clear from *Figure 7* that there was very little change in anxiety levels in either group of the period of follow-up.

EuroQol health tariff

Baseline data

The EQ-5D comprises five questions about different aspects of QoL, each with three response options. Responses to these questions are converted into a single health tariff on a scale where '1' corresponds to perfect health and '0' corresponds to being dead.[87]

Patients from Newcastle/Durham reported slightly higher QoL (mean 0.41, SD 0.30) than patients from Leeds/Bradford (mean 0.34, SD 0.31) and Manchester (mean 0.34, SD 0.31). Patients originating from primary care had a higher mean tariff (mean 0.40, SD 0.30) than patients originating from secondary care (mean 0.35, SD 0.31); however, the difference of 0.06 (95% CI –0.02 to 0.16) does not differ significantly from zero ($p = 0.17$).

Effect of metyrapone on EQ-5D health tariffs 5 weeks post randomisation

Quality-of-life scores appear to improve in both groups between baseline and the visit 5 weeks post randomisation (*Figure 8*).

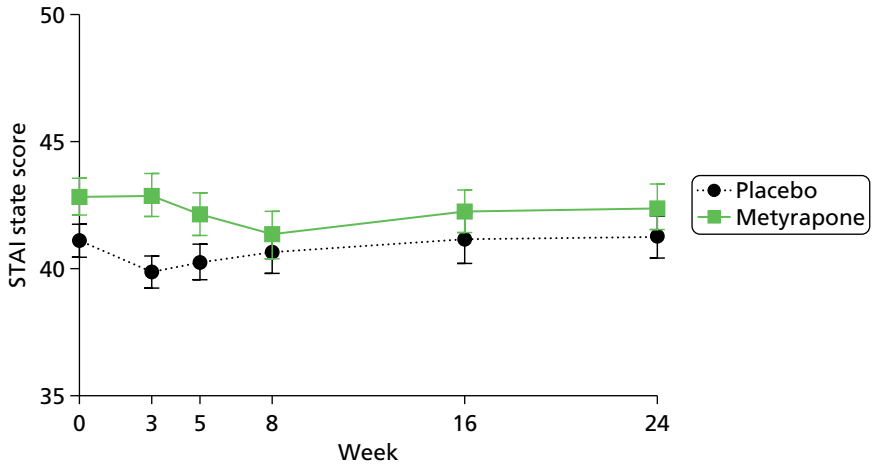


FIGURE 7 State anxiety scores over time for patients treated with either metyrapone or placebo. Data plotted as mean with error bars representing the SEM.

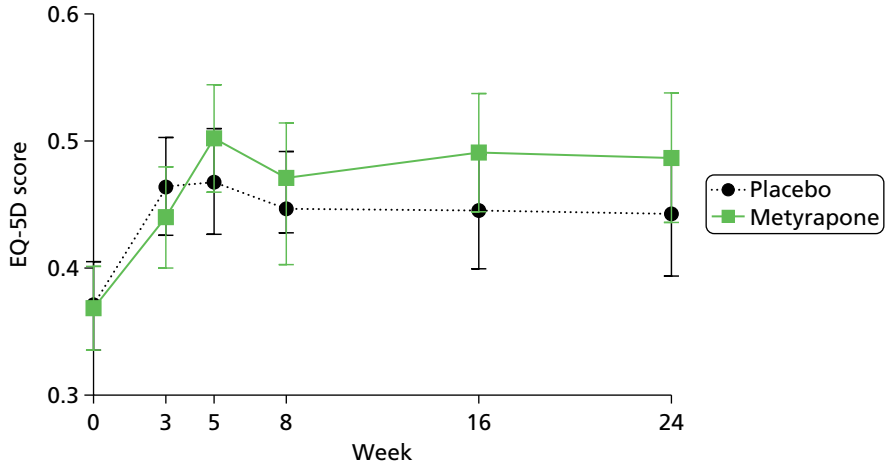


FIGURE 8 European Quality of Life-5 Dimensions tariffs over time for patients treated with either metyrapone or placebo. Data plotted as mean with error bars representing the SEM.

Based on a simple ANCOVA model with baseline tariff value included as a covariate, the estimated effect of metyrapone was a change in tariff value of 0.014 (95% CI –0.073 to 0.101). Adjusting for variation between sites and origin of patient care, the estimated effect of metyrapone was 0.015 (95% CI –0.069 to 0.099). There was no evidence of a beneficial effect of treatment with metyrapone on EQ-5D QoL tariff value.

EuroQol visual analogue scale

Baseline data

The EuroQol visual analogue scale (EQ-VAS) records the respondent's self-rated health on a 20-cm vertical, visual analogue scale with end points labelled 'the best health you can imagine' and 'the worst health you can imagine'. This information can be used as a quantitative measure of health, as judged by the individual respondents. The response is recorded as a number between '0' and '100', with higher scores indicating better health.

The EQ-VAS scores were higher for patients in Newcastle/Durham (mean 42.4, SD 19.4) and Leeds/Bradford (mean 44.9, SD 17.0) than for patients recruited in Manchester (mean 38.8, SD 16.8). Consistent with this pattern, patients recruited from primary care reported better health (mean 44.9, SD 19.7) than patients recruited from secondary care (mean 39.2, SD 16.4), although the mean difference (5.70; 95% CI –0.07 to 11.3) was of only borderline significance ($p = 0.053$).

Effect of metyrapone on EuroQol visual analogue scale at 5 weeks post randomisation

The EQ-VAS scores appear to improve in both groups between baseline and the visit 3 weeks post randomisation (*Figure 9*).

The estimated effect of metyrapone at 5 weeks post randomisation (ANCOVA with baseline scores included as a covariate) was 5.7 (95% CI –0.8 to 12.1). Adjusting for origin of patient and variation between centres, the adjusted estimate is 5.6 (95% CI –0.7 to 12.0).

Persistence of effect

It is interesting to observe that the largest difference between groups was observed at week +5. It is likely that this is a chance result, with the difference between groups being very much smaller at subsequent visits.

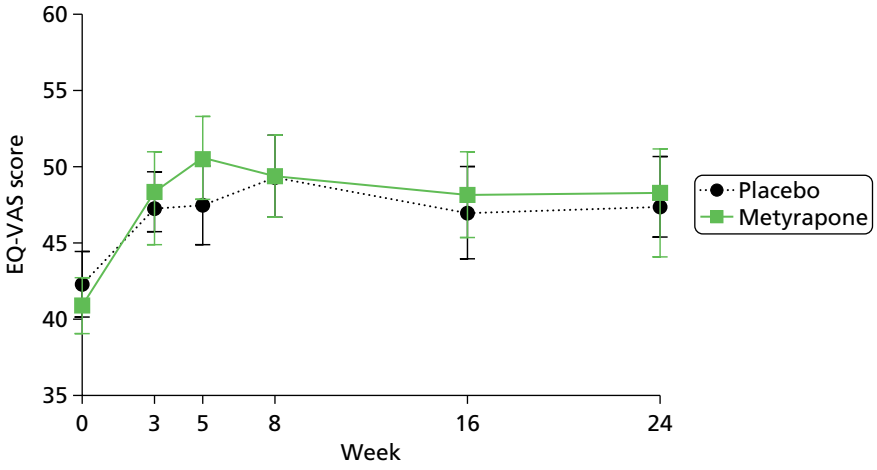


FIGURE 9 EuroQol visual analogue scale scores over time for patients treated with either metyrapone or placebo. Data plotted as mean with error bars representing the SEM.

Young Mania Rating Scale

Baseline data

The YMRS is used to assess manic symptoms, with a score of '0' being indicative of being symptom free. Mean scores at baseline were similar in the three centres. In Leeds/Bradford the mean score was 2.1 (SD 1.8), in Manchester the mean score was 2.4 (SD 1.4), and in Newcastle/Durham the mean score was 2.5 (SD 2.0). The difference in mean scores between patients recruited from primary and secondary care was 0.1 (95% CI –0.4 to 0.7), a non-significant difference.

Effect of metyrapone on Young Mania Rating Scale scores

Mean and SD YMRS scores for each randomised group are reported for all visits (*Figure 10*).

The estimated difference between patients randomised to metyrapone and placebo at 5 weeks post randomisation (ANCOVA of week +5 scores with baseline scores included as a covariate) was –0.04 (95% CI –0.54 to 0.46). Adjusting for origin of patient and variation between centres resulted in almost no change to the estimate; the adjusted estimate is –0.04 (95% CI –0.52 to 0.45). There is no evidence of different levels of manic symptoms between patients randomised to metyrapone and placebo.

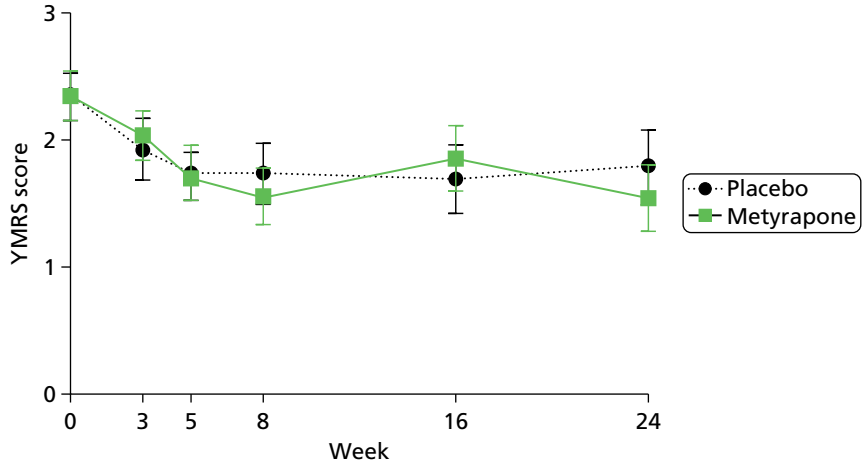


FIGURE 10 Young Mania Rating Scale scores over time for patients treated with either metyrapone or placebo. Data plotted as mean with error bars representing the SEM.

Ancillary analyses

Missing data

There may be some suggestion that patients randomised to metyrapone were less inclined to attend follow-up visits than patients randomised to placebo. This is shown in *Figure 11*.

Fitting a Cox proportional hazards model, the estimated hazard ratio was 0.57 (95% CI 0.35 to 0.93); patients randomised to metyrapone were less likely to be retained in the study than other patients. It is not clear whether this is a chance finding (we were unlucky with the randomisation) or a genuine effect of taking the active drug.

In order to investigate whether or not retention was related to baseline depression, the baseline MADRS score was included in the Cox regression model. The hazard ratio corresponding to a unit increase in MADRS score was 1.01 (95% CI 0.97 to 1.04). This suggest that time to drop out was not related to baseline depression.

To further explore the reasons for non-attendance at follow-up visits, the relationship between change in symptoms between baseline and week +3 (the period of treatment with metyrapone or placebo) and attendance at the primary outcome time point at week +5 was examined. *Figure 12* shows the change in MADRS scores between baseline and 3 weeks for patients who did or did not attend at week +5.

The premise behind joint modelling is that whether or not a patient responds is influenced by outcome; specifically patients with poorer outcomes are less likely to respond than other patients. The expectation would be that the lines in the upper half of *Figure 12* (corresponding to non-responders at week +5) would have a greater downwards gradient than those in the lower plot (responders at week +5). There is very little evidence of such a pattern either from inspection of the plot or examination of the mean changes in depression from baseline to the visit at week +3, which are shown in *Table 13*.

Those who attended at week +5 had a slightly greater reduction in depression (mean change = 6.5) than those who failed to attend (mean change = 4.8). The difference in these changes was 1.7 (95% CI –4.0 to 7.5). Although not statistically significant, the direction of effect is as hypothesised – there is a tendency for patients who drop out to be those with worse outcomes. In practice, dropout rates were very similar in

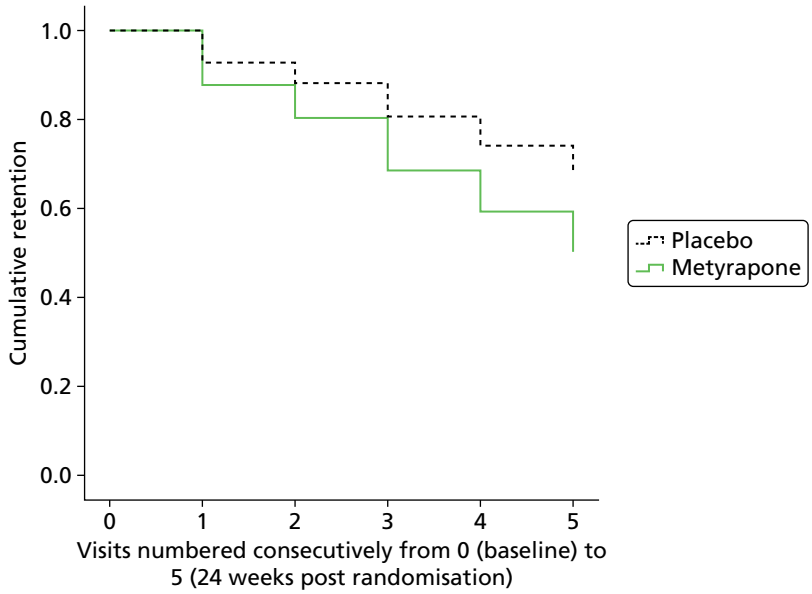


FIGURE 11 Cumulative survival (or retention) against visit number (visits at which MADRS was recorded are numbered sequentially).

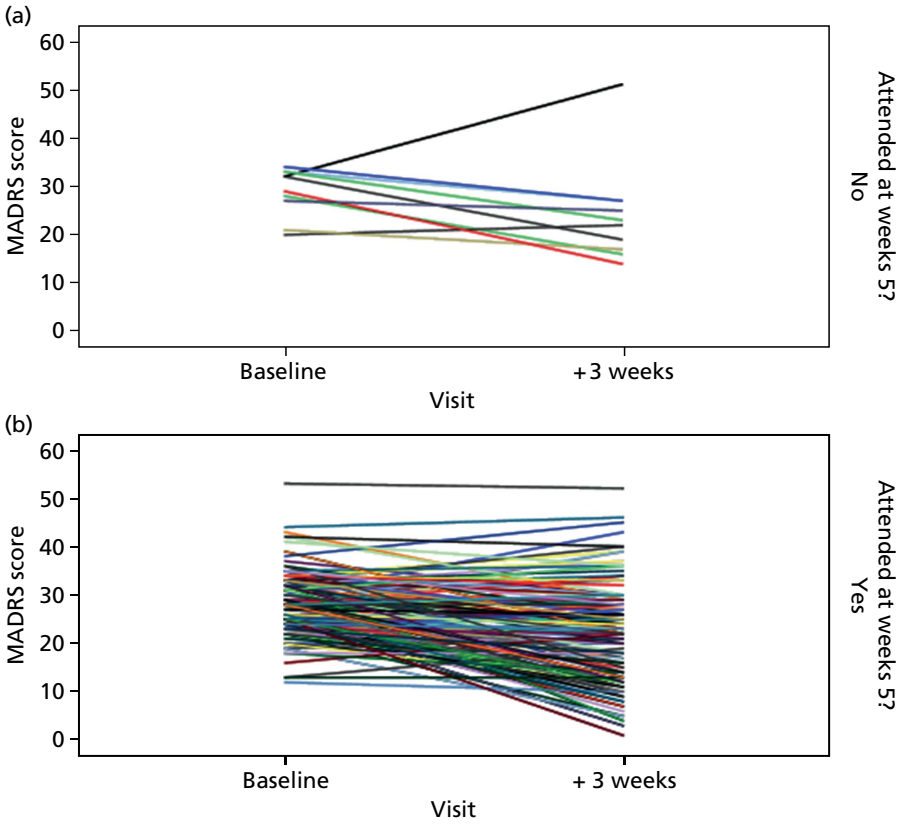


FIGURE 12 Change in depression between baseline and 3 weeks post randomisation by whether or not the patient attended the primary end point 5 weeks post randomisation. Lines correspond with individual patients.

TABLE 13 Reduction in depression from baseline to week +3

	Reduction in depression baseline to week +3			
			95% confidence limits of the difference	
Status at week 5	Mean	SD	Lower	Upper
Non-attenders	4.80	9.88	–2.26	11.86
Attenders	6.54	8.86	5.05	8.03

the two groups and if anything were higher in patients randomised to metyrapone. Thus any adjustment to the estimated impact of metyrapone based on joint modelling would be to further reduce the size of the estimated reduction in depression. There is no evidence from the joint modelling to suggest a clinically beneficial effect of metyrapone on depression.

Mechanistic outcomes

Hypothalamic–pituitary–adrenal axis function

Hypothalamic–pituitary–adrenal axis function was determined in order to gain an understanding of the mechanism of action of the drug. Salivary cortisol was measured at 11 p.m. and then on five occasions, 15 minutes apart from wakening, on the following day, as per the standard protocol for the CAR. Patients collected such samples at baseline, immediately after cessation of treatment and 2 weeks later. Healthy control subjects collected samples at baseline. Areas under the curve [area under the curve ground (AUCg) and area under the curve increase (AUCi)] were determined using trapezoidal method for the morning cortisols.[117] The control subjects and patients did not differ in their cortisol parameters (*Figure 13* and *Table 14*).

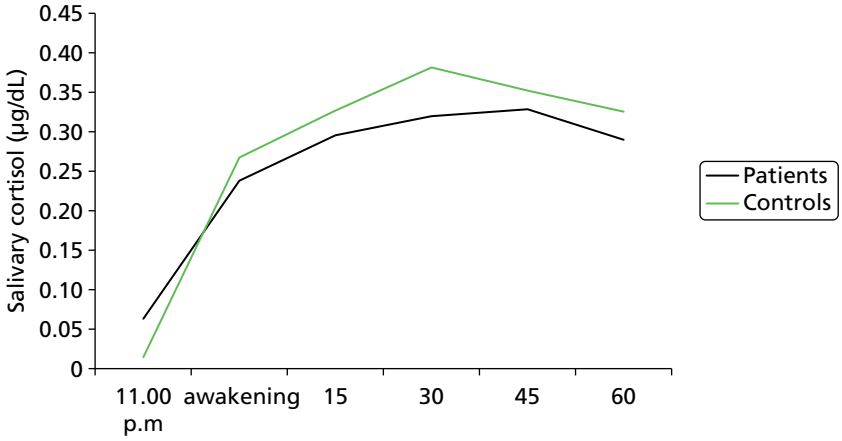


FIGURE 13 Salivary cortisol (µg/dl) at 11 p.m. and on awakening the following morning, and at 15-minute intervals for 60 minutes thereafter in the two groups.

TABLE 14 Hypothalamic–pituitary–adrenal axis results in the two groups. No significant differences

Measures of cortisol	Control subjects	Patients	Significance[a]
11 p.m.	Mean = 0.042 µg/dl, $n = 47$	Mean = 0.064 µg/dl, $n = 132$	$p = 0.084$
AUCi	Mean = 0.274, $n = 45$	Mean = 0.256, $n = 114$	$p = 0.523$
AUCg	Mean = 1.34, $n = 45$	Mean = 1.21, $n = 114$	$p = 0.848$

a Independent samples Mann–Whitney *U*-test.

Baseline AUCg did not impact on clinical response to treatment. This was analysed using repeated measures ANCOVA with 'visit' as the within-subject factor (with MADRS score at week +3 and MADRS score at week +5 as the two levels), 'treatment' (metyrapone and placebo) as the between-subject factor, and baseline MADRS score and baseline AUCg as the covariates ($F_{1, 93} = 1.275$ df; $p = 0.26$).

Metyrapone did not significantly impact cortisol in the short or longer term. Change in HPA axis function (measured using δAUCg, i.e. the difference between AUCg at baseline and AUCg at week +3) did not impact on clinical response to treatment. This was analysed using repeated measures ANCOVA with 'visit' as the within-subject factor (with MADRS score at week +3 and MADRS score at week +5 as the two levels), 'treatment' (metyrapone and placebo) as the between-subject factor and baseline MADRS score and δAUCg as the covariates ($F_{1, 81} < 1$; not significant).

Childhood trauma

The CTQ revealed that patients had experienced more childhood adversity than control subjects (*Table 15*).

TABLE 15 Scores for individual items of CTQ in patients and control subjects

Subscale	Patients	Control subjects	Significance
Emotional abuse	12.3	6.7	$p < 0.0005$
Emotional neglect	13.6	8	$p < 0.0005$
Sexual abuse	8.1	5.5	$p = 0.004$
Physical abuse	7.9	5.4	$p < 0.0005$
Physical neglect	8.6	5.8	$p < 0.0005$

Emotional abuse, emotional neglect, physical abuse, physical neglect and sexual abuse did not predict response to medication (analysed using repeated measures ANOVA with these subscores on the CTQ entered as covariates: data not shown).

Neuropsychology

Spatial working memory

Groups at baseline

Valid data were available for 55 HVs and 65 depressed patients.

For BSE there was a significant main effect of group ($F_{1,116} = 8.609$; $p = 0.004$), with patients making more errors than control subjects (*Table 16*). There was also a significant group × level interaction ($F_{2,232} = 4.978$; $p = 0.014$). Comparison of patients and control subjects at each separate level of difficulty indicated significantly poorer performance in patients at levels 6 ($p = 0.001$) and 8 ($p = 0.004$). There was a non-significant trend towards poorer performance in patients at level 4 ($p = 0.055$). For WSE there was no significant main effect of group ($F_{1,116} = 0.332$; $p = 0.566$) and no significant group by level interaction ($F_{2,232} = 0.015$; $p = 0.950$).

Treatment (patients)

Valid data were available for 43 patients (20 metyrapone, 23 placebo) tested at baseline and at 5 weeks. This analysis was conducted on totals for both the BSE and WSE. Results of the ANCOVA are presented in *Table 17*. There were no significant main effects of treatment for either outcome measure.

