NIHR | National Institute for Health and Care Research

Journals Library

Check for updates

Health Technology Assessment

Volume 29 • Issue 12 • May 2025

ISSN 2046-4924

Clinical and cost-effectiveness of lithium versus quetiapine augmentation for treatment-resistant depression in adults: LQD a pragmatic randomised controlled trial

Jess Kerr-Gaffney, Zohra Zenasni, Kimberley Goldsmith, Nahel Yaziji, Huajie Jin, Alessandro Colasanti, John Geddes, David Kessler, R Hamish McAllister-Williams, Allan H Young, Alvaro Barrera, Lindsey Marwood, Rachael W Taylor, Helena Tee and Anthony J Cleare on behalf of the LQD Study Group

DOI 10.3310/YQVF5347

NIHR | National Institute for Health and Care Research

Health Technology Assessment

Check for updates

Extended Research Article

Clinical and cost-effectiveness of lithium versus quetiapine augmentation for treatment-resistant depression in adults: LQD a pragmatic randomised controlled trial

Jess Kerr-Gaffney[1] Zohra Zenasni[1] Kimberley Goldsmith[1] Nahel Yaziji[1] Huajie Jin[1] Alessandro Colasanti[2,3] John Geddes[4] David Kessler[5] R Hamish McAllister-Williams[6,7] Allan H Young[1,8] Alvaro Barrera[4,9] Lindsey Marwood[1] Rachael W Taylor[1] Helena Tee[1] and Anthony J Cleare[1,8*] on behalf of the LQD Study Group

[1]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
[2]Brighton and Sussex Medical School, University of Sussex, Brighton, UK
[3]Sussex Partnership NHS Foundation Trust, Sussex, UK
[4]Department of Psychiatry, University of Oxford, Oxford, UK
[5]Bristol Medical School, University of Bristol, Bristol, UK
[6]Translational and Clinical Research Institute, and Northern Centre for Mood Disorders, Newcastle University, Newcastle, UK
[7]Cumbria, Northumberland, Tyne and Wear NHS Foundation Trust, Newcastle, UK
[8]South London and Maudsley NHS Foundation Trust, London, UK
[9]Oxford Health NHS Foundation Trust, Oxford, UK

[*]Corresponding author anthony.cleare@kcl.ac.uk

Published May 2025
DOI: 10.3310/YQVF5347

This report should be referenced as follows:

Kerr-Gaffney J, Zenasni Z, Goldsmith K, Yaziji N, Jin H, Colasanti A, *et al*. Clinical and cost-effectiveness of lithium versus quetiapine augmentation for treatment-resistant depression in adults: LQD a pragmatic randomised controlled trial. *Health Technol Assess* 2025;**29**(12). https://doi.org/10.3310/YQVF5347

Health Technology Assessment

ISSN 2046-4924 (Online)

Impact factor: 3.5

A list of Journals Library editors can be found on the NIHR Journals Library website

Launched in 1997, *Health Technology Assessment* (HTA) has an impact factor of 3.5 and is ranked 30th (out of 174 titles) in the 'Health Care Sciences & Services' category of the Clarivate 2022 Journal Citation Reports (Science Edition). It is also indexed by MEDLINE, CINAHL (EBSCO Information Services, Ipswich, MA, USA), EMBASE (Elsevier, Amsterdam, the Netherlands), NCBI Bookshelf, DOAJ, Europe PMC, the Cochrane Library (John Wiley & Sons, Inc., Hoboken, NJ, USA), INAHTA, the British Nursing Index (ProQuest LLC, Ann Arbor, MI, USA), Ulrichsweb™ (ProQuest LLC, Ann Arbor, MI, USA) and the Science Citation Index Expanded™ (Clarivate™, Philadelphia, PA, USA).

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

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

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta.

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

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

HTA programme

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

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

This article

The research reported in this issue of the journal was funded by the HTA programme as award number HTA 14/222/02. The contractual start date was in May 2016. The draft manuscript began editorial review in June 2023 and was accepted for publication in July 2024. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' manuscript and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this article.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Newgen Digitalworks Pvt Ltd, Chennai, India (www.newgen.co).