TABLE 16 Between-search errors and WSE for each group

Outcome	Control subjects: adjusted mean[a] (SEM)	Patients: adjusted mean[a] (SEM)
BSE total[b]	22.21 (2.63)	34.83 (2.54)
Four box	0.65 (0.22)	1.42 (0.33)
Six box	5.37 (0.93)	10.23 (1.04)
Eight box	16.19 (1.83)	23.18 (1.52)
WSE total[a]	1.42 (0.38)	1.74 (0.47)
Four box	0.04 (0.02)	0.12 (0.05)
Six box	0.23 (0.08)	0.42 (0.16)
Eight box	1.16 (0.34)	1.20 (0.35)

a Estimated marginal mean from ANCOVA model.
b Data from additional $n = 2$ control subjects available which were omitted from ANCOVA owing to missing age/gender data.

TABLE 17 Spatial working memory performance at week +5 (covaried for baseline only) with ANCOVA main effect of group

Outcome	Placebo: adjusted mean[a] (SEM)	Metyrapone: adjusted mean[a] (SEM)	$F_{1,40}$	*p*-value
BSE total	26.13 (2.58)	24.70 (2.76)	0.143	0.707
WSE total	0.85 (0.35)	1.67 (0.37)	2.546	0.118

a Estimated marginal mean from ANCOVA model.

Attentional Network Test

Groups at baseline

Valid data were available for 54 HVs and 65 depressed patients. There was a significant trend in alerting, no significant effect for orientating and a significant main effect of conflict (*Table 18*).

For the conflict effect, examination of the RTs that comprise this measure indicate that patients were slower to respond overall but were slowed to a greater extent by incongruent flankers than were healthy control subjects.

Treatment (patients)

Valid data were available for 38 patients (17 metyrapone, 21 placebo) tested at baseline and at 5 weeks. This analysis was conducted on the alerting, orientating and conflict indices. Results of the ANCOVA are presented in *Table 19*. There were no significant main effects of treatment on any ANT measure.

Object-location memory

Groups at baseline

Valid data were available for 56 HVs and 64 depressed patients. Patients were found to have significantly poorer performance on all three primary outcome measures than control subjects (*Table 20*).

Treatment (patients only)

Valid data were available for 38 patients (16 metyrapone, 22 placebo) tested at baseline and at 5 weeks. This analysis was conducted on the POM, OLB and COM. Results of the ANCOVA are presented in *Table 21*.

TABLE 18 Attentional Network Test performance for each group (covaried for age and sex)

Outcome	Control subjects: adjusted mean[a] (SEM)	Patients: adjusted mean[a] (SEM)	$F_{1,115}$	*p*-value
Alerting	28.43 (4.25)	16.90 (3.86)	3.882	0.051
Orienting	33.39 (4.09)	35.17 (3.71)	0.099	0.753
Conflict	107.90 (8.86)	155.87 (8.05)	15.459	< 0.001

a Estimated marginal mean from ANCOVA model.

TABLE 19 Attentional Network Test performance at week +5 (covaried for baseline)

Outcome	Placebo: adjusted mean[a] (SEM)	Metyrapone: adjusted mean[a] (SEM)	$F_{1,35}$	*p*-value
Alerting	27.09 (7.37)	43.77 (8.20)	2.269	0.141
Orienting	28.37 (6.80)	37.66 (7.56)	0.834	0.367
Conflict	133.84 (12.12)	129.03 (13.47)	0.070	0.792

a Estimated marginal mean from ANCOVA model.

TABLE 20 Object-location memory performance for each group (covaried for age and sex)

Outcome	Control subjects: adjusted mean[a] (SEM)	Patients: adjusted mean[a] (SEM)	$F_{1,116}$	*p*-value
Object memory (%)	2.64 (0.81)	5.11 (0.75)	4.802	0.030
Visuospatial reconstruction (mm)	107.50 (9.03)	95.68 (8.42)	0.879	0.351
POM (mm)	142.13 (6.80)	165.94 (6.34)	6.287	0.014
OLB (%)	20.47 (3.33)	32.17 (3.10)	6.354	0.013
COM (mm)	237.23 (17.92)	288.43 (16.71)	4.189	0.043

a Estimated marginal mean from ANCOVA model.

Accuracy

TABLE 21 Object-location memory performance at week +5 (covaried for baseline)

Outcome	Placebo: adjusted mean[a] (SEM)	Metyrapone: adjusted mean[a] (SEM)	$F_{1,35}$	*p*-value
POM (mm)	152.89 (8.34)	162.09 (9.79)	0.510	0.480
OLB (%)	28.47 (4.09)	28.04 (4.80)	0.005	0.946
COM (mm)	284.90 (21.81)	311.08 (25.71)	0.583	0.450

a Estimated marginal mean from ANCOVA model.

Digit span

Groups at baseline

Valid data were available for 56 HVs and 70 depressed patients. Patients performed more poorly than control subjects on all outcome measures (*Table 22*).

TABLE 22 Digit span performance for each group (covaried for age and sex)

Outcome	Control subjects: adjusted mean[a] (SEM)	Patients: adjusted mean[a] (SEM)	$F_{1,122}$	*p*-value
Experimental forwards	9.10 (0.33)	7.60 (0.30)	10.917	0.001
Backwards	7.73 (0.33)	6.56 (0.29)	6.860	0.010
Total	16.83 (0.60)	14.17 (0.53)	10.695	0.001
Span forwards	7.18 (0.20)	6.40 (0.18)	8.360	0.005
Backwards	5.43 (0.20)	4.79 (0.18)	5.588	0.020

a Estimated marginal mean from ANCOVA model.

Treatment (patients only)

Valid data were available for 45 patients (22 metyrapone, 23 placebo) tested at baseline and at 5 weeks. This analysis was conducted on all indices. Results of the ANCOVA are presented in *Table 23*. There were no significant main effects of treatment on any measure.

Face Emotion Recognition Task

Groups at baseline

Valid data were available for only 43 HVs and 39 depressed patients as a result of technical problems with the version of the task initially used.

Accuracy

For accuracy there was a significant effect of emotion type ($F_{6,468} = 5.777$; $p < 0.001$) and group ($F_{1,78} = 6.179$; $p = 0.015$) but no 'emotion × group' interaction ($F_{6,468} = 1.245$; $p = 0.39$).

Happiness and neutral face emotion were the best recognised, and anger and fear the worst. Patients were less accurate overall, but on post hoc testing the difference was statistically significant only for disgust with a trend for anger (*Table 24*).

TABLE 23 Digit span performance at week +5 (covaried for baseline)

Outcome	Placebo: adjusted mean[a] (SEM)	Metyrapone: adjusted mean[a] (SEM)	$F_{1,42}$	*p*-value
Experimental forwards	8.05 (0.29)	8.04 (0.30)	0.001	0.970
Backwards	6.60 (0.49)	7.19 (0.50)	0.724	0.400
Total	14.65 (0.67)	15.23 (0.69)	0.356	0.554
Span forwards	6.50 (0.20)	6.75 (0.21)	0.705	0.406
Backwards	4.83 (0.30)	5.00 (0.31)	0.164	0.688

a Estimated marginal mean from ANCOVA model.

TABLE 24 Accuracy (hit rate) by emotion and group (covaried for age and sex)

Emotion	Control subjects: adjusted mean[a] (SEM)	Patients: adjusted mean[a] (SEM)	$F_{1,78}$	*p*-value
All emotions	0.61 (0.02)	0.56 (0.02)	6.179	0.02
Anger	0.37 (0.03)	0.28 (0.03)	3.838	0.05
Disgust	0.58 (0.03)	0.46 (0.03)	7.599	0.01
Fear	0.35 (0.03)	0.36 (0.03)	0.024	0.88
Happiness	0.80 (0.02)	0.76 (0.03)	1.774	0.19
Sadness	0.61 (0.04)	0.57 (0.04)	0.513	0.48
Surprise	0.71 (0.03)	0.67 (0.03)	0.733	0.39
Neutral	0.86 (0.02)	0.81 (0.03)	1.756	0.19

a Estimated marginal mean from ANCOVA model.

Misattribution

For misattribution there was a significant effect of emotion type ($F_{6,468} = 2.791$; $p = 0.03$) and group ($F_{1,78} = 5.436$; $p = 0.02$) but no emotion × group interaction ($F_{6,468} = 0.729$; $p = 0.56$).

Most misattributions were made for neutral, followed by surprise and sadness, and the fewest for happiness. Patients misattributed emotions more often than control subjects, but on post hoc testing the difference was only statistically significant for fear (*Table 25*).

Treatment (patients only)

Valid data were available for 22 patients (8 metyrapone, 14 placebo) tested at baseline and at 5 weeks. Results of the ANCOVA are presented in *Tables 26* and *27*.

TABLE 25 Misattribution (false alarm rate) by emotion and group

Emotion	Control subjects: adjusted mean[a] (SEM)	Patients: adjusted mean[a] (SEM)	$F_{1,78}$	*p*-value
All emotions	0.064 (0.003)	0.073 (0.003)	5.436	0.02
Anger	0.040 (0.005)	0.051 (0.006)	2.027	0.16
Disgust	0.040 (0.006)	0.040 (0.007)	0.003	0.96
Fear	0.030 (0.006)	0.054 (0.007)	7.185	0.01
Happiness	0.010 (0.003)	0.010 (0.003)	0.089	0.77
Sadness	0.066 (0.008)	0.081 (0.009)	1.688	0.20
Surprise	0.075 (0.006)	0.079 (0.007)	0.190	0.66
Neutral	0.192 (0.012)	0.193 (0.013)	0.008	0.93

a Estimated marginal mean from ANCOVA model.

Accuracy

Patients receiving placebo had a lower hit rate at baseline than those who received metyrapone; at week +5, patients receiving placebo increased their hit rate while those on metyrapone decreased it.

TABLE 26 Accuracy (hit rate) by emotion at week +5 (covaried for baseline)

Emotion	Placebo: adjusted mean[a] (SEM)	Metyrapone: adjusted mean[a] (SEM)	$F_{1,19}$	*p*-value
Anger	0.39 (0.04)	0.24 (0.05)	5.255	0.03
Disgust	0.54 (0.04)	0.53 (0.05)	0.031	0.86
Fear	0.35 (0.04)	0.38 (0.05)	0.102	0.75
Happiness	0.68 (0.06)	0.68 (0.04)	0.004	0.95
Sadness	0.59 (0.07)	0.57 (0.09)	0.210	0.89
Surprise	0.67 (0.04)	0.74 (0.06)	1.040	0.32
Neutral	0.77 (0.05)	0.82 (0.06)	0.263	0.61

a Estimated marginal mean from ANCOVA model.

Misattribution

TABLE 27 Misattribution (false alarm rate) at week +5 (covaried for baseline)

Emotion	Placebo: adjusted mean[a] (SEM)	Metyrapone: adjusted mean[a] (SEM)	$F_{1,19}$	*p*-value
Anger	0.061 (0.011)	0.049 (0.015)	0.354	0.56
Disgust	0.047 (0.013)	0.077 (0.17)	1.905	0.18
Fear	0.051 (0.010)	0.050 (0.013)	0.003	0.96
Happiness	0.011 (0.003)	0.020 (0.004)	2.480	0.13
Sadness	0.079 (0.018)	0.070 (0.023)	0.099	0.76
Surprise	0.078 (0.009)	0.067 (0.012)	0.457	0.51
Neutral	0.180 (0.012)	0.184 (0.018)	0.053	0.82

a Estimated marginal mean from ANCOVA model.

Emotional Memory Task

Groups at baseline

Valid data were available for 52 HVs and 65 depressed patients for recall, and 51 HVs and 63 depressed patients for recognition.

Emotional word immediate recall

There was a significant effect of group ($F_{1,114} = 9.859$; $p = 0.002$), with a trend to a significant effect of valence ($F_{2,228} = 2.871$; $p = 0.062$) and a significant group x valence interaction ($F_{2,228} = 3.133$; $p = 0.049$). Post hoc analysis of emotional bias is shown in *Table 28*. Patients remembered fewer words overall than control subjects but a larger proportion of these were negative.

Emotional word delayed recognition: hits

There was no significant effect of group ($F_{1,111} = 1.835$; $p = 0.18$), valence ($F_{2,222} = 1.030$; $p = 0.36$) or a significant group × valence interaction ($F_{2,222} = 1.389$; $p = 0.25$). Post hoc analysis of emotional bias is shown in *Table 29*. Although patients recognised fewer words overall than control subjects and a larger proportion of these were negative, the differences were small and non-significant.

TABLE 28 Immediate recall at baseline by group

Emotion	Control subjects: adjusted mean[a] (SEM)	Patients: adjusted mean[a] (SEM)	$F_{1,114}$	*p*-value
All words[b]	10.91 (0.42)	9.14 (0.38)	9.592	0.002
Positive[c]	3.20 (0.16)	2.92 (0.14)	1.622	0.21
Negative[c]	3.19 (0.14)	3.68 (0.13)	6.249	0.01
Neutral[c]	3.57 (0.19)	3.33 (0.17)	0.024	0.35

a Estimated marginal mean from ANCOVA model.
b Covaried for zero-centred IQ. Values taken from a univariate analysis on total words rather than repeated measures ANCOVA in order to show total number of words rather than mean across valences.
c Covaried for total number of words recalled.

TABLE 29 Delayed recognition: hits by group

Emotion	Control subjects: adjusted mean[a] (SEM)	Patients: adjusted mean[a] (SEM)	$F_{1,111}$	*p*-value
All words[b]	23.75 (0.64)	22.64 (0.56)	1.983	0.16
Positive[c]	7.75 (0.15)	7.62 (0.14)	0.389	0.53
Negative[c]	7.72 (0.16)	7.95 (0.14)	1.132	0.29
Neutral[c]	7.60 (0.15)	7.51 (0.14)	0.217	0.64

a Estimated marginal mean from ANCOVA model.
b Covaried for zero-centred IQ. Values taken from a univariate analysis on total words rather than repeated measures ANCOVA in order to show total number of words rather than mean across valences.
c Covaried for total number of words recognised.

Emotional word delayed recognition: false alarms

There was a strong trend towards a significant effect of group ($F_{1,111} = 3.931$; $p = 0.05$), a significant effect of valence ($F_{2,222} = 8.387$; $p < 0.001$) but no significant group × valence interaction ($F_{2,222} = 1.235$; $p = 0.29$). Post hoc analysis of emotional bias is shown in *Table 30*. There were more false alarms for emotional than neutral words, and patients had more false alarms overall. Although patients had relatively fewer positive and more negative false alarms than control subjects, these were not significant.

Treatment (patients only)

Valid data were available for 41 patients (19 metyrapone, 22 placebo) tested at baseline and at 5 weeks.

Recall: No significant effect of treatment was found. Results of the ANCOVA are presented in *Table 31*.

Recognition: No significant effects of treatment were found. Results of the ANCOVAs are presented in *Tables 32* and *33*.

TABLE 30 Delayed recognition: false alarms by group

Emotion	Control subjects: adjusted mean[a] (SEM)	Patients: adjusted mean[a] (SEM)	$F_{1,111}$	*p*-value
All words[b]	4.16 (0.64)	5.49 (0.45)	3.931	0.05
Positive[c]	2.00 (0.14)	1.72 (0.13)	2.210	0.14
Negative[c]	1.68 (0.16)	1.89 (0.14)	1.006	0.32
Neutral[c]	1.21 (0.13)	1.28 (0.12)	0.158	0.69

a Estimated marginal mean from ANCOVA model.
b Covaried for zero-centred IQ. Values taken from a univariate analysis on total words rather than repeated measures ANCOVA in order to show total number of words rather than mean across valences.
c Covaried for total number of false alarms.

TABLE 31 Immediate recall at week +5 by treatment (covaried for baseline)

Emotion	Placebo: adjusted mean[a] (SEM)	Metyrapone: adjusted mean[a] (SEM)	$F_{1,37}$	*p*-value
All words	10.18 (0.62)	9.44 (0.65)	0.671	0.42
Positive	3.10 (0.35)	2.52 (0.37)	1.287	0.26
Negative	3.64 (0.28)	3.35 (0.30)	0.505	0.48
Neutral	3.59 (0.42)	3.41 (0.45)	0.083	0.78

a Estimated marginal mean from ANCOVA model.

TABLE 32 Delayed recognition: hits at week +5 by treatment (covaried for baseline)

Emotion	Placebo: adjusted mean[a] (SEM)	Metyrapone: adjusted mean[a] (SEM)	$F_{1,38}$	*p*-value
All words	21.96 (1.10)	22.63 (1.19)	0.171	0.68
Positive	6.95 (0.46)	7.32 (0.37)	0.304	0.59
Negative	7.97 (0.41)	7.93 (0.44)	0.007	0.94
Neutral	7.06 (0.47)	7.35 (0.50)	0.183	0.67

a Estimated marginal mean from ANCOVA model.

TABLE 33 Delayed recognition: false alarms at week +5 by treatment (covaried for baseline)

Emotion	Placebo: adjusted mean[a] (SEM)	Metyrapone: adjusted mean[a] (SEM)	$F_{1,38}$	*p*-value
All words	5.95 (0.89)	6.32 (0.95)	0.083	0.78
Positive	1.98 (0.39)	2.60 (0.42)	1.198	0.28
Negative	2.33 (0.32)	1.93 (0.34)	0.715	0.40
Neutral	1.62 (0.36)	1.81 (0.39)	0.121	0.73

a Estimated marginal mean from ANCOVA model.

Affective GoNoGoTask

Groups at baseline

Valid data were available for 47 HVs and 57 depressed patients.

Analysis by mean Go blocks

There was no effect of group ($F_{1,101} = 0.942$; $p = 0.33$), a significant effect of valence ($F_{2,202} = 104.773$; $p < 0.001$), but no significant group × valence interaction ($F_{2,202} = 2.185$; $p = 0.12$). RTs were longer for the blocks with neutral targets than for those with emotional targets. Patients had longer RTs than control subjects for positive targets. Post hoc analysis is shown in *Table 34*.

Paired-sample *t*-tests in each group separately showed that GoPos and GoNeg differed significantly in patients ($t_{56} = 3.706$; $p < 0.001$) but not control subjects ($t_{46} = 0.0569$; $p = 0.57$).

Analysis by mean NoGo blocks

There was no effect of group ($F_{1,102} = 0.942$; $p = 0.33$), a significant effect of valence ($F_{2,202} = 20.228$; $p < 0.001$) but no significant group × valence interaction ($F_{2,202} = 0.330$; $p = 0.72$). RTs were shorter for the blocks with neutral distractors than for those with emotional distractors. Patients had slightly longer RTs than control subjects but this was not significant. Post hoc analysis is shown in *Table 35*.

Treatment (patients only)

Valid data were available for 38 patients (17 metyrapone, 21 placebo) tested at baseline and at 5 weeks. No significant effect of treatment was found. Results of the post hoc ANCOVAs are presented in *Table 36*.

Functional magnetic resonance imaging

In the Newcastle centre patient sample

In total, 22 patients and 15 healthy control subjects participated in this aspect of the study at week 0. Three patients and two control subjects were not included in the analysis due to excessive head motion.

TABLE 34 Mean Go blocks by emotion: RTs (milliseconds) at baseline by group

Emotion	Control subjects: adjusted mean[a] (SEM)	Patients: adjusted mean[a] (SEM)	$F_{1,101}$	*p*-value
GoPos[b]	856 (15)	900 (14)	4.444	0.04
GoNeg[b]	848 (17)	860 (15)	0.314	0.58
GoNeut[b]	998 (18)	997 (16)	0.001	0.98
GoPos[c]	865 (8)	892 (8)	5.458	0.02
GoNeg[c]	860 (8)	854 (7)	0.320	0.57
GoNeut[c]	1008 (11)	988 (10)	1.992	0.16

a Estimated marginal mean from ANCOVA model.
b Covaried for zero-centred age.
c Covaried for mean RT.

TABLE 35 Mean NoGo blocks by emotion: RTs (milliseconds) at baseline by group

Emotion	Control subjects: adjusted mean[a] (SEM)	Patients: adjusted mean[a] (SEM)	$F_{1,101}$	*p*-value
NoGoPos[b]	915 (17)	930 (15)	0.445	0.51
NoGoNeg[b]	923 (15)	944 (13)	1.052	0.31
NoGoNeut[b]	865 (16)	891 (15)	1.384	0.24
NoGoPos[c]	925 (8)	917 (7)	0.545	0.46
NoGoNeg[c]	932 (7)	934 (7)	0.030	0.86
NoGoNeut[c]	875 (8)	882 (7)	0.377	0.54

a Estimated marginal mean from ANCOVA model.
b Covaried for zero-centred age.
c Covaried for overall RT.

TABLE 36 Mean Go and NoGo blocks by emotion: RTs (milliseconds) at week +5 by treatment (covaried for baseline)

Emotion	Placebo: adjusted mean[a] (SEM)	Metyrapone: adjusted mean[a] (SEM)	$F_{1,35}$	*p*-value
GoPos	871 (16)	895 (18)	0.936	0.34
GoNeg	866 (18)	852 (20)	0.292	0.59
GoNeut	1,010 (17)	988 (20)	0.765	0.39
NoGoPos	941 (15)	910 (17)	1.843	0.18
NoGoNeg	916 (16)	944 (18)	1.285	0.27
NoGoNeut	893 (18)	880 (20)	0.211	0.65

a Estimated marginal mean from ANCOVA model.

Facial emotion processing task

Whole-brain analysis showed a cluster of significantly higher ($p < 0.05$, FWE corrected) BOLD signal in the right amygdala in both the EM condition and the EL condition than the SM condition (*Figure 14*). *Table 37* shows the estimated marginal means of the BOLD signal in the three conditions in this cluster. Controlling for age, sex and premorbid IQ [National Adult Reading Test (NART)], there was a significant difference between patients and control subjects in the EM condition, but not in the EL or SM conditions, with patients showing higher BOLD values than control subjects.

Estimated marginal means for the BOLD signal of the three conditions in anatomically defined amygdala regions, separately for the left and right hemisphere, are shown in *Table 38*. There were no significant differences between the two groups in any of the conditions in either region, although there was a statistical trend for higher BOLD signal for patients in the left amygdala for the EM condition.

Emotional enhancement of memory task

A whole-brain analysis did not show clusters of differential BOLD signal between the three conditions (negative arousing, negative non-arousing, neutral) in the amygdala across both groups (*Table 39*).

No significant differences in activation between patients and control subjects in any of the conditions were found.

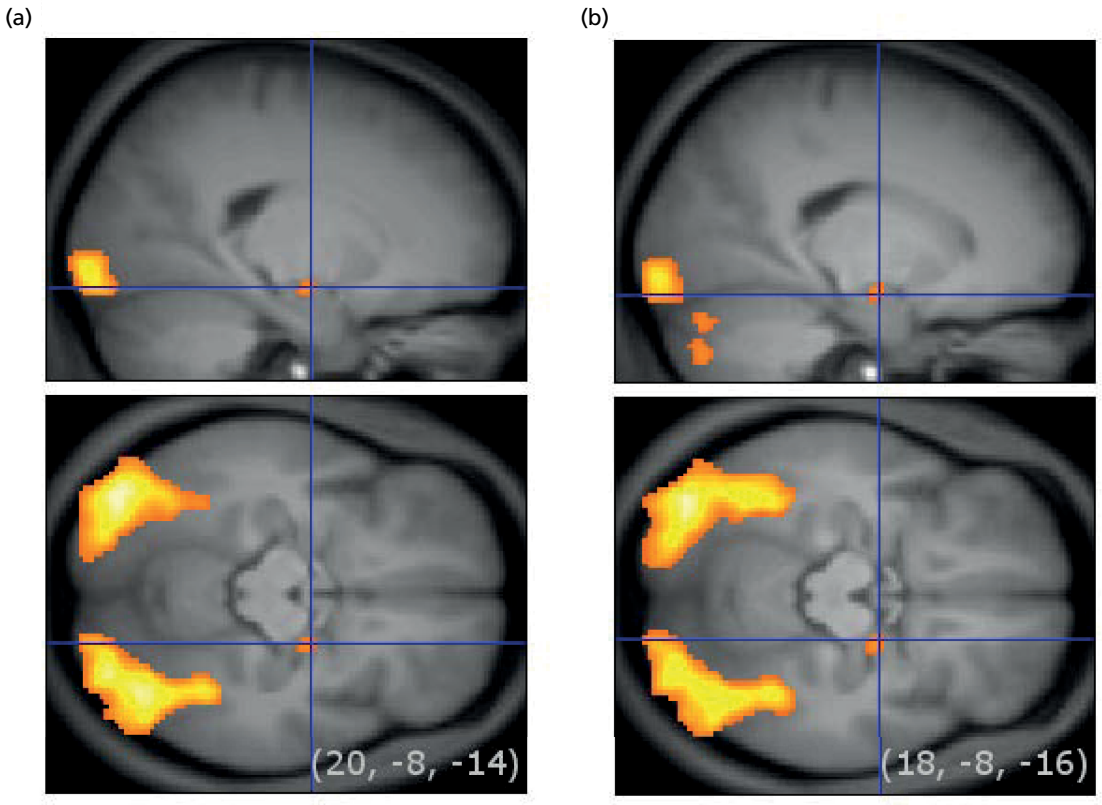


FIGURE 14 Significant cluster of BOLD signal increases in the EM condition (a) and EL condition (b) over the control condition SM across the entire sample ($p < 0.05$, FEW corrected, $k = 10$). Numbers in parentheses are MNI coordinates (in mm) of the local signal peaks.

TABLE 37 Average BOLD signals (and SD) in the FEP task from a cluster showing significant differences between the facial processing conditions (EM, EL) and the control condition in the right amygdala (MNI 18, –8, –16; $k = 96$)

	EM	EL	SM
Patients	0.352 (0.371)	0.408 (0.446)	–0.120 (0.295)
Control subjects	0.067 (0.371)	0.334 (0.449)	–0.227 (0.298)
Group effect	$F = 5.076$; $p = 0.033$	$F < 1$ (n.s.)	$F = 1.112$; $p = 0.301$

n.s., not significant.

TABLE 38 Average BOLD signals (and SD) in the FEP task for anatomically defined regions of interest of the left and right amygdala

	EM	EL	SM
Left amygdala			
Patients	0.280 (0.389)	0.450 (0.492)	0.244 (0.328)
Control subjects	0.018 (0.391)	0.373 (0.496)	0.108 (0.329)
Group effect	$F = 3.838$; $p = 0.061$	$F < 1$ (n.s.)	$F = 1.455$; $p = 0.238$
Right amygdala			
Patients	0.317 (0.319)	0.307 (0.384)	0.114 (0.253)
Control subjects	0.141 (0.321)	0.357 (0.387)	0.072 (0.251)
Group effect	$F = 2.605$; $p = 0.118$	$F < 1$ (n.s.)	$F < 1$ (n.s.)

n.s., not significant.

TABLE 39 Estimated marginal means for the anatomically defined amygdala ROI

		Negative	
	Neutral	Arousing	Non-arousing
Left amygdala			
Patients	–0.075 (0.250)	–0.201 (0.201)	–0.019 (0.244)
Control subjects	0.037 (0.313)	–0.015 (0.252)	0.020 (0.305)
Group effect	$F < 1$ (n.s.)	$F < 1$ (n.s.)	$F < 1$ (n.s.)
Right amygdala			
Patients	–0.323 (0.131)	–0.577 (0.155)	–0.349 (0.151)
Control subjects	–0.222 (0.164)	–0.323 (0.194)	–0.240 (0.189)
Group effect	$F < 1$ (n.s.)	$F = 1.007$; $p = 0.324$	$F < 1$ (n.s.)

n.s., not significant.

In the Manchester centre patient sample

Twenty-seven patients were eligible for this study, one of whom subsequently withdrew consent. Twenty-six patients participated in the final study (13 females). Twenty patients were right handed, six ambidextrous and 1 left handed. Given the literature regarding laterality[118] and hand preference combined with the relative difficulty to recruit and retain patients with TRD, it was decided to take a pragmatic approach and recruit patients with a preference for left-handedness and perform a post hoc sensitivity analysis. One scan from the patient group was excluded from analysis owing to movement artefact, and another subject was withdrawn from the study owing to the discovery of a structural abnormality; therefore, data are presented for 24 patients.

Thirty-two right-handed HVs were recruited from advertisements by the University of Manchester. Thirty were eligible (15 females). All of the female subjects were on long-term hormonal contraception to control for the effects of menstrual cycle on endogenous cortisol levels to allow reliable within-subject comparison[119] but subjects were otherwise medication free, drank < 21 units of alcohol/week and were caffeine free on the scan days.

One male HV was later excluded owing to a head-coil malfunction; therefore data are reported from 29 subjects.

The mean age of patients was 44.1 years (SD 7.7 years; range 26–57 years) compared with mean age 36.47 years (SD 5.94 years; range 31–55 years) for the HVs. The mean NART IQ score for the patient group was 109.7 (SD 10.0) and for the HVs 113.3 (SD 10.4). There was no significant difference between the patient and HV groups in either age or IQ ($p = 0.22$ and $p = 0.15$, respectively). The baseline GRID-HAMD score for the patient group was 23.74 (SD 3.7) compared with 0.28 (SD 0.5) for the HVs. There was a significant difference in handedness between the HV group and the patients, with one left-handed individual and five ambidextrous individuals in the TRD group; however, the results were not substantially altered by the removal of this individual in post hoc sensitivity analysis.

Behavioural results

Emotional pictures encoding task

There was no significant difference in the number of responses made between participant groups ($F_{1,51} = 0.02$; $p = 0.88$) or RT ($F_{1,51} = 0.84$; $p = 0.36$). There was a main effect for group for subjective emotionality ($F_{1,51} = 5.14$; $p = 0.03$), with patients classifying fewer positive images as 'emotional'; however, there was no significant participant group × valence effect ($F_{1,51} = 1.19$; $p = 0.28$).

Emotional pictures retrieval task

There was no significant effect of the emotional valence (positive or neutral image) in the number of words omitted in the retrieval tasks ($F_{1,50} = 2.40$; $p = 0.13$) and no effect of group (TRD or HV) by emotional valence on number omitted ($F_{1,50} = 2.03$; $p = 0.16$). There was also no effect of emotion on RT ($F_{1,50} = 0.91$; $p = 0.35$) and no effect of group by emotion on RT ($F_{1,50} = 0.12$; $p = 0.73$). There was also no effect of emotion on number of words correctly recalled ($F_{1,50} = 0.01$; $p = 0.94$) and no effect of emotion by group on correct recognition ($F_{1,50} = 10.51$; $p = 0.25$).

n-back task

There was no significant main effect of group on number correct ($F_{1,53} = 0.09$; $p = 0.76$), omitted ($F_{1,53} = 1.19$; $p = 0.28$) or RT ($F_{1,53} = 1.12$; $p = 0.29$). There was also no effect of level of difficulty or level by group on any of these measures.

Functional magnetic resonance imaging results

Emotional pictures encoding task

The main effect of task (all participants, all encoding images minus rest) showed increased BOLD signal in bilateral parahippocampal, bilateral inferior frontal gyrus (BA47, BA9), left temporal gyrus (BA20) and right hippocampus in keeping with findings of meta-analyses of encoding tasks literature.[120,121]

When the HVs were compared with the patients, patients with TRD showed a significant decreased BOLD signal at whole-brain level in the posterior cingulate compared with the HVs during the encoding task [HVs minus patients, all encoding images minus rest contrast, BA31; $x = 9.5$, $y = -35$, $z = 35$, $Z = 3.47$ p(FWE) (whole brain corrected) = 0.043. There were no significant findings in the pre-hypothesised ROIs]. (see *Table 40* and *Figure 15*). There were no significant results in the opposite subtraction.

Emotional pictures retrieval task

The main effect of task (all participants, all retrieval tasks minus rest) showed increased BOLD signal in the parahippocampus bilaterally, the left inferior temporal lobe (BA36), the left lingual gyrus (BA18) and left BA9 and BA6. There was no significant change in BOLD signal in the hippocampus.

In the a priori ROI of the insula, the BOLD signal was increased in the patients compared with the control subjects while completing the retrieval task, compared with the rest condition (patients minus control subjects, retrieval images minus rest contrast) [posterior insula, BA40; $x = -46.5$, $y = -21$, $z = 20$, $Z = 3.85$ p(FWE)(ROI analysis) = 0.046] (see *Table 40* and *Figure 16*). In the anterior cingulate, another a priori ROI, the BOLD signal was reduced in the patients compared with the control subjects while retrieving positive images, compared with neutral images (control subjects minus patients, positive images minus neutral images contrast) [anterior cingulate, BA33; $x = -4.5$, $y = 17.5$, $z = 25$, $Z = 3.95$ p(FWE)(ROI analysis) = 0.033] (see *Table 40* and *Figure 17*).

TABLE 40 Results of the three fMRI tasks

			Coordinates (MNI)			Cluster		
Area	BA	Side	*x*	*y*	*z*	K	*Z*	*p*(FWE)
Emotional Pictures Encoding task *Encode–rest HV–Pts*								
Posterior cingulate	31	R	9.5	–35	35	77	3.47	0.043[a]
Emotional Pictures Retrieval task *Retrieve–rest HV–Pts*								
Insula	40	L	–46.5	–21	20	56	3.85	0.046[b]
Positive–neutral HV–Pts								
Anterior cingulate	33	L	–4.5	17.5	25	18	3.95	0.033[b]
n-back task *2back–0back HV–Pts*								
dlPFC	9	L	–25.5	7	50	68	4.53	< 0.001[a]

BA, Brodmann area; K, cluster size at $p < 0.005$; L, left; Pts, patients; R, right.
a Whole-brain corrected p(FWE) < 0.05.
b p(FWE) < 0.05 following a priori ROI analysis.

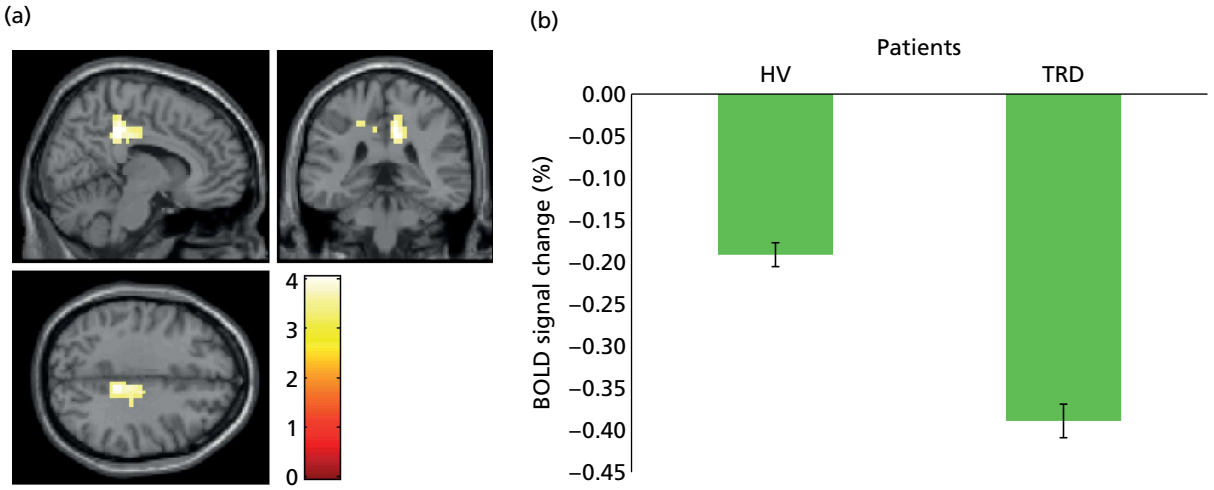


FIGURE 15 Significant BOLD decrease in the patients compared with the control subjects in the posterior cingulate (9.5, –35, 35), *p*(FWE) whole brain = 0.043 during the encoding task. Control subjects minus patients, all encoding images minus rest, error bars indicate SEM.

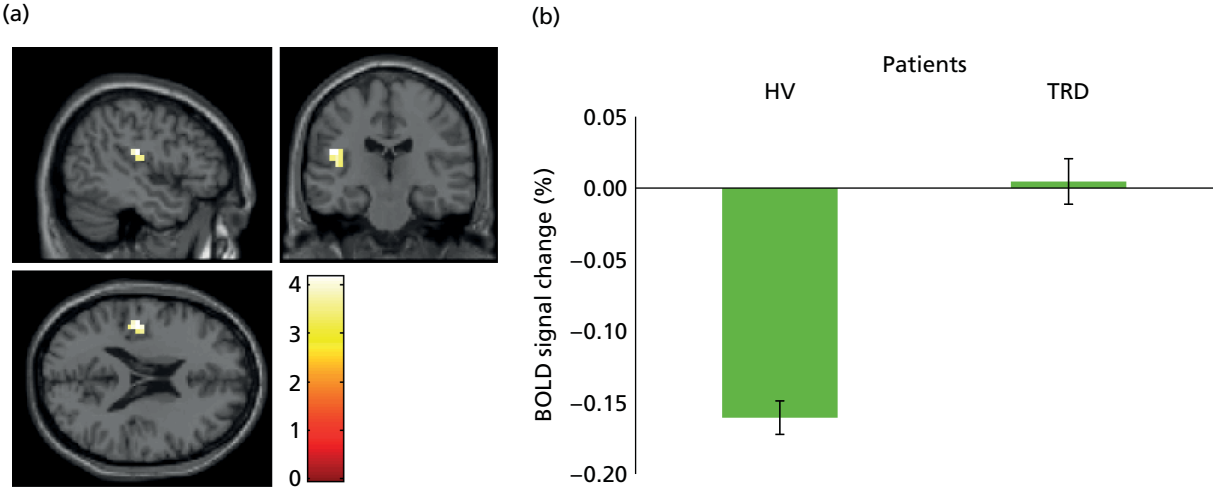


FIGURE 16 Significant BOLD increase in the patients compared with the control subjects in the posterior insula (–46.5, –21, 20), ROI *p*(FWE) = 0.046 during the retrieval task. Patients minus control subjects, all retrieval images minus rest, error bars indicate SEM.

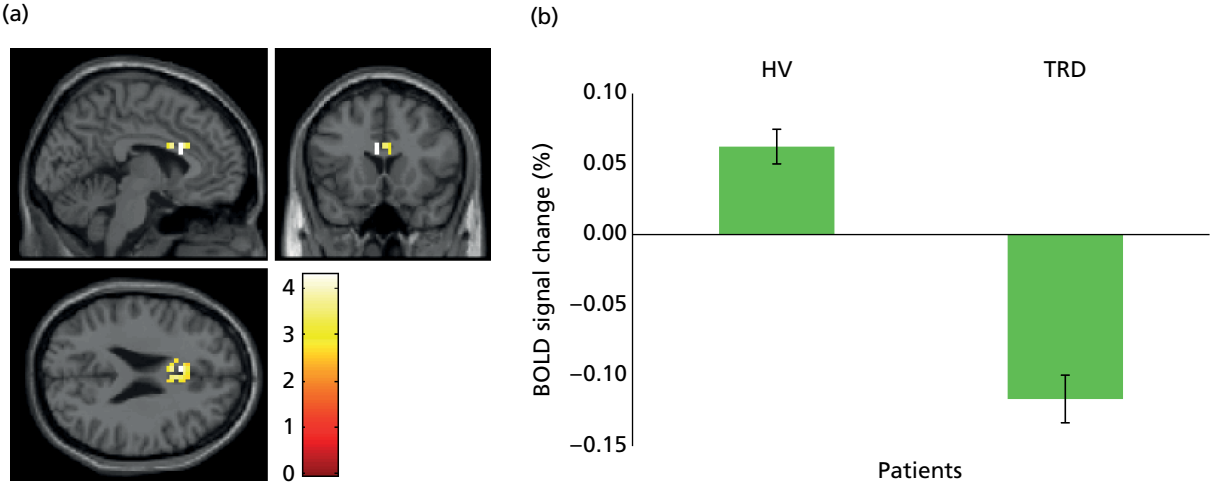


FIGURE 17 Significant BOLD increase in the patients compared with the control subjects in the anterior cingulate (–4.5, 17.5, 25) ROI *p*(FWE) = 0.033 during the retrieval task. Control subjects minus patients, positive minus neutral, error bars indicate SEM.

n-back task

The main effect of the task (using all participant scans), the two-back compared with zero-back subtraction showed increased BOLD signal bilaterally in the dlPFC, medial posterior parietal and lateral premotor cortices, in keeping with the areas reported in a previous meta-analysis.[122] In the control subjects minus patients two-back minus zero-back contrast, the patients showed a decrease in BOLD signal compared with the control subjects in the dlPFC [BA6; $x = -25.5$, $y = 7$, $z = 25$, $Z = 50$ p(FWE)(whole brain corrected) = < 0.001] (see *Table 40* and *Figure 18*). However, there was no difference seen in our ROI areas.

Acute effects of hydrocortisone in treatment-resistant depression using pharmacological functional magnetic resonance imaging

The final sample consisted of 20 patients with TRD and 27 HVs. The a priori regions of interest were the hippocampus, hypothalamus, amygdala and dorsolateral and ventrolateral prefrontal cortex (all areas rich in corticosteroid receptors). We failed to demonstrate a difference between the TRD and HV groups, and also a main effect of hydrocortisone, in contrast with our previous study.[107] Reasons that may have accounted for this result include lack of power and artefact.

Electroencephalography

At study initiation it had been intended to recruit 50 patients to the EEG mechanistic studies, giving approximately 25 treated with metyrapone and 25 with placebo. Given the slow recruitment to the study initially, two decisions were made – first, there was an acceptance of a lower sample size than the originally planned 190, and, second, emphasis was given to recruitment to the main treatment element of the study, including patients who did not wish to consent to recruitment to the mechanistic studies. A total of 34 patients were recruited for the EEG studies. Twenty-five of these had MADRS data at the primary end point at week +5. Of these, 14 were treated with placebo and 11 with metyrapone, with five responders and six non-responders. These numbers did not provide sufficient power to warrant examination of the EEG variables suggested to predict response to active treatment, and therefore this was not pursued further.

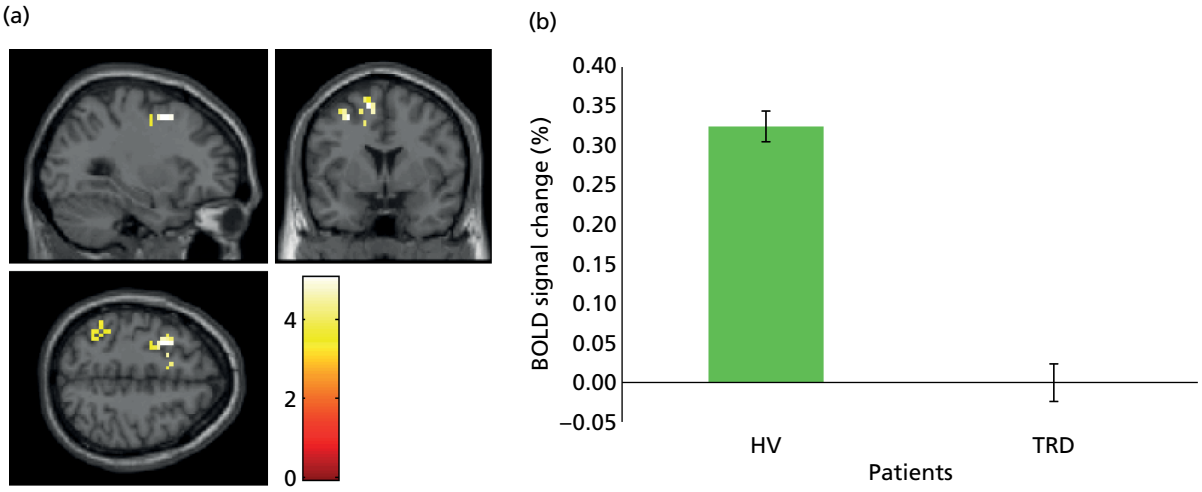


FIGURE 18 Significant BOLD increase in the control subjects compared with patients in the dlPFC (–25.5, 7, 50) whole-brain p(FWE) < 0.001 in the two-back minus zero-back contrast. Error bars indicate SEM.