Abstract

Background: Lithium and several atypical antipsychotics are the recommended first-line augmentation options for treatment-resistant depression; however, few studies have compared them directly, and none for longer than 8 weeks. Consequently, there is little evidence-based guidance for clinicians when choosing an augmentation option for patients with treatment-resistant depression.

Objectives: This trial examined whether it is more clinically and cost-effective to prescribe lithium or quetiapine augmentation therapy for patients with treatment-resistant depression over 12 months.

Design: This was a parallel group, multicentre, pragmatic, open-label superiority trial comparing the clinical and cost-effectiveness of lithium versus quetiapine augmentation of antidepressant medication in treatment-resistant depression. Participants were randomised 1 : 1 at baseline to the decision to prescribe either lithium or quetiapine.

Setting: Six National Health Service trusts in England.

Participants: Eligible participants were aged ≥ 18 years, met *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition criteria for major depressive disorder, scored ≥ 14 on the 17-item Hamilton Depression Rating Scale and whose depression had had an inadequate response to at least two therapeutic antidepressant treatment trials in the current episode, with a current antidepressant treatment at or above the therapeutic dose for ≥ 6 weeks. Patients with a history of psychosis or bipolar disorder were excluded. Patients were judged suitable for either treatment.

Interventions: After randomisation, pre-prescribing safety checks were undertaken as per standard care and trial clinicians decided whether to proceed with prescribing the allocated medication. Trial clinicians received recommendations for titration and dosing in line with current clinical guidelines; however, dosing regimens could be altered according to tolerability and response. Participants were followed up using weekly self-report questionnaires and 8-, 26- and 52-week research visits.

Main outcome measures: The co-primary outcome measures were depressive symptom severity over 52 weeks, measured weekly using the self-rated Quick Inventory of Depressive Symptomatology, and time to all-cause treatment discontinuation of the trial medication. Economic analyses compared costs between the two treatment arms over 52 weeks, from a National Health Service and Personal Social Services perspective, and a societal perspective.

Results: Two hundred and twelve participants were randomised, 107 to quetiapine and 105 to lithium. The quetiapine arm showed a significantly greater reduction in depressive symptoms than the lithium arm over 52 weeks (quetiapine vs. lithium area under the differences curve = −68.36, 95% confidence interval: −129.95 to −6.76, $p = 0.0296$). Median days to discontinuation did not significantly differ between the two arms (quetiapine = 365.0, interquartile range = 57.0–365.0, lithium = 212.0, interquartile range = 21.0–365.0), $p = 0.1196$. Quetiapine was more cost effective than lithium. Thirty-two serious adverse events were recorded, only one of which was deemed possibly related to the intervention (lithium).

Limitations: The trial was unblinded, therefore expectancies regarding the trial medications may have influenced the results. Further, there was substantial missing data for some of the secondary outcome measures.

Conclusions: As well as being more cost-effective, quetiapine may be a more clinically effective augmentation option for treatment-resistant depression.

Future work: Examining predictors of treatment response, including clinical, sociodemographic and biological factors, will help establish whether there are additional factors to consider when choosing an augmentation treatment for treatment-resistant depression.

Trial registration: This trial is registered as ISRCTN16387615.

Funding: This award was funded by the National Institute for Health and Care Research (NIHR) Health Technology Assessment programme (NIHR award ref: 14/222/02) and is published in full in *Health Technology Assessment*; Vol. 29, No. 12. See the NIHR Funding and Awards website for further award information.

Contents

List of tables

List of figures

List of supplementary material

Report Supplementary Material 1	Statistical analysis plan
Report Supplementary Material 2	Health economics analysis plan

Supplementary material can be found on the NIHR Journals Library report page (https://doi.org/10.3310/YQVF5347).