With regard to the ESMT, the issue of the small patient sample size was compounded by a methodological issue. A large amount of horizontal eye movement was found in patients during the test phase. This led to a high epoch reject rate. EEG analysis of the ESMT task was therefore not conducted owing to an insufficient number of both patients and control subjects meeting EEG signal-to-noise ratio inclusion criteria. For a satisfactory signal–noise ratio, at least 10 epochs would need to be present in each of three task outcome measures (neutral, negative and positive correctly recalled items). Only three patients and one control met this criterion.

With regard to the LTP analysis, this was conducted on 32 patients and 20 control subjects. The mean VEP waveforms for patients and control subjects from the first pre-tetanus and first post-tetanus assessment are shown in *Figure 19*. Data for the individual components of the VEP are shown in *Figures 20–22*.

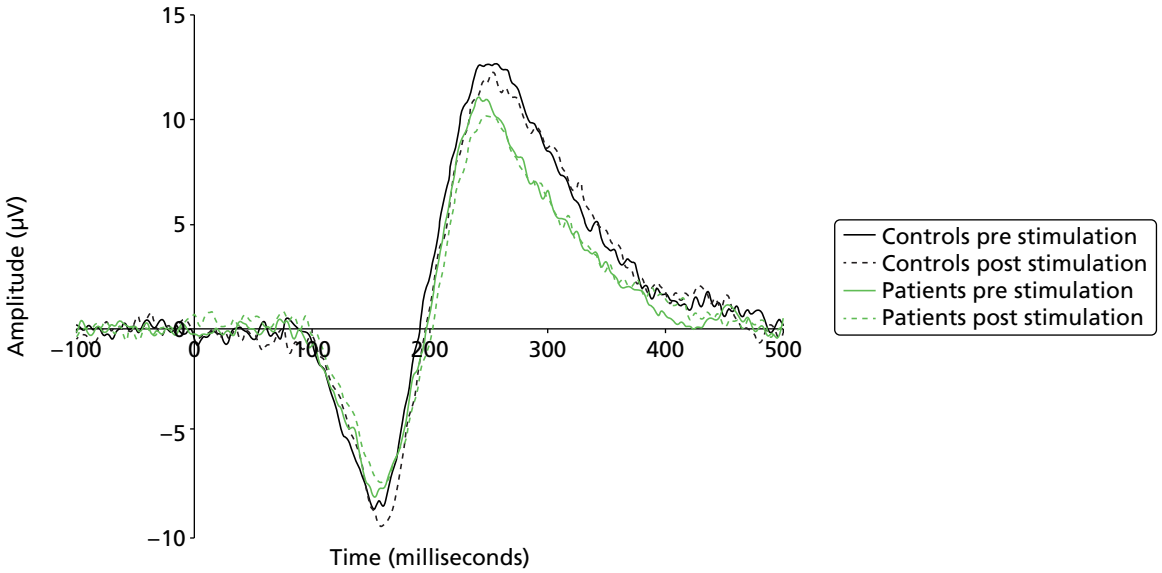


FIGURE 19 Mean VEP waveforms of patients (green lines) and control subjects (black lines) before (solid lines) and after (dotted lines) visual tetanic stimulation.

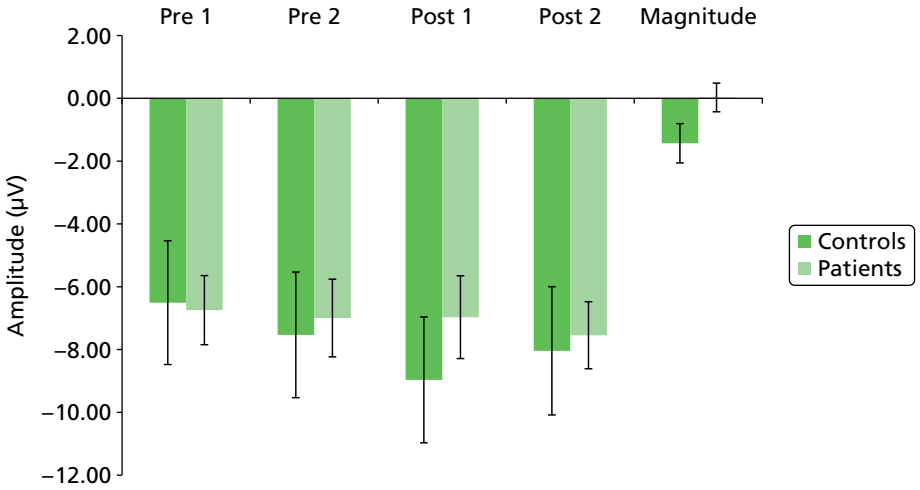


FIGURE 20 Shows N1 amplitude, together with the magnitude of enhancement from the mean of the two pre-tetanus amplitudes to the first post-tetanus amplitude (error bars are SEM).

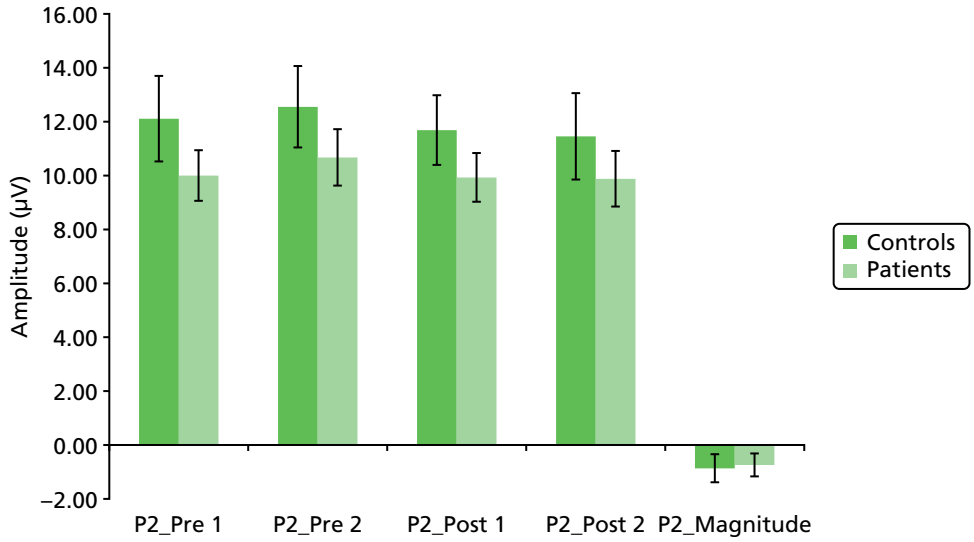


FIGURE 21 Shows P2 amplitude, respectively, together with the magnitude of enhancement from the mean of the two pre-tetanus amplitudes to the first post-tetanus amplitude (error bars are SEM).

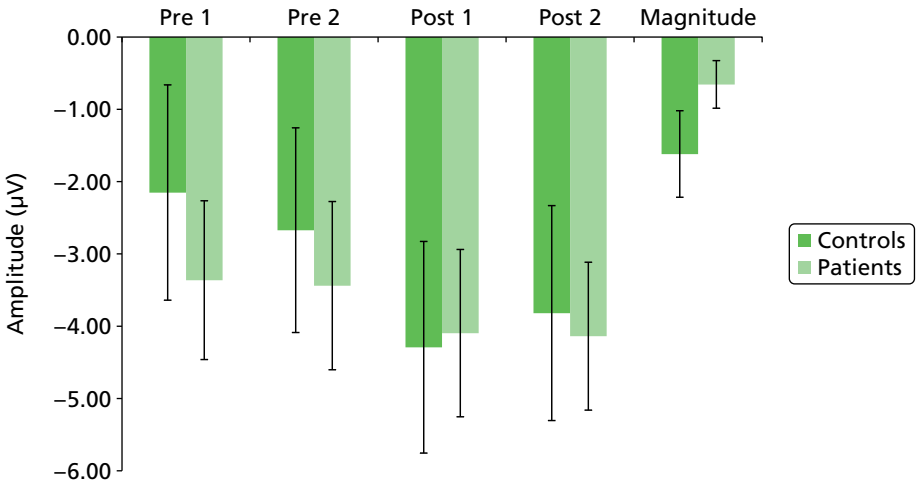


FIGURE 22 Shows N1b amplitude, respectively, together with the magnitude of enhancement from the mean of the two pre-tetanus amplitudes to the first post-tetanus amplitude (error bars are SEM).

There was no significant difference in magnitude of either the first or second pre-tetanus N1, P2 or N1b components between the patients and the healthy control subjects (all p-values > 0.2). As can be seen, there was an increase in amplitude of N1 and N1b post tetanus, especially in the healthy control subjects, although little change in amplitude of P2. In the healthy control subjects there was a significant difference in N1 and N1b amplitudes in the post-tetanus test 1 block compared with the mean of the pre-tetanus blocks ($t = 2.17$, $p = 0.035$; $t = 3.51$; $p = 0.001$ respectively), but this was not seen for the P2 component, or for any component in the patient group ($p > 0.2$ for all comparisons). Comparison of the magnitude of the increase in N1 amplitude showed a significantly greater increase for the control subjects than the patients ($t = 2.01$; $p = 0.043$). The difference in magnitude of increase in N1b between patients and control subjects, however, did not reach significance ($t = 1.52$; $p = 0.146$).

It was not possible to examine the effect of metyrapone on the LTP EEG measure owing to insufficient data being available.

Analyses not yet complete

The data from, and analyses of, the following measures are not included in this report, as either the data are not yet to hand or analyses are not yet complete: negative life events questionnaire [adapted from the List of Threatening Experiences;[77] the Social Circumstances Questionnaire, which is an adaptation of the Social Support Questionnaire[78] as used in the NewMood study;[75] the Ruminative Responses Scale;[80] the 11-deoxycortisol results (data not yet to hand due to delays in optimising assay sensitivity, and therefore the secondary analyses bases on this measure of apparent concordance with the drug)] and the genetic results (data not yet to hand). We have not been able to carry out a per-protocol analysis on those patients who took medication for 3 weeks, as this information was not available. A number of analyses that were to be based on responders were not carried out, as there were insufficient numbers of responders for analysis and no metyrapone-versus-placebo difference in the numbers of responders. A number of the cross-sectional analyses and the effect of differences in ratings of depression between various scales remain to be fully evaluated. It is anticipated all the analyses will be complete by the end of 2015.

Harms

Serious adverse events

A total of 14 serious adverse events (SAEs) were reported for 11 of the patients randomised to study medication (five in the group randomised to metyrapone and six in the group randomised to placebo). Events before and after randomisation are shown in *Tables 41* and *42*.

None of the above SAEs was deemed to be related to medication by the PI in the appropriate centre. Most occurred well after the medication period and/or were related to pre-existing conditions.

TABLE 41 Events between screening and randomisation

Event	Allocation
Committed suicide prior to randomisation	Not randomised
Intracranial lesion discovered during scanning procedure	Metyrapone[a]

a This patient did attend visit 2 and was randomised to metyrapone. However, after the lesion came to light (scan taken at screening visit), this patient was not given any study medication. The patient attended no follow-up visits.

TABLE 42 Events recorded after commencement of study medication

Event description	Allocation
Breast cancer (diagnosis preceded randomisation), dehydration/leucopenia	Placebo
Groin infection/abscess	Placebo
Overdose (tramadol, pregabalin, amitriptyline and alcohol)	Metyrapone
Migraine with medication over-use headaches, right Holmes–Adie pupil	Metyrapone
Increased suicidal thoughts	Placebo
Patient fell off bike – dislocated knee/broken leg	Placebo
Overactive bladder	Metyrapone
Mood deterioration following screening	Metyrapone
Patient took unknown amount of ketamine	Placebo
Increase in symptoms of depressed mood and re-emergence of suicidal ideation	Placebo

Adverse events

One AE was reported between screening and randomisation; the patient was not randomised and it is excluded in the following analyses. Multiple entries with the same date at onset for a particular patient are considered to be a single event. A total of 229 AEs were reported by patients recruited and randomised. Of these, 83 (38%, 95% CI 32% to 45%) were thought to be possibly related to study medication and 18 (8%, 95% CI 4% to 12%) were thought to be definitely related to study medication. Twelve (5%, 95% CI 2% to 9%) of the events led to adjustment, interruption or discontinuation of study medication.

Of the 165 patients recruited and randomised, 105 (63.6%, 95% CI 55.8% to 71.0%) had at least one AE. The mean number of events reported was 2.0 (3.2 if we consider only those 105 patients who reported at least one). These data are shown in *Table 43*.

There were significant differences between centres in terms of the proportion of patients reporting AEs and this is shown in *Table 44*.

There was very little difference in the likelihood of patients recruited in secondary care reporting an AE compared with patients recruited in primary care (*Table 45*) but any difference may be masked by the differences between centres.

The number of AEs was analysed using negative binomial regression. The incidence rate ratios (risk in group randomised to metyrapone divided by risk in group randomised to placebo) were as follows: unadjusted estimate = 1.34 (95% CI 0.90 to 1.98); estimate adjusted for centre and origin of patient = 1.41 (95% CI 0.98 to 2.03). In each case the 95% CI spans one corresponding to equal risk of the reporting of an AE in the two treatment groups. However, the upper limits are very close to two; we cannot exclude the possibility that treating patients with metyrapone may increase the risk of AEs. If the analysis is restricted to events classified as 'possibly' or 'definitely' related to study medication the corresponding estimates are as follows: unadjusted estimate = 1.71 (95% CI 0.98 to 3.01); estimate

TABLE 43 Numbers of AEs in randomised groups

	Treatment		
No. of AEs	Placebo	Metyrapone	Total
0	36	24	60
1	14	16	30
2	13	13	26
3	6	11	17
4	4	6	10
5	3	4	7
6	1	2	3
7	1	2	3
8	1	2	3
9	0	2	2
10	1	0	1
11	2	0	2
13	0	1	1
Total	82	83	165

TABLE 44 Numbers reporting AEs in each centre

Centre		Any AE?		Total
		No	Yes	
Leeds/Bradford	Count	21	17	38
	%	55.3	44.7	100.0
Manchester	Count	26	33	59
	%	44.1	55.9	100.0
Newcastle/Durham	Count	13	55	68
	%	19.1	80.9	100.0
Total	Count	60	105	165
	%	36.4	63.6	100.0

TABLE 45 Numbers reporting AEs in groups depending on origin of referral

Origin of patient		Any AE?		Total
		No	Yes	
Primary care	Count	23	47	70
	%	32.9	67.1	100.0
Secondary care	Count	37	58	95
	%	38.9	61.1	100.0
Total	Count	60	105	165
	%	36.4	63.6	100.0

adjusted for centre and origin of patient = 1.92 (95% CI 1.14 to 3.24). There were six instances of hypocortisolaemia during the study (one on placebo and five on metyrapone). In all cases this was asymptomatic. As part of a standard operating procedure, all patients with hypocortisolaemia had their lying and standing blood pressure and urea and electrolytes checked but no abnormalities were detected in these measures. Medication was continued and repeat cortisol measurements at later dates returned to normal.

The DMEC requested that the risk of events indicating raised levels of suicidality be compared between the two trial arms. The incidence rate ratio (metyrapone–placebo) based on a negative binomial regression model adjusting for centre and origin of patient care was 0.47 (95% CI 0.17 to 1.32); there was no evidence of an increased risk of events associated with suicidality in patients randomised to metyrapone.

Toronto Side Effects Scale

Across all 32 symptoms rated by this scale, there was no difference between the randomised groups in terms of incidence. There were two symptoms where the difference was significant at the 5% level (delayed ejaculation and weight loss both being greater in the placebo group), but this is only to be expected given that there are 32 symptoms rated. No differences were seen in the incidence of central nervous system side effects. The data for this are shown in *Table 46*. It is particularly noteworthy that the rates of dizziness and postural hypotension are very similar in both randomised groups.

TABLE 46 Toronto Side Effects Scale: central nervous system side effects – incidence, frequency and severity

	Incidence							Frequency						Severity					
	Metyrapone		Placebo		Metyrapone/placebo			Metyrapone			Placebo			Metyrapone			Placebo		
Side effect	*n*	%	*n*	%	RR	95% confidence limits		μ	95% confidence limits		μ	95% confidence limits		μ	95% confidence limits		μ	95% confidence limits	
						Lower	*Upper*		*Lower*	*Upper*		*Lower*	*Upper*		*Lower*	*Upper*		*Lower*	*Upper*
Nervousness	53	77	53	72	1.1	0.9	1.3	2.9	2.5	3.3	2.5	2.2	2.8	2.6	2.3	2.9	2.4	2.1	2.7
Agitation	55	80	52	70	1.1	0.9	1.4	3.1	2.7	3.5	2.6	2.3	3.0	2.8	2.4	3.1	2.5	2.2	2.8
Tremor	28	41	21	28	1.4	0.9	2.3	1.9	1.5	2.2	1.6	1.4	1.9	1.7	1.4	2.0	1.5	1.2	1.7
Myoclonus	25	36	32	43	0.8	0.6	1.3	1.9	1.6	2.3	1.9	1.6	2.2	1.6	1.4	1.9	1.6	1.4	1.8
Weakness or fatigue	56	80	61	82	1.0	0.8	1.2	3.6	3.2	4.0	3.6	3.3	4.0	3.1	2.8	3.5	3.5	3.1	3.8
Dizziness	24	53	25	34	1.0	0.7	1.6	1.7	1.4	2.0	1.7	1.4	1.9	1.7	1.4	2.0	1.7	1.4	1.9
Postural hypotension	25	36	32	43	0.8	0.6	1.3	1.6	1.4	1.8	2.0	1.7	2.3	1.6	1.4	1.90	1.7	1.4	2.0
Drowsiness	48	70	55	74	0.9	0.8	1.2	2.9	2.5	3.3	2.8	2.5	3.2	2.6	2.2	3.0	2.5	2.2	2.8
Increased sleep	15	22	18	24	0.9	0.5	1.6	1.5	1.2	1.8	1.6	1.3	1.9	1.3	1.1	1.6	1.4	1.2	1.7
Decreased sleep	46	67	55	74	0.90	0.72	1.11	3.09	2.67	3.51	3.09	2.71	3.47	2.90	2.48	3.31	2.85	2.48	3.22
Sweating	39	56	43	58	0.97	0.73	1.29	2.20	1.87	2.54	2.27	1.94	2.60	2.13	1.81	2.45	2.11	1.80	2.42
Flushing	31	45	29	39	1.15	0.78	1.69	2.00	1.67	2.33	1.73	1.47	1.99	1.81	1.52	2.10	1.70	1.42	1.98
Oedema	17	25	17	23	1.07	0.60	1.93	1.72	1.38	2.07	1.57	1.29	1.84	1.48	1.22	1.73	1.50	1.25	1.75
Headache	47	68	57	77	0.88	0.72	1.08	2.35	2.03	2.66	2.68	2.36	2.99	2.26	1.95	2.57	2.55	2.25	2.86
Blurred vision	23	33	22	30	1.12	0.69	1.82	1.67	1.39	1.95	1.62	1.35	1.89	1.54	1.30	1.77	1.49	1.27	1.70

Chapter 4 Discussion

Limitations

In general there are a few limitations to the current study. Although the trial did not reach its original target of 90% power, it achieved 84% power. In terms of the binary outcomes of response and remission, the very wide CI suggests that the study is not adequately powered to detect differences in these measures of outcome (and these were not the outcomes on which sample size was based). However, response rates were almost identical in the two groups, suggesting that if there is any difference in efficacy between metyrapone and placebo it is too small to be of clinical significance. This is also consistent with a lack of effect of metyrapone on the primary outcome measure. There are no obvious biases in terms of those patients who were randomised and the groups were very well balanced in terms of key demographic and clinical variables (see *Table 5*). The numbers retained in the study to 24 weeks are reasonable, given the complexity and severity of the underlying condition. The attrition rate in the follow-up phase was somewhat higher in the metyrapone group than in the placebo group (see *Figure 10*) but there was no evidence from joint modelling that this had any impact on the findings. Blinding was fully maintained throughout the study (see *Chapter 2*, *Randomisation and blinding*). There was no limitation with multiplicity of analyses, with the possible exception of the TSES, which involved the analysis of a large number of variables. There was a suggestion of patients being recruited from one centre (Manchester) having more severe depression that those recruited elsewhere. This probably relates to the significantly more severe depression seen in patients recruited from secondary care compared with primary care (as might be expected): in Manchester there was a relatively low rate of recruitment from primary care. Interestingly, despite the differences in severity of depression in primary care compared with secondary care patients, there was no difference in severity of anxiety in these two groups. Nevertheless, analysis has included centre and source of patient as covariates.

A current limitation of this report is that we do not have good evidence of whether or not the patients adhered to their medication. Medication was posted to patients and that system went well. Patients were reminded to take medication and asked to return any unused capsules (which was a rare event). Thus apparent concordance was good but it is known that, despite such measures, there can be problems in this domain. Although the ADD Study did not include direct measures of metyrapone plasma concentrations in patients, measurement of plasma 11-deoxycortisol after 1 week's treatment serves as a sensitive assessment of both compliance and pharmacodynamic effect of the treatment. Previous data show that metyrapone administration is associated with highly significant increases in 11-deoxycortisol[33,91] owing to its effect of blocking 11β-hydroxylase. A prespecified explanatory analysis excluding patients who did not take or discontinued treatment, assessed utilising the 11-deoxycortisol measure of apparent concordance with treatment, will be performed and results published separately.

Previous studies have suggested that patients with baseline HPA axis dysregulation may preferentially respond to metyrapone; this effect is not apparent in this data set.

Generalisability

The ADD Study is the largest RCT of antiglucocorticoid treatment of major depression in the UK and one of the largest worldwide. In general terms, a strength of the current study is that it is generalisable to the large numbers of patients in the NHS and in the community who have TRD, as defined using the criteria in this study. Compared with many previous studies, including a relatively large proof-of-concept study of metyrapone augmentation of conventional antidepressants,[33] the ADD Study recruited from a broad population of patients, including those from both primary and secondary care. Inclusion criteria were intentionally kept broad, and exclusion criteria to a minimum, in order to explore the efficacy of metyrapone treatment in as 'real world' a setting as possible. This means that, for example, patients with significant suicidal ideation, who are often excluded from RCTs in depression, were included in the ADD Study. The help of the service users on the research team and that of the MHRN Service User Forum in accessing such a broad group of patients cannot be underestimated. However, although the referral system to this study was open to patients with this disorder from all types of settings (as discussed further below), in practice 58% of randomised patients came from secondary care outpatient settings, which is an over-representation, as most such cases are managed in the community. In addition, there was a large number ($n = 708$) of dropouts between referral and randomisation. Some were not included because they failed to meet criteria for inclusion, some as a result of meeting exclusion criteria, but others for unknown reasons (e.g. did not attend) so we cannot be completely confident that our sample is fully representative of the clinical condition in the community at large.

The patients included in the ADD Study were well characterised at baseline in relation to a number of factors known to relate to TRD, and no marked baseline imbalances were noted, increasing confidence that the observed outcomes were indeed attributable to intervention. The degree of treatment refractoriness of patients was carefully assessed using the MGH-TRD staging scale,[111] which included taking a history of medication prescription and included perusal of hospital and GP records. There is a lack of consensus as to which staging scale for refractory depression should be used in studies. We chose the MGH-TRD scale owing to its ease of use, especially taking into account scoring for dose optimisation and augmentation/combination treatment. We chose a minimum MGH-TRD score of 2 points for inclusion. Use of a single antidepressant at an effective dose scores 1 point and hence the cut-off of 2 points represents a failure to respond to at least two antidepressants, given usual UK practice in primary care of not augmenting or combining medications for depression until after this stage.[1] Beyond this point in the treatment algorithm for individual patients there is great divergence in practice with patients being referred to secondary care at different stages by individual clinicians. The maximum MGH-TRD score for inclusion was set as 10 points. In practice, this means five or six trials of different antidepressants, allowing for dose optimisation and augmentation/combination strategies used for some of these trials. Use of electroconvulsive therapy scores 3 points in the MGH-TRD scale. As such, this, in itself, was not an exclusion criterion in the ADD Study. However, as most patients receiving ECT will also have had at least two antidepressants (often with dose optimisation and augmentation), in practice, few patients treated with ECT have a score under the maximum cut-off of 10 points. The mean (SD) MGH-TRD score for the randomised patients was 4.6 (1.9), indicating that most patients were well in the range of treatment refractoriness but not in the later stages of treatment refractoriness.

In broad terms, the patients studied here could be described as having moderate depression. In terms of generalisability, no patients could be described as having mild depression and very few as having severe depression. This reflects the referral strategy from outpatient clinics, primary care and the community. Although about 55% of patients randomised came from a secondary care setting, all of these patients were outpatients. The remaining 45% of patients came from primary care or were self-referrals. It is important to note that the recruitment to this study was therefore very different from that of the positive proof-of-concept study of metyrapone augmentation of conventional antidepressants carried out by Jahn *et al.*,[33] which exclusively studied inpatients. The depression rating scale results further demonstrate differences between the current study and that of Jahn *et al.*[33] For example, the metyrapone group in the Jahn study[33] had a mean (SD) MADRS score at baseline of 31.5 (7.6) compared with 28.1 (5.5) in the

current study, and the corresponding scores on the HDRS were 26.6 (6.5) and 23.3 (3.9), although it should be noted the Jahn *et al.* study[33] utilised the 21-item version of the HDRS, in contrast with the 17-item version used here. It is of interest that baseline BDI scores in the metyrapone group in the current study were higher at 35.6 (10.9) than those seen in the Jahn *et al.* study[33] (8.4). High BDI scores and a high BDI/MADRS ratio have been associated with poor outcome in TRD.[123] Our patients therefore may have clinical characteristics known to be associated with worse outcome, which may link to the low response and remission rates seen in the current study.

There was evidence that many patients had a family history of depression and high levels of anxiety, as reflected in the CAS and STAI scores, as is commonly found in such patients in routine care in the NHS.

The dosing schedule for metyrapone selected for the ADD Study was 500 mg twice a day, administered in the morning and at noon. The total daily dose matches that given in previous studies.[33] The rationale for the timing of administration was physiological, to coincide with the portion of the day associated with the highest cortisol concentrations. The study is novel in its length of follow-up and in its inclusion of a range of mechanistic studies exploring how such treatments might work.

Interpretation

The overall conclusion from this study is that metyrapone augmentation of antidepressants is not efficacious for moderately depressed patients in outpatient clinics and in the community who have failed to respond to at least two antidepressants. There was no obvious benefit to its use either on the primary outcome or over the period of follow-up and this negative result extended to other secondary outcomes, such as the CAS, the BDI and measures of QoL. Metyrapone was generally well tolerated by this group as there were no SAEs attributable to it and AEs occurring during its use were as common in the placebo group.

As discussed above, there is preliminary evidence for metyrapone being efficacious in TRD in a smaller RCT carried out on more severely ill inpatients in Germany.[33] It may be that the drug works only for more severe depression but, unfortunately, our study was not powered for this type of subgroup analysis. Alternatively, the positive result in the Jahn *et al.* study[33] is a type I error. This TRD population was characterised by increased exposure to childhood adversity (compared with the control subjects) and normal HPA axis function. This finding accords with the existing literature in chronic populations;[124,125] there is evidence that the former predicts non-response to treatment.[126] None of baseline HPA axis function, change in CAR in response to drug treatment or severity of childhood trauma predicted clinical response to metyrapone.

The impact of metyrapone on the HPA axis is relevant to its potential mechanism of action. In the Jahn *et al.* study,[33] it exerted a minimal impact on circulating cortisol levels, this is presumably because cortisol activation of mineralocorticoid receptor (MR) and glucocorticoid receptor (GR) is a major homeostatic regulator of HPA axis activity. Thus in the Jahn *et al.* study[33] – using the same dose as this study – there was a marked increase in other HPA axis hormones, notably ACTH and 11-deoxycortisol. The HPA axis response to metyrapone will be more evident when the 11-deoxycortisol data are available. The issue of whether or not HPA axis dysregulation at baseline would be expected to predict response to metyrapone is of interest and relevance. Translational studies from our group show that antiglucocorticoid strategies – such as GR antagonists – engender an increase in the prefrontal cortex serotonin response to SSRIs, even in rats with normal HPA axis function,[17] suggesting that HPA axis dysregulation is not a requirement for therapeutic gain from HPA axis manipulation.

There are very few data specifically on the neuropsychology of treatment-resistant depressed groups. Those that have been conducted suggest that deficits are restricted to tests of processing speed.[127] In this TRD sample, we see broad deficits in verbal and visuospatial working memory compared with healthy control subjects. Deficits in attention were not general and, instead, were restricted to the executive control of attention. These findings are indicative of impairment in effortful processing in TRD. Cognitive deficits have important clinical and functional consequences for depressed patients. They are associated with poor functional outcomes, particularly employment, and with a greater risk of relapse and poorer clinical outcomes.[128] Patients also complain of the effects of limited cognitive function and find this dispiriting. For these reasons, strategies to improve cognitive function in depression are an important target for intervention. In patients with bipolar depression, the glucocorticoid receptor antagonist has been shown to improve cognitive function, specifically spatial working memory.[129] However, this strategy was ineffective in schizophrenia.[130] In the current study, metyrapone was not associated with improvement in cognitive measures but, as has previously been demonstrated,[125] there was no evidence of elevated corticosteroids in this group of patients with TRD. It may be that antiglucocorticoid therapies improve cognition only when dysregulation of the HPA axis leads to elevated corticosteroid levels, as is found in bipolar disorder[131] but not TRD[125] or schizophrenia.[130] However, remediation strategies are showing more promise. Recent work has demonstrated that cognitive remediation with supplemental 'internet-based homework' led to significant improvements in aspects of cognition, especially attention/processing speed and verbal memory. Improved cognition was also predictive of improvement in general functioning.[132]

Patients at baseline are less accurate in identifying emotions and more likely to wrongly attribute emotions. The lack of an emotion by group interaction means that there is a lack of evidence of emotional bias, but the greatest differences were seen in the negative emotions of anger, disgust and fear. These results are broadly in line with the literature for depression and confirm for the first time that similar findings are evident in TRD. The only differential effect of treatment allocation was seen for anger accuracy; however, this is likely to be a chance finding and partly accounted for by baseline differences. One limitation is the relatively small numbers due to loss of participants for technical reasons and dropout from follow-up.

Patients at baseline had poorer immediate recall than healthy control subjects, with evidence of a negative emotional bias: negative words were recalled equally by control subjects but positive and negative words less so. Control subjects did not show a positive bias in this study. No significant results for accuracy of delayed recognition were seen, although the pattern was the same as for recall. However, there was a strong trend for patients to wrongly identify words as having been seen before, driven most strongly by negative words, suggesting there may be an alteration in emotional recognition discrimination or bias. No significant effects of treatment were seen. The delayed recognition data need to be viewed as exploratory, given the lack of a significant group or group by valence interaction. A limitation to the task is that delayed recall was not measured.

Word valence affected rate of responding to targets, with neutral targets and emotional distractors having the slowest rate and emotional targets and neutral distractors the fastest rates. This suggests that words having an emotional valance capture attentional resources, indicating that the task appeared to work. The main finding is that patients are slower to react to positive targets in the presence of neutral and negative distractors, suggesting positive words capture less attentional resource and have less salience for patients than they do for control subjects. There was no evidence that negative words captured more attentional resource in patients and no positive bias was seen in control subjects (GoPos and GoNeg values are very similar). No effects of treatment were seen. The lack of a significant group by valence interaction makes the results for the GoPos condition exploratory. The task itself may have been limited by its complexity and number of conditions presented.

The results show similar hyperactivation in the amygdala in response to processing faces showing negative emotions in patients with TRD, as previously reported in the literature.[34] However, these hyperactivations (compared with healthy individuals) were present only in the condition that did not require explicit processing of the emotional content of the face. This may indicate that healthy individuals were better able to ignore the displayed negative emotion when the task could be performed without explicitly referring to the emotionality in the face, whereas patients may have had the tendency to process the emotion regardless of its relevance for the task. When explicit processing of facial emotions was required, patients with TRD and healthy control subjects showed similar levels of amygdala activation.

Contrary to previous work,[101] there were no differential activations in the amygdala between the three conditions of the EEM task across all participants. In general, brain activation differences between the conditions were relatively small. Similarly, there were no differences in amygdala responses between patients and control subjects. However, this might be related to the fact that we also did not see any evidence for improved recognition memory for emotional over neutral words (results not shown here). In addition, our implementation of the EEM task failed to elicit reliable amygdala activity. Various methodological alterations in this task that were necessary for the current study and these negative findings may be related to those alterations. Even though we tried to eliminate the amount of time that participants had to explicitly rehearse the words, contrary to previous studies, they were still aware that they would be tested on their memory performance during the encoding, which may have altered their processing of the words. Additionally, as a result of time constraints, the number of words presented to the participants was smaller than in other studies.

During the n-back task, the patients with TRD showed a reduced activation of the dorsolateral prefrontal cortex compared with the control subjects. In the emotional memory task, the TRD group had less activation of the posterior cingulate cortex while encoding both positive and neutral images, and reduced anterior cingulate cortex activation while retrieving positive images compared with neutral images. While retrieving images irrespective of valence, the TRD group demonstrated an increased activation of the posterior insula compared with control subjects. The decreased activation of the prefrontal cortex during the n-back task was demonstrated here in a treatment-resistant population. This study also suggests that there is an alteration in the functioning of the cingulate and insular cortex in the encoding and retrieval of positive emotional memories in this group of patients with TRD. Further study is needed to determine whether or not there is alteration of negative emotional memory processing, and what relationship this has to the degree of depression or a failure to respond to treatment.

The primary finding of the EEG substudies is that a putative marker of long-term potentiation, the enhancement of VEPs following tetanic visual stimulation, is impaired in patients with TRD. Following tetanus, a significant increase in the magnitude of the N1 and N1b components of VEPs were seen in control subjects, in line with previous published data demonstrating this phenomenon.[69,133–136] However, no significant enhancement was seen in depressed patients in line with a previous report in non-treatment-refractory depressed patients,[137] patients with bipolar disorder[138] and patients with schizophrenia.[139] This suggests that impaired visual cortical plasticity is a phenomenon shared across a range of psychiatric disorders. Further research is required to examine if this represents a wider impairment in cortical plasticity and its aetiology and is of pathophysiological significance.

In summary, regarding the mechanistic outcomes, HPA axis measures were not different between the groups and did not predict any clinical response to metyrapone. A number of abnormalities in the neuropsychology and neural activity were found in the TRD population. Many of these were known from previous work in depression but this study showed them to be present in TRD and quantified them. No specific effects of metyrapone on these variables have been shown.

Clinical implications and future research directions

The firm conclusion based on the primary outcome measure of the ADD Study is that in the population of depressed patients studied, the addition of metyrapone to standard serotonergic antidepressants is ineffective. A question remains as to why the ADD Study was negative, without any suggestion whatsoever of an effect of metyrapone at any time point, when an effect of antiglucocorticoid treatment is supported by preclinical data[17] and with a previous German RCT of metyrapone augmentation being positive.[33] This may relate to the nature of the patients studied by Jahn *et al.*[33] They studied inpatients, which are usually, but not always, severely and acutely ill. We included many primary care patients and outpatients with a high degree of treatment refractoriness and chronicity of depressive symptoms. A study of HPA axis function in chronic depressed patients suggests that no abnormality is found,[125] contrary to findings in patients with more acute illnesses.[18,19] Similarly, the ADD Study patients do not demonstrate HPA axis abnormalities, at least as indexed by the CAR. Chronic depression has been shown previously to be associated with normal HPA axis function,[125] and it has been argued that normocortisolaemic depression may predict non-response to biological treatments and hence may predict the development of a chronic illness. Conversely, it has been suggested that the initial hypercortisolaemia of depression may normalise with time in patients who continue to demonstrate symptoms; hence normal cortisol levels may be a consequence or a cause of chronic TRD. This is an important issue for future research to clarify. In addition, it would be of interest to explore the efficacy of metyrapone augmentation in acutely depressed patients, particularly those with demonstrable hypercortisolaemia.

The research questions above notwithstanding, the clinical implication of the ADD Study is that management of treatment refractory and chronically depressed patients should not include the use of metyrapone as an augmentation strategy.

Acknowledgements

Contributions of authors

All authors reviewed, revised and approved the final version of the manuscript.

I Nicol Ferrier* (Professor, Psychiatry) was the chief investigator.

Ian M Anderson (Professor, Psychiatry) was co-applicant, a PI for the Manchester site, and was involved in design of the study and recruitment.

Jane Barnes (Trial Manager) was involved in all aspects of the study's management and monitoring.

Peter Gallagher (Lecturer, Psychology) conducted the review of all neuropsychological measurements used in the study.

Heinz CR Grunze (Professor, Psychiatry) was co-applicant, a PI for the Newcastle site, and was involved in design of the study and recruitment.

Peter M Haddad (Consultant, Psychiatry) was co-applicant, a PI for the Manchester site, and was involved in design of the study and recruitment.

Allan O House (Professor, Psychiatry) was co-applicant, a PI for the Leeds site, and was involved in design of the study and recruitment via the CLRN.

Tom Hughes (Consultant, Psychiatry) was co-applicant, a PI for the Leeds site, and was involved in design of the study and recruitment.

Adrian J Lloyd (Consultant, Psychiatry) was co-applicant, a PI for the Newcastle site, and was involved in the design of the imaging components of the study.

Chrysovalanto Mamasoula* (Statistician) conducted all of the statistics for the study.

Elaine McColl* (Professor, Health Services Research and Director of Newcastle Clinical Trials Unit) was co-applicant, a PI for the Newcastle site, and was involved in the design of the study and its protocols and governance.

Simon Pearce (Professor, Endocrinology) was co-applicant, a PI for the Newcastle site, and was involved in the design and supervision of the endocrine components of the study.

Najma Siddiqi (Consultant, Psychiatry) was a PI for the Bradford site, and was involved in the design of the study and recruitment.

Baxi Sinha (Consultant, Psychiatry) was a PI for the Durham/Teesside site, and was involved in the design of the study and recruitment.

Chris Speed (Senior Trial Manager) was involved in the study's governance and management.

Nick Steen* (Statistician) was co-applicant, and designed and oversaw all of the statistics for the study.

June Wainwright (service user and carer) was co-applicant, a PI for the Newcastle site, and was involved in design of the study and recruitment and liaison with the MHRN.

Stuart Watson* (Senior Lecturer, Psychiatry) was co-applicant, a PI for the Newcastle site, and was involved in design of the study, particularly reviewing its endocrine aspects and recruitment.

Fiona H Winter (Service User) was co-applicant, a PI for the Newcastle site and was involved in design of the study and recruitment.

R Hamish McAllister-Williams* (Reader, Psychiatry) was co-applicant, a PI for the Newcastle site, and was involved in design of the study, particularly reviewing its EEG aspects and recruitment.

The ADD Study Team: Tanefa Apekey, Sam Bulmer, Stephanie Clutterbuck, Andreas Finkelmeyer, John Hiley, Brett Jones, Sophie Landa, Faye Ryles, Paul Sigalas, Eleanor Smith, Lucy Stevens, Roger Sturrock, Cathy Symonds, Kate Williams and Sarah Yates.

(*Denotes members of group who wrote first draft.)

This project was supported by the Efficacy and Mechanism Evaluation (EME) programme** (reference number 08/43/39). It is funded by the Medical Research Council (MRC) and managed by the NIHR on behalf of the MRC–NIHR partnership. Further support was obtained from Northumberland and Tyne and Wear NHS Foundation Trust with funding for a Clinical Research Fellow and for the metyrapone. The study was sponsored by Northumberland and Tyne and Wear NHS Foundation Trust. The views expressed in this publication are those of the author(s) and not necessarily those of the MRC, NHS, NIHR or the Department of Health.

The authors would like to thank the following for their invaluable help: Isabel Adeyemi, Hamid Alhaj, Sally Barker, Jill Barlow, Rob Baskind, Vicky Bell, Karen Bibbings, Susan Bonner, Alexandros Chatziagorakis, Emily Clare, Lyndsey Dixon, Michael Dixon, Julie Doughty, Claire Farrow, Cheri Fletcher, Nicola Gill, John Grey, Wendy Hall, Marco Haring, Prakash Hosali, Shola Johnson, Alice Kennedy, Steve Lankshear, Fiona McKenzie, Elaine McMullen, Anna Massey, Carole Milburn, Zarina Mirza, Mohammad Musabir, Catherine O'Neil, Ella Roelant, Rebecca Savage, Frances Sherratt, Lisa Svennson, Nick Venters, Helen Watkinson, Stephen Wright and Rashmi Yadav. We would also like to express thanks for the support from the North East and North West hubs of the MHRN, Primary Care Research Networks and CLRNs, and the Manchester Wellcome Trust Clinical Research Facility. We would also like to thank members of the Trial Steering Committee (chairperson, Professor P Cowen) and the DMEC (chairperson, Dr D Christmas) for their support and guidance.

(**The EME programme is funded by the MRC and NIHR, with contributions from the CSO in Scotland and NISCHR in Wales and the HSC R&D Division, Public Health Agency in Northern Ireland. It is managed by the NIHR Evaluation, Trials and Studies Coordinating Centre (NETSCC) based at the University of Southampton.)

References

1. National Institute for Health and Clinical Excellence (NICE). *Depression: the Treatment and Management of Depression in Adults (Update)*. NICE Clinical Guideline 91. London: NICE; 2009.
2. Murray C, Lopez A. *The Global Burden of Disease*. Harvard, CT: Harvard University Press; 1996.
3. The Centre for Economic Performance's Mental Health Policy Group. *The Depression Report: A New Deal for Depression and Anxiety Disorders*. London: London School of Economics and Political Science; 2006.
4. Wulsin LR, Vaillant GE, Wells VE. A systematic review of the mortality of depression. *Psychosom Med* 1999;**61**:6–17. http://dx.doi.org/10.1097/00006842-199901000-00003
5. Anderson IM, Ferrier IN, Baldwin RC, Cowen PJ, Howard L, Lewis G, *et al.* Evidence-based guidelines for treating depressive disorders with antidepressants: a revision of the 2000 British Association for Psychopharmacology guidelines. *J Psychopharm* 2008;**22**:343–96. http://dx.doi.org/10.1177/0269881107088441
6. Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, *et al.* Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry* 2006;**163**:28–40. http://dx.doi.org/10.1176/appi.ajp.163.1.28
7. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, *et al.* Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry* 2006;**163**:1905–17. http://dx.doi.org/10.1176/ajp.2006.163.11.1905
8. Smith KA, Fairburn CG, Cowen PJ. Relapse of depression after rapid depletion of tryptophan. *Lancet* 1997;**349**:915–19. http://dx.doi.org/10.1016/S0140-6736(96)07044-4
9. Ruhe HG, Mason NS, Schene AH. Mood is indirectly related to serotonin, norepinephrine and dopamine levels in humans: a meta-analysis of monoamine depletion studies. *Mol Psychiatry* 2007;**12**:331–59. http://dx.doi.org/10.1038/sj.mp.4001949
10. Vallejo J, Gasto C, Catalan R, Bulbena A, Menchon JM. Predictors of antidepressant treatment outcome in melancholia: psychosocial, clinical and biological indicators. *J Affect Disord* 1991;**21**:151–62. http://dx.doi.org/10.1016/0165-0327(91)90036-R
11. McAllister-Williams RH, Massey AE, Fairchild G. Repeated cortisol administration attenuates the EEG response to buspirone in healthy volunteers: evidence for desensitization of the 5-HT1A autoreceptor. *J Psychopharmacol* 2007;**21**:826–32. http://dx.doi.org/10.1177/0269881107078292
12. Fairchild G, Leitch MM, Ingram CD. Acute and chronic effects of corticosterone on 5-HT1A receptor-mediated autoinhibition in the rat dorsal raphe nucleus. *Neuropharmacology* 2003;**45**:925–34. http://dx.doi.org/10.1016/S0028-3908(03)00269-7
13. Leitch MM, Ingram CD, Young AH, McQuade R, Gartside SE. Flattening the corticosterone rhythm attenuates 5-HT1 A autoreceptor function in the rat: relevance for depression. *Neuropsychopharmacology* 2003;**28**:119–25. http://dx.doi.org/10.1038/sj.npp.1300016
14. Man MS, Young AH, McAllister-Williams RH. Corticosterone modulation of somatodendritic 5-HT1A receptor function in mice. *J Psychopharm* 2002;**16**:245–52. http://dx.doi.org/10.1177/026988110201600310
15. Blier P. Pharmacology of rapid-onset antidepressant treatment strategies. *J Clin Psychiatry* 2001;**62**(Suppl. 15):12–17.

16. Gartside SE, Leitch MM, Young AH. Altered glucocorticoid rhythm attenuates the ability of a chronic SSRI to elevate forebrain 5-HT: implications for the treatment of depression. *Neuropsychopharmacology* 2003;**28**:1572–8. http://dx.doi.org/10.1038/sj.npp.1300201

17. Johnson DA, Grant EJ, Ingram CD, Gartside SE. Glucocorticoid receptor antagonists hasten and augment neurochemical responses to a selective serotonin reuptake inhibitor antidepressant. *Biol Psychiatry* 2007;**62**:1228–35. http://dx.doi.org/10.1016/j.biopsych.2007.05.003

18. Wong ML, Kling MA, Munson PJ, Listwak S, Licinio J, Prolo P, *et al.* Pronounced and sustained central hypernoradrenergic function in major depression with melancholic features: relation to hypercortisolism and corticotropin-releasing hormone. *Proc Natl Acad Sci USA* 2000;**97**:325–30. http://dx.doi.org/10.1073/pnas.97.1.325

19. Heuser I, Yassouridis A, Holsboer F. The combined dexamethasone/CRH test: a refined laboratory test for psychiatric disorders. *J Psychiatr Res* 1994;**28**:341–56. http://dx.doi.org/10.1016/0022-3956(94)90017-5

20. Young EA, Altemus M, Lopez JF, Kocsis JH, Schatzberg AF, DeBattista C, *et al.* HPA axis activation in major depression and response to fluoxetine: a pilot study. *Psychoneuroendocrinology* 2004;**29**:1198–204. http://dx.doi.org/10.1016/j.psyneuen.2004.02.002

21. Juruena MF, Pariante CM, Papadopoulos AS, Poon L, Lightman S, Cleare AJ. Prednisolone suppression test in depression: prospective study of the role of HPA axis dysfunction in treatment resistance. *Br J Psychiatry* 2009;**194**:342–9. http://dx.doi.org/10.1192/bjp.bp.108.050278

22. Ribeiro SC, Tandon R, Grunhaus L, Greden JF. The DST as a predictor of outcome in depression: a meta-analysis. *Am J Psychiatry* 1993;**150**:1618–29. http://dx.doi.org/10.1176/ajp.150.11.1618

23. Zobel AW, Nickel T, Sonntag A, Uhr M, Holsboer F, Ising M. Cortisol response in the combined dexamethasone/CRH test as predictor of relapse in patients with remitted depression. A prospective study. *J Psychiatr Res* 2001;**35**:83–94. http://dx.doi.org/10.1016/S0022-3956(01)00013-9

24. Appelhof BC, Huyser J, Verweij M, Brouwer JP, van Dyck R, Fliers E, *et al.* Glucocorticoids and relapse of major depression (dexamethasone/corticotropin-releasing hormone test in relation to relapse of major depression). *Biol Psychiatry* 2006;**59**:696–701. http://dx.doi.org/10.1016/j.biopsych.2005.09.008

25. Aubry JM, Gervasoni N, Osiek C, Perret G, Rossier MF, Bertschy G, *et al.* The DEX/CRH neuroendocrine test and the prediction of depressive relapse in remitted depressed outpatients. *J Psychiatr Res* 2007;**41**:290–4. http://dx.doi.org/10.1016/j.jpsychires.2006.07.007

26. Vreeburg SA, Hoogendijk WJ, DeRijk RH, van DR, Smit JH, Zitman FG, *et al.* Salivary cortisol levels and the 2-year course of depressive and anxiety disorders. *Psychoneuroendocrinology* 2013;**38**:1494–502. http://dx.doi.org/10.1016/j.psyneuen.2012.12.017

27. Gallagher P, Malik N, Newham J, Young AH, Ferrier IN, Mackin P. Antiglucocorticoid treatments for mood disorders. *Cochrane Database Syst Rev* 2008;**1**:CD005168. http://dx.doi.org/10.1002/14651858.CD005168.pub2

28. Ghadirian AM, Engelsmann F, Dhar V, Filipini D, Keller R, Chouinard G, *et al.* The psychotropic effects of inhibitors of steroid biosynthesis in depressed patients refractory to treatment. *Biol Psychiatry* 1995;**37**:369–75. http://dx.doi.org/10.1016/0006-3223(94)00150-2

29. O'Dwyer AM, Lightman SL, Marks MN, Checkley SA. Treatment of major depression with metyrapone and hydrocortisone. *J Affect Disord* 1995;**33**:123–8. http://dx.doi.org/10.1016/0165-0327(94)00082-K

30. Raven PW, O'Dwyer AM, Taylor NF, Checkley SA. The relationship between the effects of metyrapone treatment on depressed mood and urinary steroid profiles. *Psychoneuroendocrinology* 1996;**21**:277–86. http://dx.doi.org/10.1016/0306-4530(95)00057-7

31. Murphy BE, Ghadirian AM, Dhar V. Neuroendocrine responses to inhibitors of steroid biosynthesis in patients with major depression resistant to antidepressant therapy. *Can J Psychiatry* 1998;**43**:279–86.

32. Rogoz Z, Skuza G, Wojcikowski J, Daniel WA, Wrobel A, Dudek D, *et al.* Effect of metyrapone supplementation on imipramine therapy in patients with treatment-resistant unipolar depression. *Pol J Pharmacol* 2004;**56**:849–55.

33. Jahn H, Schick M, Kiefer F, Kellner M, Yassouridis A, Wiedemann K. Metyrapone as additive treatment in major depression: a double-blind and placebo-controlled trial. *Arch Gen Psychiatry* 2004;**61**:1235–44. http://dx.doi.org/10.1001/archpsyc.61.12.1235

34. Groenewold NA, Opmeer EM, de Jonge P, Aleman A, Costafreda SG. Emotional valence modulates brain functional abnormalities in depression: evidence from a meta-analysis of fMRI studies. *Neurosci Biobehav Rev* 2013;**37**:152–63. http://dx.doi.org/10.1016/j.neubiorev.2012.11.015

35. Demenescu LR, Renken R, Kortekaas R, van Tol M-J, Marsman JBC, van Buchem MA, *et al.* Neural correlates of perception of emotional facial expressions in out-patients with mild-to-moderate depression and anxiety. A multicenter fMRI study. *Psychol Med* 2011;**41**:2253–64. http://dx.doi.org/10.1017/S0033291711000596

36. Arce E, Simmons AN, Lovero KL, Stein MB, Paulus MP. Escitalopram effects on insula and amygdala BOLD activation during emotional processing. *Psychopharmacology* 2008;**196**:661–72. http://dx.doi.org/10.1007/s00213-007-1004-8

37. Alhaj H, Wisniewski G, McAllister-Williams RH. The use of the EEG in measuring therapeutic drug action: focus on depression and antidepressants. *J Psychopharm* 2011;**25**:1175–91. http://dx.doi.org/10.1177/0269881110388323

38. Ulrich G, Renfordt E, Zeller G, Frick K. Interrelation between changes in the EEG and psychopathology under pharmacotherapy for endogenous depression. A contribution to the predictor question. *Pharmacopsychiatry* 1984;**17**:178–83. http://dx.doi.org/10.1055/s-2007-1017433

39. Ulrich G, Frick K. A new quantitative approach to the assessment of stages of vigilance as defined by spatiotemporal EEG patterning. *Percept Mot Skills* 1986;**62**:567–76. http://dx.doi.org/10.2466/pms.1986.62.2.567

40. Suffin SC, Emory WH. Neurometric subgroups in attentional and affective disorders and their association with pharmacotherapeutic outcome. *Clin Electroencephalogr* 1995;**26**:76–83. http://dx.doi.org/10.1177/155005949502600204

41. Knott VJ, Telner JI, Lapierre YD, Browne M, Horn ER. Quantitative EEG in the prediction of antidepressant response to imipramine. *J Affect Disord* 1996;**39**:175–84. http://dx.doi.org/10.1016/0165-0327(96)00003-1

42. Bruder GE, Sedoruk JP, Stewart JW, McGrath PJ, Quitkin FM, Tenke CE. Electroencephalographic alpha measures predict therapeutic response to a selective serotonin reuptake inhibitor antidepressant: pre- and post-treatment findings. *Biol Psychiatry* 2008;**63**:1171–7. http://dx.doi.org/10.1016/j.biopsych.2007.10.009

43. Ulrich G, Haug HJ, Stieglitz RD, Fahndrich E. Are there distinct biochemical subtypes of depression? EEG characteristics of clinically defined on-drug responders and non-responders. *J Affect Disord* 1988;**15**:181–5. http://dx.doi.org/10.1016/0165-0327(88)90088-2

44. Bruder GE, Stewart JW, Tenke CE, McGrath PJ, Leite P, Bhattacharya N, *et al.* Electroencephalographic and perceptual asymmetry differences between responders and nonresponders to an SSRI antidepressant. *Biol Psychiatry* 2001;**49**:416–25. http://dx.doi.org/10.1016/S0006-3223(00)01016-7

45. Bochkarev VK, Avedisova AS, Liupaeva NV. Placebo to antidepressant effects ratio by electroencephalographic data. *Zh Nevrol Psikhiatr Im S S Korsakova* 2004;**104**:42–8.

46. Iosifescu DV, Greenwald S, Devlin P, Mischoulon D, Denninger JW, Alpert JE, *et al.* Frontal EEG predictors of treatment outcome in major depressive disorder. *Eur Neuropsychopharmacol* 2009;**19**:772–7. http://dx.doi.org/10.1016/j.euroneuro.2009.06.001

47. Mayberg HS. Limbic-cortical dysregulation: a proposed model of depression. *J Neuropsychiatry Clin Neurosci* 1997;**9**:471–81. http://dx.doi.org/10.1176/jnp.9.3.471

48. Brody AL, Saxena S, Silverman DH, Alborzian S, Fairbanks LA, Phelps ME, *et al.* Brain metabolic changes in major depressive disorder from pre- to post-treatment with paroxetine. *Psychiatry Res* 1999;**91**:127–39. http://dx.doi.org/10.1016/S0925-4927(99)00034-7

49. Saxena S, Brody AL, Ho ML, Zohrabi N, Maidment KM, Baxter LR Jr. Differential brain metabolic predictors of response to paroxetine in obsessive–compulsive disorder versus major depression. *Am J Psychiatry* 2003;**160**:522–32. http://dx.doi.org/10.1176/appi.ajp.160.3.522

50. Buchsbaum MS, Wu J, Siegel BV, Hackett E, Trenary M, Abel L, *et al.* Effect of sertraline on regional metabolic rate in patients with affective disorder. *Biol Psychiatry* 1997;**41**:15–22. http://dx.doi.org/10.1016/S0006-3223(96)00097-2

51. Davidson RJ, Irwin W, Anderle MJ, Kalin NH. The neural substrates of affective processing in depressed patients treated with venlafaxine. *Am J Psychiatry* 2003;**160**:64–75. http://dx.doi.org/10.1176/appi.ajp.160.1.64

52. Pizzagalli DA, Oakes TR, Davidson RJ. Coupling of theta activity and glucose metabolism in the human rostral anterior cingulate cortex: an EEG/PET study of normal and depressed subjects. *Psychophysiology* 2003;**40**:939–49. http://dx.doi.org/10.1111/1469-8986.00112

53. Pizzagalli D, Pascual-Marqui RD, Nitschke JB, Oakes TR, Larson CL, Abercrombie HC, *et al.* Anterior cingulate activity as a predictor of degree of treatment response in major depression: evidence from brain electrical tomography analysis. *Am J Psychiatry* 2001;**158**:405–15. http://dx.doi.org/10.1176/appi.ajp.158.3.405

54. Mulert C, Juckel G, Brunnmeier M, Leicht SK, Mergl R, Moller HJ, *et al.* Rostral anterior cingulate cortex activity in the theta band predicts response to antidepressive medication. *Clin EEG Neurosci* 2007;**38**:78–81. http://dx.doi.org/10.1177/155005940703800209

55. Korb AS, Hunter AM, Cook IA, Leuchter AF. Rostral anterior cingulate cortex theta current density and response to antidepressants and placebo in major depression. *Clin Neurophysiol* 2009;**120**:1313–19. http://dx.doi.org/10.1016/j.clinph.2009.05.008

56. Hegerl U, Juckel G. Intensity dependence of auditory evoked potentials as an indicator of central serotonergic neurotransmission: a new hypothesis. *Biol Psychiatry* 1993;**33**:173–87. http://dx.doi.org/10.1016/0006-3223(93)90137-3

57. O'Neill BV, Croft RJ, Nathan PJ. The loudness dependence of the auditory evoked potential (LDAEP) as an in vivo biomarker of central serotonergic function in humans: rationale, evaluation and review of findings. *Hum Psychopharmacol* 2008;**23**:355–70. http://dx.doi.org/10.1002/hup.940

58. Paige SR, Fitzpatrick DF, Kline JP, Balogh SE, Hendricks SE. Event-related potential amplitude/intensity slopes predict response to antidepressants. *Neuropsychobiology* 1994;**30**:197–201. http://dx.doi.org/10.1159/000119161

59. Paige SR, Hendricks SE, Fitzpatrick DF, Balogh S, Burke WJ. Amplitude/intensity functions of auditory event-related potentials predict responsiveness to bupropion in major depressive disorder. *Psychopharm Bull* 1995;**31**:243–8.

60. Gallinat J, Bottlender R, Juckel G, Munke-Puchner A, Stotz G, Kuss HJ, *et al.* The loudness dependency of the auditory evoked N1/P2-component as a predictor of the acute SSRI response in depression. *Psychopharmacology* 2000;**148**:404–11. http://dx.doi.org/10.1007/s002130050070

61. Lee TW, Yu YW, Chen TJ, Tsai SJ. Loudness dependence of the auditory evoked potential and response to antidepressants in Chinese patients with major depression. *J Psychiatry Neurosci* 2005;**30**:202–5.

62. Linka T, Muller BW, Bender S, Sartory G. The intensity dependence of the auditory evoked N1 component as a predictor of response to Citalopram treatment in patients with major depression. *Neurosci Lett* 2004;**367**:375–8. http://dx.doi.org/10.1016/j.neulet.2004.06.038

63. Linka T, Muller BW, Bender S, Sartory G, Gastpar M. The intensity dependence of auditory evoked ERP components predicts responsiveness to reboxetine treatment in major depression. *Pharmacopsychiatry* 2005;**38**:139–43. http://dx.doi.org/10.1055/s-2005-864126

64. Juckel G, Pogarell O, Augustin H, Mulert C, Muller-Siecheneder F, Frodl T, *et al.* Differential prediction of first clinical response to serotonergic and noradrenergic antidepressants using the loudness dependence of auditory evoked potentials in patients with major depressive disorder. *J Clin Psychiatry* 2007;**68**:1206–12. http://dx.doi.org/10.4088/JCP.v68n0806

65. Jaworska N, Blondeau C, Tessier P, Norris S, Fusee W, Blier P, *et al.* Response prediction to antidepressants using scalp and source-localized loudness dependence of auditory evoked potential (LDAEP) slopes. *Prog Neuropsychopharmacol Biol Psychiatry* 2013;**44**:100–7. http://dx.doi.org/10.1016/j.pnpbp.2013.01.012

66. Mulert C, Juckel G, Brunnmeier M, Karch S, Leicht G, Mergl R, *et al.* Prediction of treatment response in major depression: integration of concepts. *J Affect Disord* 2007;**98**:215–25. http://dx.doi.org/10.1016/j.jad.2006.07.021

67. Smith AP, Dolan RJ, Rugg MD. Event-related potential correlates of the retrieval of emotional and nonemotional context. *J Cogn Neurosci* 2004;**16**:760–75. http://dx.doi.org/10.1162/089892904970816

68. Smith AP, Henson RN, Dolan RJ, Rugg MD. fMRI correlates of the episodic retrieval of emotional contexts. *NeuroImage* 2004;**22**:868–78. http://dx.doi.org/10.1016/j.neuroimage.2004.01.049

69. Teyler TJ, Hamm JP, Clapp WC, Johnson BW, Corballis MC, Kirk IJ. Long-term potentiation of human visual evoked responses. *Eur J Neurosci* 2005;**21**:2045–50. http://dx.doi.org/10.1111/j.1460-9568.2005.04007.x

70. Hirata R, Togashi H, Matsumoto M, Yamaguchi T, Izumi T, Yoshioka M. Characterization of stress-induced suppression of long-term potentiation in the hippocampal CA1 field of freely moving rats. *Brain Res* 2008;**1226**:27–32. http://dx.doi.org/10.1016/j.brainres.2008.06.004

71. Rotllant D, Armario A. A single dose of metyrapone caused long-term dysregulation of the hypothalamic–pituitary–adrenal axis in the rat. *Neuroscience* 2005;**130**:427–34. http://dx.doi.org/10.1016/j.neuroscience.2004.09.007

72. Gallagher P, Watson S, Elizabeth DC, Young AH, Ferrier IN. Persistent effects of mifepristone (RU-486) on cortisol levels in bipolar disorder and schizophrenia. *J Psychiatr Res* 2008;**42**:1037–41. http://dx.doi.org/10.1016/j.jpsychires.2007.12.005

73. McAllister-Williams RH, Smith E, Anderson IM, Barnes J, Gallagher P, Grunze HC, *et al.* Study protocol for the randomised controlled trial: antiglucocorticoid augmentation of anti-Depressants in Depression (The ADD Study). *BMC Psychiatry* 2013;**13**:205. http://dx.doi.org/10.1186/1471-244X-13-205

74. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979;**134**:382–9. http://dx.doi.org/10.1192/bjp.134.4.382

75. Juhasz G, Chase D, Pegg E, Downey D, Toth ZG, Stones K, *et al. CNR1* gene is associated with high neuroticism and low agreeableness and interacts with recent negative life events to predict current depressive symptoms. *Neuropsychopharmacology* 2009;**34**:2019–27. http://dx.doi.org/10.1038/npp.2009.19

76. Srivastava S, John OP, Gosling SD, Potter J. Development of personality in early and middle adulthood: set like plaster or persistent change? *J Pers Soc Psychol* 2003;**84**:1041–53. http://dx.doi.org/10.1037/0022-3514.84.5.1041

77. Brugha T, Bebbington P, Tennant C, Hurry J. The List of Threatening Experiences: a subset of 12 life event categories with considerable long-term contextual threat. *Psychol Med* 1985;**15**:189–94. http://dx.doi.org/10.1017/S003329170002105X

78. Sarason IG, Sarason BR, Sherin EN, Pierce GR. A brief measure of social support: practical and theoretical implications. *J Soc Pers Relat* 1987;**4**:497–510. http://dx.doi.org/10.1177/0265407587044007

79. Bernstein DP, Fink L, Handelsman L, Foote J, Lovejoy M, Wenzel K, *et al.* Initial reliability and validity of a new retrospective measure of child abuse and neglect. *Am J Psychiatry* 1994;**151**:1132–6. http://dx.doi.org/10.1176/ajp.151.8.1132

80. Nolen-Hoeksema S, Larson J, Grayson C. Explaining the gender difference in depressive symptoms. *J Pers Soc Psychol* 1999;**77**:1061–72. http://dx.doi.org/10.1037/0022-3514.77.5.1061

81. Williams JB, Kobak KA, Bech P, Engelhardt N, Evans K, Lipsitz J, *et al.* The GRID-HAMD: standardization of the Hamilton Depression Rating Scale. *Int Clin Psychopharmacol* 2008;**23**:120–9. http://dx.doi.org/10.1097/YIC.0b013e3282f948f5

82. Carmody TJ, Rush AJ, Bernstein I, Warden D, Brannan S, Burnham D, *et al.* The Montgomery Asberg and the Hamilton ratings of depression: a comparison of measures. *Eur Neuropsychopharm* 2006;**16**:601–11. http://dx.doi.org/10.1016/j.euroneuro.2006.04.008

83. Snaith RP, Baugh SJ, Clayden AD, Husain A, Sipple MA. The Clinical Anxiety Scale: an instrument derived from the Hamilton Anxiety Scale. *Br J Psychiatry* 1982;**141**:518–23. http://dx.doi.org/10.1192/bjp.141.5.518

84. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry* 1961;**4**:561–71. http://dx.doi.org/10.1001/archpsyc.1961.01710120031004

85. Kendall PC, Finch AJ Jr., Auerbach SM, Hooke JF, Mikulka PJ. The State-Trait Anxiety Inventory: a systematic evaluation. *J Consult Clin Psychol* 1976;**44**:406–12. http://dx.doi.org/10.1037/0022-006X.44.3.406

86. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. *Br J Psychiatry* 1978;**133**:429–35. http://dx.doi.org/10.1192/bjp.133.5.429

87. EuroQol Group. EuroQol: a new facility for the measurement of health-related quality of life. *Health Policy* 1990;**16**:199–208. http://dx.doi.org/10.1016/0168-8510(90)90421-9

88. 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.

89. Wust S, Federenko I, Hellhammer DH, Kirschbaum C. Genetic factors, perceived chronic stress, and the free cortisol response to awakening. *Psychoneuroendocrinology* 2000;**25**:707–20. http://dx.doi.org/10.1016/S0306-4530(00)00021-4

90. Gallagher P, Leitch MM, Massey AE, McAllister-Williams RH, Young AH. Assessing cortisol and dehydroepiandrosterone (DHEA) in saliva: effects of collection method. *J Psychopharmacol* 2006;**20**:643–9. http://dx.doi.org/10.1177/0269881106060585

91. Otte C, Lenoci M, Metzler T, Yehuda R, Marmar CR, Neylan TC. Effects of metyrapone on hypothalamic–pituitary–adrenal axis and sleep in women with post-traumatic stress disorder. *Biol Psychiatry* 2007;**61**:952–6. http://dx.doi.org/10.1016/j.biopsych.2006.08.018

92. Fan J, McCandliss BD, Sommer T, Raz A, Posner MI. Testing the efficiency and independence of attentional networks. *J Cogn Neurosci* 2002;**14**:340–7. http://dx.doi.org/10.1162/089892902317361886

93. Posner MI, Petersen SE. The attention system of the human brain. *Annu Rev Neurosci* 1990;**13**:25–42. http://dx.doi.org/10.1146/annurev.ne.13.030190.000325

94. Kessels RPC, Postma A, de Haan EHF. Object Relocation: a program for setting up, running, and analyzing experiments on memory for object locations. *Behav Res Methods Instrum Comput* 1999;**31**:423–8. http://dx.doi.org/10.3758/BF03200721

95. Kessels RPC, Postma A, Kappelle LJ, de Haan EHF. Spatial memory impairment in patients after tumour resection: evidence for a double dissociation. *J Neurol Neurosurg Psychiatry* 2000;**69**:389–91. http://dx.doi.org/10.1136/jnnp.69.3.389

96. Lezak MD, Howieson DB, Loring DW. *Neuropsychological Assessment*. 4th edn. New York, NY: Oxford University Press; 2004.

97. Anderson IM, Shippen C, Juhasz G, Chase D, Thomas E, Downey D, *et al.* State-dependent alteration in face emotion recognition in depression. *Br J Psychiatry* 2011;**198**:302–8. http://dx.doi.org/10.1192/bjp.bp.110.078139

98. Tranter R, Bell D, Gutting P, Harmer C, Healy D, Anderson IM. The effect of serotonergic and noradrenergic antidepressants on face emotion processing in depressed patients. *J Affect Disord* 2009;**118**:87–93. http://dx.doi.org/10.1016/j.jad.2009.01.028

99. Roiser JP, Sahakian BJ. Hot and cold cognition in depression. *CNS Spectr* 2013;**18**:139–49. http://dx.doi.org/10.1017/S1092852913000072

100. Hariri AR, Bookheimer, SY, Mazziotta JC. Modulating emotional responses: effects of a neocortical network on the limbic system. *NeuroReport* 2000;**11**:43–8. http://dx.doi.org/10.1097/00001756-200001170-00009

101. Kensinger EA, Corkin S. Two routes to emotional memory: distinct neural processes for valence and arousal. *Proc Natl Acad Sci USA* 2004;**101**:3310–15. http://dx.doi.org/10.1073/pnas.0306408101

102. Ekman P, Friesen WV. *Pictures of Facial Affect*. Palo Alto, CA: Consulting Psychologists Press; 1976.

103. Amunts K, Kedo O, Kindler M, Pieperhoff P, Mohlberg H, Shah NJ, *et al.* Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps. *Anat Embryol* 2005;**210**:343–52. http://dx.doi.org/10.1007/s00429-005-0025-5

104. Whalley HC, McKirdy J, Romaniuk L, Sussmann J, Johnstone EC, Wan HI, *et al.* Functional imaging of emotional memory in bipolar disorder and schizophrenia. *Bipolar Disord* 2009;**11**:840–56. http://dx.doi.org/10.1111/j.1399-5618.2009.00768.x

105. Koychev I, McMullen K, Lees J, Dadhiwala R, Grayson L, Perry C, *et al.* A validation of cognitive biomarkers for the early identification of cognitive enhancing agents in schizotypy: a three-center double-blind placebo-controlled study. *Eur Neuropsychopharmacol* 2012;**22**:469–81. http://dx.doi.org/10.1016/j.euroneuro.2011.10.005

106. Talairach J, Tournoux P. *Co-Planar Stereotaxic Atlas of the Human Brain*. New York, NY: Thieme; 1998.

107. Symonds CS, McKie S, Elliot R, Deakin JFW, Anderson IM. Detection of the acute effects of hydrocortisone in the hippocampus using pharmacological fMRI. *Eur Neuropsychopharmacol* 2012;**22**:867–74. http://dx.doi.org/10.1016/j.euroneuro.2012.03.008

108. American Psychiatric Association (APA). *Diagnostic and Statistical Manual of Mental Disorders*. 4th edn (DSM-IV). Washington, DC: APA; 1994.

109. Spitzer RL, Williams JB, Gibbon M, First MB. The Structured Clinical Interview for DSM-III-R (SCID). I: History, rationale, and description. *Arch Gen Psychiatry* 1992;**49**:624–9. http://dx.doi.org/10.1001/archpsyc.1992.01820080032005

110. Hamilton M. Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol* 1967;**6**:278–96. http://dx.doi.org/10.1111/j.2044-8260.1967.tb00530.x

111. Fava M. Diagnosis and definition of treatment-resistant depression. *Biol Psychiatry* 2003;**53**:649–59. http://dx.doi.org/10.1016/S0006-3223(03)00231-2

112. Hazari H, Christmas D, Matthews K. The clinical utility of different quantitative methods for measuring treatment resistance in major depression. *J Affect Disord* 2013;**150**:231–6. http://dx.doi.org/10.1016/j.jad.2013.03.030

113. Van Breukelen GJ. ANCOVA versus change from baseline: more power in randomized studies, more bias in nonrandomized studies [corrected]. *J Clin Epidemiol* 2006;**59**:920–5. http://dx.doi.org/10.1016/j.jclinepi.2006.02.007

114. Senn S. Change from baseline and analysis of covariance revisited. *Stat Med* 2006;**25**:4334–44. http://dx.doi.org/10.1002/sim.2682

115. Parzen M, Lipsitz SR, Dear KBG. Does clustering affect the usual test statistics of no treatment effect in a randomized clinical trial? *Biom J* 1998;**40**:385–402. http://dx.doi.org/10.1002/(SICI)1521-4036(199808)40:4<385::AID-BIMJ385>3.3.CO;2-R

116. Kahan BC, Morris TP. Improper analysis of trials randomised using stratified blocks or minimisation. *Stat Med* 2012;**31**:328–40. http://dx.doi.org/10.1002/sim.4431

117. Pruessner Jens C, Kirschbaum C, Meinlschmid G, Hellhammer, DH. Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology* 2003;**28**:916–31. http://dx.doi.org/10.1016/S0306-4530(02)00108-7

118. Bruder GE, Stewart JW, Hellerstein D, Alvarenga JE, Alschuler D, McGrath PJ. Abnormal functional brain asymmetry in depression: evidence of biologic commonality between major depression and dysthymia. *Psychiatry Res* 2012;**196**:250–4. http://dx.doi.org/10.1016/j.psychres.2011.11.019

119. Symonds CS, Gallagher P, Thompson JM, Young A. Effects of the menstrual cycle on mood, neurocognitive and neuroendocrine function in healthy premenopausal women. *Psychol Med* 2004;**34**:93–102. http://dx.doi.org/10.1017/S0033291703008535

120. Skinner EI, Fernandes MA. Neural correlates of recollection and familiarity: a review of neuroimaging and patient data. *Neuropsychologia* 2007;**45**:2163–79. http://dx.doi.org/10.1016/j.neuropsychologia.2007.03.007

121. Spaniol J, Davidson PSR, Kim ASN, Han H, Moscovitch M, Grady CL. Event-related fMRI studies of episodic encoding and retrieval: meta-analyses using activation likelihood estimation. *Neuropsychologia* 2009;**47**:1765–79. http://dx.doi.org/10.1016/j.neuropsychologia.2009.02.028

122. Owen AM, McMillan KM, Laird AR, Bullmore E. N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Hum Brain Map* 2005;**25**:46–59. http://dx.doi.org/10.1002/hbm.20131

123. Rane LJ, Fekadu A, Wooderson S, Poon L, Markopoulou K, Cleare AJ. Discrepancy between subjective and objective severity in treatment-resistant depression: prediction of treatment outcome. *J Psychiatr Res* 2010;**44**:1082–7. http://dx.doi.org/10.1016/j.jpsychires.2010.03.020

124. Douglas KM, Porter RJ. The effect of childhood trauma on pharmacological treatment response in depressed inpatients. *Psychiatry Res* 2012:**18**;1–4. http://dx.doi.org/10.1016/j.psychres.2012.06.015

125. Watson S, Gallagher P, Del-Estal D, Hearn A, Ferrier IN, Young AH. Hypothalamic–pituitary–adrenal axis function in patients with chronic depression. *Psychol Med* 2002;**32**:1021–8. http://dx.doi.org/10.1017/S0033291702005998

126. Nanni V, Uher R, Danese A. Childhood maltreatment predicts unfavorable course of illness and treatment outcome in depression: a meta-analysis. *Am J Psychiatry* 2012;**169**:141–51. http://dx.doi.org/10.1176/appi.ajp.2011.11020335

127. Moreines JL, McClintock SM, Kelley ME, Holtzheimer PE, Mayberg HS. Neuropsychological function before and after subcallosal cingulate deep brain stimulation in patients with treatment-resistant depression. *Depress Anxiety* 2014;**31**:690–8. http://dx.doi.org/10.1002/da.22263

128. Porter RJ, Bourke C, Gallagher P. Neuropsychological impairment in major depression: its nature, origin and clinical significance. *Aust N Z J Psychiatry* 2007;**41**:115–28. http://dx.doi.org/10.1080/00048670601109881

129. Watson S, Gallagher P, Porter RJ, Smith MS, Herron LJ, Bulmer S, *et al.* A randomized trial to examine the effect of mifepristone on neuropsychological performance and mood in patients with bipolar depression. *Biol Psychiatry* 2012;**72**:943–9. http://dx.doi.org/10.1016/j.biopsych.2012.05.029

130. Gallagher P, Watson S, Smith MS, Ferrier IN, Young AH. Effects of adjunctive mifepristone (RU-486) administration on neurocognitive function and symptoms in schizophrenia. *Biol Psychiatry* 2005;**57**:155–61. http://dx.doi.org/10.1016/j.biopsych.2004.10.017

131. Watson S, Gallagher P, Ritchie JC, Ferrier IN, Young AH. Hypothalamic–pituitary–adrenal axis function in patients with bipolar disorder. *Br J Psychiatry* 2004;**184**:496–502. http://dx.doi.org/10.1192/bjp.184.6.496

132. Bowie CRI, Gupta M, Holshausen K, Jokic R, Best M, Milev R. Cognitive remediation for treatment-resistant depression: effects on cognition and functioning and the role of online homework. *J Nerv Ment Dis* 2013;**201**:680–5. http://dx.doi.org/10.1097/NMD.0b013e31829c5030

133. Clapp WC, Zaehle T, Lutz K, Marcar VL, Kirk IJ, Hamm JP, *et al.* Effects of long-term potentiation in the human visual cortex: a functional magnetic resonance imaging study. *Neuroreport* 2005;**16**:1977–80. http://dx.doi.org/10.1097/00001756-200512190-00001

134. Clapp WC, Kirk IJ, Hamm JP, Shepherd D, Teyler TJ. Induction of LTP in the human auditory cortex by sensory stimulation. *Eur J Neurosci* 2005;**22**:1135–40. http://dx.doi.org/10.1111/j.1460-9568.2005.04293.x

135. McNair NA, Clapp WC, Hamm JP, Teyler TJ, Corballis MC, Kirk IJ. Spatial frequency-specific potentiation of human visual-evoked potentials. *Neuroreport* 2006;**17**:739–41. http://dx.doi.org/10.1097/01.wnr.0000215775.53732.9f

136. Ross RM, McNair NA, Fairhall SL, Clapp WC, Hamm JP, Teyler TJ, *et al.* Induction of orientation-specific LTP-like changes in human visual evoked potentials by rapid sensory stimulation. *Brain Res Bull* 2008;**76**:97–101. http://dx.doi.org/10.1016/j.brainresbull.2008.01.021

137. Teyler TJ, Cavus I. Depressed neuroplasticity in major depressive disorder? *Biol Psychiatry* 2007;**62**:371–2. http://dx.doi.org/10.1016/j.biopsych.2007.07.008

138. Elvsashagen T, Moberget T, Boen E, Boye B, Englin NO, Pedersen PO, *et al.* Evidence for impaired neocortical synaptic plasticity in bipolar II disorder. *Biol Psychiatry* 2012;**71**:68–74. http://dx.doi.org/10.1016/j.biopsych.2011.09.026

139. Cavus I, Reinhart RM, Roach BJ, Gueorguieva R, Teyler TJ, Clapp WC, *et al.* Impaired visual cortical plasticity in schizophrenia. *Biol Psychiatry* 2012;**71**:512–20. http://dx.doi.org/10.1016/j.biopsych.2012.01.013

Appendix 1 Statistical analysis plan

Background

Antidepressant drugs have established efficacy versus placebo in clinical trials but in naturalistic settings, many patients have unsatisfactory outcomes and may be referred to be suffering from treatment-refractory depression (TRD). However, a Cochrane review demonstrated the efficacy of antiglucocorticoid augmentation of antidepressants in patients with TRD [1]. In the Cochrane meta-analysis of antiglucocorticoid treatments, the largest effect size was seen with metyrapone [1], a cortisol synthesis inhibitor which crosses the blood–brain barrier. In addition, a successful study has been conducted by Jahn and colleagues [2] in a centre in Germany, with 63 depressed inpatients. Patients receiving metyrapone were significantly more likely to respond than those receiving placebo, with an effect size of $d = 0.63$ without serious side effects. The study will extend research in this area by exploring the translatability of the findings to a primary and secondary care, UK NHS population and also extending the follow-up period for 6 months after metyrapone treatment has ceased, in comparison with pre-existing antiglucocorticoid trials [1]. Additionally, an important element of the proposed study is the further exploration of the mechanisms underlying the beneficial effects of metyrapone augmentation on antidepressant response using neuropsychological outcome measures which may be markers of its mechanism of action.

This document should be read in conjunction with the (to be) published ADD Study Protocol paper.

Study objectives

Primary clinical objective

- The primary objective is to determine whether metyrapone (500 mg twice a day) for 21 days is efficacious in augmenting conventional serotonergic antidepressants in TRD in a UK NHS primary and secondary care setting. This is to be assessed by Montgomery–Åsberg Depression Rating Scale (MADRS) scores measured at baseline and two weeks post treatment (week +5 from randomisation), comparing patients treated with metyrapone to those treated with placebo.

Secondary clinical objectives

1. To determine the clinical effect size at two weeks post-completion of treatment of a three-week course of metyrapone (vs placebo) augmentation of antidepressants in depressed patients who have failed to respond to at least two courses of antidepressants, in primary care and psychiatric outpatient clinics in the UK.
2. To assess whether the response is sustained for up to 21 weeks post cessation of metyrapone.
3. To assess whether metyrapone augmentation improves patients' quality of life.
4. To assess the tolerability and safety of metyrapone augmentation in a large sample taken from a representative population of psychiatric outpatients and primary care patients with TRD.

Mechanistic objectives

In the full sample

1. To assess whether metyrapone changes patients' HPA axis function.
2. To assess whether changes in HPA axis function are seen 2 weeks after metyrapone treatment.
3. To assess whether the change in HPA axis function correlates with clinical response.
4. To assess whether baseline HPA axis function predicts clinical response.
5. To assess if type and severity of childhood trauma [as assessed by the Childhood Trauma Questionnaire (CTQ) scores] predicts clinical response.
6. To assess whether the change in, or baseline assessment of, HPA axis function has specific genetic underpinning and whether this relates to, or explains better, the clinical response.

In subsamples

1. To determine the nature and extent of neuropsychological abnormalities in patients with TRD compared with healthy control subjects.
2. To determine whether neuropsychological performance improves with metyrapone treatment.
3. To determine whether clinical improvements with metyrapone augmentation correlate with improvements in neuropsychological function and particularly (a) changes in cortisol sensitive tasks, e.g. spatial working memory and/or (b) 5-HT sensitive tasks, e.g. emotional processing.
4. To compare visual cortical long-term potentiation (LTP) in patients with TRD and healthy control subjects.
5. To examine if visual cortical LTP is altered by treatment with metyrapone.
6. To compare emotional source memory performance, and its electrophysiological underpinnings, in patients with TRD and healthy control subjects.
7. To examine if emotional source memory performance, and its electrophysiological underpinnings, is altered by treatment with metyrapone.
8. To determine whether EEG predictors of conventional antidepressant response, and in particular specific predictors of response to serotonergic antidepressants, predict response to metyrapone augmentation.
9. To compare the degree of activation of the amygdala in response to emotional faces and words in patients with TRD and healthy control subjects.
10. To determine whether abnormalities in the degree of activation of the amygdala in response to emotional faces pretreatment is modified by metyrapone treatment.
11. To determine if changes in the degree of activation of the amygdala with metyrapone post-treatment correlate with clinical outcome.
12. To compare neural activity during episodic and working memory tasks in patients with TRD and healthy control subjects.
13. To determine whether abnormalities in neural activity during episodic and working memory tasks is modified by metyrapone treatment.
14. To determine if changes in neural activity during episodic and working memory tasks correlate with clinical outcome.
15. To determine whether patients with TRD have altered hippocampal response to hydrocortisone compared with healthy control subjects.
16. To determine if the hippocampal response to hydrocortisone is modified in patients by metyrapone treatment.
17. To determine the effect of genetic variations on cognitive and emotional processing and on predictors of treatment outcome.
18. To examine the degree to which differences in ratings of depression between the scales used to assess mood at baseline predict response to metyrapone, including a focus on the degree of disparity between observer and self-ratings of mood.

Study design

Patient randomised double-blind randomised controlled parallel trial with half the patients receiving metyrapone and half the patients receiving a placebo.

Sample size

Two groups of 85 patients give 90% power to detect an effect size of 0.5 assuming a type 1 error rate of 5%. Allowing for 10% attrition during the trial, 95 per group would need to be randomised; 190 in total aged 18–65 years. 55 healthy control subjects, aged 18–55, with no personal or family history of mental illness are included.

The target number in the patient group was amended following a delay in starting recruitment and a slow early rate of recruitment, following agreement from the Data Monitoring and Ethics Committee (DMEC), the funder and the sponsor. A reduced power of 80% was agreed. Using the same assumptions as above, and allowing for 10% attrition prior to the primary outcome measure, the aim was to randomise 70 per group; 140 in total.

Study population

This is more fully defined in the ADD Study protocol publication. Patients included had to meet criteria including i) having a DSM-IV confirmed diagnosis of a major depressive episode; ii) current symptoms of moderate to severe severity as defined by Hamilton Depression Rating Scale-17 item (HDRS17) score of ≥ 18 at week –2 and 0; iii) have TRD as defined by a Massachusetts General Hospital Treatment Resistant Depression (MGH-TRD) staging score of 2–10; iv) currently taking a serotonergic antidepressant.

CONSORT diagram

The flow of patients through the study will be described with the aid of a CONSORT diagram.

Characteristics of groups at baseline

The baseline characteristics of the study population will be summarised for each study group.

Compliance with medication and withdrawal from study medication

Adherence to medication will be assessed using measures of 11-deoxycortisol.

A 95% confidence interval for the relative odds of withdrawing from the two treatment arms will be reported.

Completeness of data

The level of missing data will be reported for each of the study measures of outcome.

General analysis considerations

Unless otherwise stated, all analyses will use two-tailed tests where appropriate with significance level set at 5%. Statistical packages used will include SPSS, STATA and R, as well as specialised packages for MRI/EEG analysis, etc.

Primary outcome

The primary outcome will be assessed using an intention-to-treat analysis (patients allocated to the group to which they were randomised regardless of whether they received the treatment).

Primary analysis

- The estimated difference between the treatment groups at week 5 adjusting for randomisation strata will be estimated using a mixed model. The dependent variable will be MADRS score at week 5. The baseline score at week 0 will be included as a covariate. Differences between centres will be incorporated as a random effect; differences between primary and secondary care origin of the patients will be included as a fixed effect. (Statistical packages: SPSS, STATA, R.)

Secondary analyses

- The difference between MADRS at week 5 and MADRS immediately before treatment (week 0) in two arms will be compared using an independent sample *t*-test utilising all patients in the study in an intention-to-treat analysis.
- The impact of missing data due to patients declining to participate in data collection will be assessed using joint modelling of 'time to withdrawal from data collection' and MADRS scores.
- In addition to analysing the data using the conventional MADRS score, the above primary and secondary analyses will be repeated utilising the addition 'atypical' depression items (rating hypersomnia and increased appetite). In this scoring of the MADRS, the highest score from the conventional sleep and atypical sleep items, and conventional appetite and atypical appetite items, will be used to calculate the total MADRS score.
- The intention-to-treat analyses above will be repeated in a 'per-protocol analysis' excluding patients who discontinued treatment during the three week treatment period. A second 'per-protocol' definition will utilise the 11-deoxycortisol measure of apparent concordance with treatment.

Secondary clinical outcomes

1. In relation to clinical effect size:

    i. Response (defined as 50% reduction in MADRS) and remission (defined as MADRS $\leq$ 10) rates at week 5 will be analysed using logistic regression procedures analogous to the mixed models described previously. Variation between centres will be incorporated as a random effect; differences between primary and secondary care origin of patients and the difference between those randomised to metyrapone and those randomised to placebo will be included as fixed effects. Results will be presented in the form of a 95% confidence interval for the odds of response/remission.

    ii. In relation to CAS, BDI and STAI measures:

    – The CAS, BDI and STAI will be analysed by fitting a sequence of nested models. The baseline model will include the score at week 5 as the dependent variable, the baseline score as a covariate and the difference between study groups as a fixed effect. Variation between centres will then be included as a random effect and the difference between primary and secondary care origin as a fixed effect. Nested models will be compared using a likelihood ratio test. The impact of compliance with treatment will be evaluated using the same approach (based on the measures of compliance previously defined). Results will be given in the form of unadjusted and adjusted 95% confidence intervals for the effect of treatment with metyrapone.

2. In relation to whether the response is sustained:

    i. The persistence of change in MADRS, CAS, BDI and STAI scores will be assessed using repeated measures ANOVA, adjusting for randomisation strata of centre and primary vs. secondary care, using all of the data points available.

    ii. The speed of response and speed of remission (both defined using MADRS as described above) with metyrapone vs. placebo will be assessed using survival analysis; time to response and time to

remission will be analysed using a Cox proportional hazards regression model. Results will be given in the form of a 95% confidence interval for the hazard ratio between the two groups.

3. In relation to whether metyrapone improves patient's quality of life:

   i. The EQ-5D data will be analysed by fitting a sequence of nested models. The baseline model will include the score at week 5 as the dependent variable, the baseline score as a covariate and the difference between study groups as a fixed effect. Variation between centres will then be included as a random effect and the difference between primary and secondary care origin as a fixed effect. Nested models will be compared using a likelihood ratio test. The impact of compliance with treatment will be evaluated using the same approach (based on the measures of compliance previously defined). Results will be given in the form of unadjusted and adjusted 95% confidence intervals for the effect of treatment with metyrapone.

4. In relation to how well metyrapone is tolerated:

   i. The Toronto Side Effects Scale scores will be analysed utilising data from all patients who had at least one dose of study medication. A second 'per-protocol' definition will utilise the 11-deoxycortisol measure of apparent concordance with treatment.

      – The difference between TSES score at week 3 and TSES score immediately before treatment (week 0) in two arms will also be compared using an independent sample *t*-test utilising all patients in the study in an intention-to-treat analysis.
      – The change in TSES score over time will be compared using a repeated measures ANOVA including all available time points.
      – Symptoms described by patients will be grouped into clinically relevant clusters and analysed. This will include a specific focus on suicidal and any changes in this during or following treatment.

   ii. The Young Mania Rating Scale (YMRS) will be analysed in a similar way to that described for the TSES.

Mechanistic outcomes

In the full patient sample

1. To assess whether metyrapone changes patients' HPA axis function:

   i. HPA function will be assessed by calculating the Area Under the Curve (AUC) for the CAR data, measuring the peak change in cortisol on waking by comparing the initial waking sample with the maximum, as well as examining the diurnal change in cortisol using the CAR and the 11 p.m. saliva sample. The effect of metyrapone will be compared with placebo in relation to HPA axis function at week 3 (end of metyrapone treatment) using ANOVA covarying for the baseline at week 0.

2. To assess whether changes in HPA axis function are seen 2 weeks after metyrapone treatment:

   i. HPA axis function assessed as above comparing data at week 5 covarying for week 0.

3. To assess whether the change in HPA axis function correlates with clinical response:

   i. The correlation between change in HPA axis function (CAR AUC and maximal peak, and diurnal cortisol, both at week 3 and week 5 compared with week 0) with change in MADRS, CAS, BDI, STAI, EQ-5D (week 5 compared with week 0) will be examined.

4. To assess whether baseline HPA axis function predicts clinical response:

    i. The correlation between HPA axis function (CAR AUC and maximal peak, and diurnal cortisol) with change in MADRS, CAS, BDI, STAI, EQ-5D (week 5 compared with week 0) will be examined.

5. To assess if type and severity of childhood trauma [as assessed by the Childhood Trauma Questionnaire (CTQ) scores] predicts clinical response:

    i. The correlation between CTQ with change in MADRS, CAS, BDI, STAI, EQ-5D (week 5 compared with week 0) will be examined.

6. To assess whether the change in, or baseline assessment of, HPA axis function has specific genetic underpinning and whether this relates to, or explains better, the clinical response:

    i. The relationship between clinical response (as described above) and HPA axis function compared with a range of candidate polymorphisms will be examined.

In subsamples of patients

1. To determine the nature and extent of neuropsychological abnormalities in patients with TRD compared with healthy control subjects:

    i. This will be studied using the various outcome measures from the neuropsychological tasks employed, comparing the patients with healthy control subjects. Analysis will include controlling for age, gender and IQ (as assessed using NART).

2. To determine whether neuropsychological performance improves with metyrapone treatment:

    i. This will be studied using the various outcome measures from the neuropsychological tasks employed, comparing performance at week 5 to that at week 0, in patients treated with metyrapone compared with those treated with placebo. This will be conducted both using an intention-to-treat population and using a 'per-protocol' population.

3. To determine whether clinical improvements with metyrapone augmentation correlate with improvements in neuropsychological function and particularly (a) changes in cortisol sensitive tasks, e.g. spatial working memory and/or (b) 5-HT sensitive tasks, e.g. emotional processing:

    i. The correlation between changes in neuropsychological function (as measured above) with changes in clinical symptoms (MADRS, CAS, BDI, STAI) at week 5 compared with week 0 will be examined.

4. To compare visual cortical long-term potentiation (LTP) in patients with TRD and healthy control subjects:

    i. The degree and persistence of LTP induced in visual cortex by the tetanic presentation of a flashing checkerboard in patients (at week 0) will be compared with that in healthy control subjects. Analysis will include controlling for age, gender and IQ (as assessed using NART).

5. To examine if visual cortical LTP is altered by treatment with metyrapone:

    i. The degree and persistence of LTP at week 5 compared with week 0 will be examined in patients treated with metyrapone and placebo. This will be conducted both using an intention-to-treat population and using a 'per-protocol' population.

6. To compare emotional source memory performance, and its electrophysiological underpinnings, in patients with TRD and healthy control subjects:

    i. Behaviour responses, and related ERPs, recorded from patients at week 0 will be compared with those from healthy control subjects. Analysis will include controlling for age, gender and IQ (as assessed using NART).

7. To examine if emotional source memory performance, and its electrophysiological underpinnings, is altered by treatment with metyrapone:

    i. The Behaviour responses, and related ERPs, at week 5 compared with week 0 will be examined in patients treated with metyrapone and placebo. This will be conducted both using an intention-to-treat population and using a 'per-protocol' population.

8. To determine whether EEG predictors of conventional antidepressant response, and in particular specific predictors of response to serotonergic antidepressants, predict response to metyrapone augmentation:

    i. The correlation between a variety of EEG variables (including LDAEP slope, alpha power, alpha hemispheric asymmetry, anterior cingulate located theta power, and combinations of these), recorded at week 0 will be correlated with the change in clinical symptoms (MADRS, CAS, BDI, STAI) at week 5 compared with week 0 will be examined.

9. To compare the degree of activation of the amygdala in response to emotional faces or emotional words in patients with TRD and healthy control subjects:

    i. The fMRI BOLD response to emotional stimuli, particularly in relation to activity in the amygdala, recorded from patients at week 0 will be compared with those from healthy control subjects. Analysis will include controlling for age, gender and IQ (as assessed using NART).

10. To determine whether abnormalities in the degree of activation of the amygdala pretreatment is modified by metyrapone treatment:

    i. The change in fMRI BOLD response to emotional stimuli, particularly in relation to activity in the amygdala, at week 5 compared with week 0, in patients treated with metyrapone compared with those treated with placebo. This will be conducted both using an intention-to-treat population and using a 'per-protocol' population.

11. To determine if changes in the degree of activation of the amygdala with metyrapone post treatment correlate with clinical outcome:

    i. The correlation between change in amygdala activation at week 5 compared with week 0 with change in MADRS, CAS, BDI, STAI, EQ-5D (week 5 compared with week 0) will be examined.

12. To compare neural activity during episodic and working memory tasks in patients with TRD and healthy control subjects:

    i. The fMRI BOLD response during episodic and working memory tasks recorded from patients at week 0 will be compared with those from healthy control subjects. Analysis will include controlling for IQ (as assessed using NART) and if required age and gender.

13. To determine whether abnormalities in neural activity during episodic and working memory tasks is modified by metyrapone treatment:

    i. The change in fMRI BOLD response during episodic and working memory tasks at week 5 compared with week 0, in patients treated with metyrapone compared with those treated with placebo.

14. To determine if changes in neural activity during episodic and working memory tasks correlate with clinical outcome:

    i. The correlation between change in BOLD response during episodic and working tasks at week 5 compared with week 0 with change in MADRS, CAS, BDI, STAI, EQ-5D (week 5 compared with week 0) will be examined.

15. To determine whether patients with TRD have altered hippocampal response to hydrocortisone compared with healthy control subjects.

    i. The phMRI BOLD response related to the administration of hydrocortisone in patients at week 0 will be compared with those from healthy control subjects.

16. To determine if the hippocampal response to hydrocortisone is modified in patients by metyrapone treatment:

    i. The change in phMRI BOLD response related to the administration of hydrocortisone at week 5 compared with week 0, in patients treated with metyrapone
    ii. The change in phMRI BOLD response related to the administration of hydrocortisone at week 5 in patients treated with metyrapone compared with those treated with placebo.

17. To determine the effect of genetic variations on cognitive and emotional processing and on predictors of treatment outcome using candidate genes with functional effects on HPA axis and monoamine function.
18. To examine the degree to which differences in ratings of depression between the scales used to assess mood at baseline predict response to metyrapone, including a focus on the degree of disparity between observer and self-ratings of mood:

    i. The difference in ratings of mood using the HDRS, MADRS and BDI at baseline will be correlated with the change in mood (assessed using MADRS) at week 5 compared with week 0.

Additional cross-sectional analyses

These will include the following analyses to explore the relationship between various factors at baseline (week 0) in the patient population (and where data are available in the healthy control subjects as well). The list is not comprehensive and further analyses may be undertaken in the light of new developments or for hypothesis testing or generating:

1. The relationship between HPA axis function (as defined above) and CTQ will be examined.
2. The relationship between HPA axis function (as defined above) and neuropsychological function, particularly spatial working memory, will be examined.
3. The relationship between HPA axis function (as defined above) and LTP will be examined.
4. The relationship between HPA axis function (as defined above) and emotional source memory (both behavior and related ERPs) will be examined.

5. The relationship between HPA axis function (as defined above) and the amygdala response to emotional faces.
6. The relationship between HPA axis function (as defined above) and the fMRI BOLD response in episodic and working memory tasks.
7. The relationship between HPA axis function (as defined above) and the fMRI bold response to hydrocortisone.
8. The relationship between LTP and neuropsychological function.

Subgroup analysis

The data will be analysed by dividing the sample into those with severe depression (defined as MADRS ≥ 30 at baseline week 0) or less-severe depression.

Missing data

Scores for the various scales used will be calculated in accordance with their author's instructions. In the absence of any rule for missing data imputation will be used provided at least half the items in any scale have been completed (imputed missing value = mean value of non-missing items). Missing survival date will be dealt with by using the date of censoring and the status of the patient at that point. The secondary outcomes are being analysed using mixed models that make use of all data available. The analysis is adjusted to take into account that some subjects have provided more information than others. If there is a difference in survival rates we will have more 'missing' outcome data in one group than another. If this situation occurs the data will be analysed using joint modelling as described above. Thus estimates of functional health status and quality of life will be adjusted for differences in survival rates. There will be no other data imputation.

Safety

The number of serious adverse events in each group will be reported along with a 95% confidence interval for the relative odds (between the two treatment groups) of at least one event being reported.

The number of non-serious adverse events will be compared using a negative binomial regression model. Results will be given in the form of a 95% confidence interval for the incidence rate ratio.

Implementation

A SAP Implementation Group will be formed to oversee the analysis of all ADD-related data. The analysis of the clinical outcomes will be undertaken by the study team based in Newcastle. The mechanistic and cross-sectional analyses will be undertaken by various individuals/centres with the agreement of the SAP Implementation Group.

Signatures

The above plan has been read and agreed by:

Name	Signature	Date
Ian Nicol Ferrier		
Ian Nick Steen		
Chrysovalanto Mamasoula		
Hamish McAllister-Williams		

References

1. Gallagher P, Malik N, Newham J, Young AH, Ferrier IN, Mackin P. Antiglucocorticoid treatments for mood disorders. *Cochrane Database Syst Rev* 2008;CD005168.
2. Jahn H, Schick M, Kiefer F, Kellner M, Yassouridis A, Wiedemann K. Metyrapone as additive treatment in major depression: a double-blind and placebo-controlled trial. *Arch Gen Psychiatry* 2004;**61**:1235–1244.
3. Bruder GE, Sedoruk JP, Stewart JW, McGrath PJ, Quitkin FM, Tenke CE. Electroencephalographic alpha measures predict therapeutic response to a selective serotonin reuptake inhibitor antidepressant: pre- and post-treatment findings. *Biol Psychiatry* 2008;**6**:1171–1177.
4. Mulert C, Juckel G, Brunnmeier M, Leicht SK, Mergl R, Moller HJ *et al.* Rostral anterior cingulate cortex activity in the theta band predicts response to antidepressive medication. *Clin EEG Neurosci* 2007;**38**:78–81.
5. Mulert C, Juckel G, Brunnmeier M, Karch S, Leicht G, Mergl R *et al.* Prediction of treatment response in major depression: integration of concepts. *J Affect Disord* 2007;**98**:215–225.
6. Henderson R, Diggle P and Dobson A. Joint modelling of longitudinal measurements and event type data. *Biostatistics* 2000;**1**:464–480.
7. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet* 2007;81.

Glossary of abbreviations

5-HT	5-hydroxytryptophan, serotonin
11-HSD	11β-hydroxysteroid dehydrogenase
ACTH	Adrenocorticotrophic hormone
AE	Adverse events
ANT	Attentional Network Test
AUCg	Area Under the Curve Ground
AUCi	Area Under the Curve Increase
BDI	Beck Depression Inventory
BFI-44	Big Five Inventory 44
BOLD	Blood Oxygen Level Dependent
BSE	Between-search errors
CANTAB	Cambridge Neuropsychological Test Automated Battery
CAR	Cortisol awakening response
CAS	Clinical Anxiety Scale
CBT	Cognitive behavioural therapy
CI	Chief investigator
COM	Combined memory condition
CRF	Case report form
CTA	Clinical trial authorisation
CTQ	Childhood Trauma Questionnaire
CTIMP	Clinical trial of an investigational medicinal product
DHEA	Dehydroepiandrosterone
DMEC	Data Monitoring and Ethics Committee
DSM-IV	Diagnostic and Statistical Manual-Fourth Edition

ECMT	Emotional Categorization and Memory Test
EEG	Electroencephalogram
EEM	Emotional enhancement of memory
EL	Emotional labelling
EM	Emotional matching
EME	Efficacy Mechanism and Evaluation
ERP	Event-related potential
ESMT	Emotional Source Memory Task
EQ-5D	EuroQol Quality of Life scale
FEER	Facial Emotional Expression Recognition
FEP	Facial emotion processing
fMRI	Functional magnetic resonance imaging
FWE	Family Wise Error
GCP	Good Clinical Practice
GR	Glucocorticoid receptors
GRID-HAMD	GRID Hamilton Depression Rating Scale
HDRS17	Hamilton Depression Rating Scale-17 item
HPA	Hypothalamic–pituitary–adrenal
HV	Healthy volunteers
HTA	Health technology assessment
IAPS	International Affective Picture Set
IMP	Investigational medicinal product
LDAEP	Loudness dependency of auditory evoked potentials
LTE	List of Life Threatening Experiences
LTP	Long-term potentiation
MADRS	Montgomery-Åsberg Depression Research Scale
MCV	Mean cell volume
MGH	Massachusetts General Hospital
MHRA	Medicines and Healthcare Regulatory Agency
MHRN	Mental Health Research Network
MR	Mineralocorticoid receptors
MRC	Medical Research Council
MTA	Material Transfer Agreement
NRES	National Research Ethics Service
NICE	National Institute of Clinical Excellence
NIHR	National Institutes of Health Research
OLB	Object-location binding
OLM	Object-location memory
PCRN	Primary Care Research Network
phMRI	Pharmacological functional magnetic resonance imaging

PI	Principal investigator
PIC	Participant Identification Centres
POM	Position-only memory
QoL	Quality of life
R&D	Research and development
RAVLT	Rey Auditory Verbal Learning Test
REC	Research Ethics Committee
ROI	Region of interest
RRS	Ruminative Responses Scale
RT	Reaction time
SAE	Serious adverse event
SARs	Serious adverse reactions
SD	Standard deviation
SEM	Standard error of the mean
SM	Shape matching
SmPC	Summary of major product characteristics
SOP	Standard operating procedure
SCID	Structured Clinical Interview for DSM
SCQ	Social Circumstances Questionnaire
SPM8	Statistical Parametric Mapping
SSI	Site-specific information
SSRI	Selective serotonin reuptake inhibitor
STAI	State–Trait Anxiety Inventory
SUSAR	Suspected unexpected serious adverse reaction
SWM	Spatial working memory
TSC	Trial Steering Committee
TSES	Toronto Side Effects Scale
VEP	Visual Evoked Potential
WSE	Within-search errors
YMRS	Young Mania Rating Scale

Appendix 2 Numbers analysed and descriptive statistics

TABLE 47 Mean MADRS scores at follow-up visits in each randomised group

	Placebo			Metyrapone		
Visit	Mean	*n*	SD	Mean	*n*	SD
Baseline	28.1	82	5.4	27.7	83	6.7
Week 3	20.8	77	9.9	22.6	72	10.9
Week 5	22.4	74	10.6	21.7	69	10.9
Week 8	22.6	66	10.8	21.2	62	10.4
Week 16	20.5	61	11.5	21.4	54	11.0
Week 24	20.0	58	11.6	21.0	46	11.1

TABLE 48 Mean BDI scores at follow-up visits in each randomised group

	Placebo			Metyrapone		
Visit	Mean	*n*	SD	Mean	*n*	SD
Baseline	34.8	81	10.3	35.6	81	10.9
Week 3	28.7	75	14.0	30.5	69	15.1
Week 5	29.6	72	14.5	27.9	67	15.3
Week 8	30.9	63	14.4	28.7	61	14.9
Week 16	28.0	61	15.7	30.5	52	14.2
Week 24	28.9	57	16.9	28.2	47	14.2

TABLE 49 Mean CAS scores at follow-up visits in each randomised group

	Placebo			Metyrapone		
Visit	Mean	*n*	SD	Mean	*n*	SD
Baseline	10.00	82	4.57	9.53	83	4.51
Week 3	7.40	77	5.21	8.46	72	6.16
Week 5	8.23	74	5.78	8.52	69	5.98
Week 8	8.53	66	5.59	8.71	62	6.02
Week 16	7.70	61	4.87	8.46	54	4.96
Week 24	7.43	58	5.29	7.87	46	5.72

TABLE 50 Mean state anxiety scores at follow-up visits in each randomised group

	Placebo			Metyrapone		
Visit	Mean	*n*	SD	Mean	*n*	SD
Baseline	41.09	81	5.83	42.83	81	6.49
Week 3	39.87	75	5.45	42.87	69	7.10
Week 5	40.26	72	6.02	42.13	67	6.88
Week 8	40.70	64	6.95	41.32	59	7.22
Week 16	41.17	58	7.32	42.25	52	6.14
Week 24	41.25	56	6.23	42.43	47	6.13

TABLE 51 Mean EQ-5D tariffs at each visit in each randomised group

	Placebo			Metyrapone		
Visit	Mean	*n*	SD	Mean	*n*	SD
Baseline	0.370	81	0.314	0.368	81	0.301
Week 3	0.464	75	0.334	0.440	69	0.332
Week 5	0.468	72	0.354	0.502	67	0.347
Week 8	0.447	64	0.357	0.471	60	0.337
Week 16	0.445	60	0.359	0.491	52	0.334
Week 24	0.443	58	0.370	0.487	47	0.348

TABLE 52 Mean EQ-VAS scores at each visit for each of the randomised groups

	Placebo			Metyrapone		
Visit	Mean	*n*	SD	Mean	*n*	SD
Baseline	42.3	79	19.4	40.9	80	16.6
Week 3	47.3	74	20.8	48.4	65	20.8
Week 5	47.5	71	22.1	50.6	63	21.9
Week 8	49.4	64	22.0	49.4	59	20.0
Week 16	47.0	61	23.9	48.2	51	20.0
Week 24	47.4	58	25.0	48.3	47	19.7

TABLE 53 Mean YMRS score at each visit for each of the randomised groups

	Placebo			Metyrapone		
Visit	Mean	*n*	SD	Mean	*n*	SD
Baseline	2.35	82	1.83	2.34	82	1.71
Week 3	1.92	77	2.19	2.03	72	1.67
Week 5	1.74	74	1.87	1.70	69	1.62
Week 8	1.74	66	1.88	1.55	62	1.73
Week 16	1.69	61	2.10	1.85	54	1.90
Week 24	1.80	59	2.10	1.54	46	1.75

EME
HS&DR
HTA
PGfAR
PHR

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

Published by the NIHR Journals Library