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

List of abbreviations

A&E	accident and emergency
AUC	area under the curve
AWP	Avon and Wiltshire Mental Health Partnership
BMI	body mass index
BNF	*British National Formulary*
CD-MCAR	covariate dependent missing completely at random
CEAC	cost-effectiveness acceptability curve
CGI	Clinical Global Impression
CSRI	Client Service Receipt Inventory
CTIMP	Clinical Trial of an Investigational Medicinal Product
DMEC	Data Monitoring and Ethics Committee
DSCT	Digit Symbol Coding Test
DSM-V	*Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition
ECG	electrocardiogram
EQ-5D	EuroQol-5 Dimensions
FDR	false discovery rate
GP	general practitioner
HAMD-17	Hamilton Depression Rating Scale-17 item
ICER	incremental cost-effectiveness ratio
INB	incremental net benefit
IR	immediate release
ITT	intention to treat
KCTU	King's Clinical Trials Unit
LQD	Lithium versus Quetiapine for Depression
MADRS	Montgomery–Åsberg Depression Rating Scale
MAR	missing at random
MARS-5	5-item Medication Adherence Report Scale
MDD	major depressive disorder
MINI 7.0	Mini International Neuropsychiatric Interview, Version 7
MSM	Maudsley Staging Method
MSM-MTR	Maudsley Staging Model of Treatment Resistance, Modified Multi-Therapy resistance version
MTI	Maudsley Treatment Inventory
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health and Care Research
OHT	Oxford Health NHS Foundation Trust
PP	per protocol
PPI	patient and public involvement
PRISE	Patient Rated Inventory of Side Effects
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
QALY	quality-adjusted life-year
QIDS-SR	self-rated Quick Inventory of Depressive Symptomatology
RMST	restricted mean survival time
SAE	serious adverse event
SAP	statistical analysis plan
SAR	serious adverse reaction
SLaM	South London and Maudsley
SPFT	Sussex Partnership NHS Foundation Trust
SUSAR	suspected unexpected serious adverse reaction
TEWV	Tees, Esk and Wear Valleys
TRD	treatment-resistant depression
TSC	Trial Steering Committee
WSAS	Work and Social Adjustment Scale
WTP	willingness to pay
XR	extended release

Plain language summary

Many people with depression experience limited benefits from initial antidepressant medications and psychological therapies. There is some evidence suggesting that adding another type of medication to an antidepressant might be beneficial for reducing depressive symptoms. Lithium and quetiapine are two of the most commonly used treatments in the National Health Service to add on to antidepressants. This study aimed to test whether adding lithium or adding quetiapine was more effective in reducing symptoms of depression, and whether there were differences in how long patients stayed on the added medications. We also compared the cost of the treatments to the National Health Service and to society (e.g. time off work due to health problems). Adults whose depression had not responded to at least two trials of antidepressants at the recommended dose and duration were eligible to take part. Two hundred and twelve participants were included in the study and had an equal chance of being prescribed either lithium or quetiapine.
We assessed participants over 12 months, including weekly assessments of their depression and several visits to the hospital. Over the 12-month study period, we found that adding quetiapine to patients' antidepressant treatment led to a greater improvement in symptoms of depression than adding lithium. There was no difference in how long the two groups stayed on the medications, or the number of side effects: around 50% of those who started taking lithium and 39% of those who started taking quetiapine stopped taking the medication within 12 months, usually due to side effects. Quetiapine provided a greater benefit for patients at a lower cost than lithium. The results suggest that overall, adding quetiapine may be a better option than adding lithium for those who are still suffering with depression after taking two or more courses of antidepressants.

Scientific summary

Background

Major depressive disorder (MDD) is a highly prevalent and disabling illness. Between 20% and 50% of those with MDD do not respond to first- and second-line treatments, termed treatment-resistant depression (TRD). Clinical guidelines recommend augmentation with lithium or atypical antipsychotics as one treatment option for TRD. However, few studies have compared these options directly, and none have included a long-term follow-up, which is imperative given the long-term course of TRD.

Objectives

This trial aimed to examine whether it is more clinically and cost-effective to prescribe lithium or quetiapine augmentation therapy for patients with TRD over the course of 12 months.

Methods

This was a phase 4, 12-month, parallel-group, pragmatic, open-label, superiority trial comparing the clinical and cost-effectiveness of lithium versus quetiapine augmentation treatment to antidepressant medication in patients with TRD. Two arms were randomised 1 : 1 to the decision to prescribe either lithium or quetiapine, stratified by baseline depression severity, TRD severity and recruitment site. Trial clinicians received information on titration and dosing in line with current clinical guidelines. After randomisation, pre-prescribing safety checks were undertaken as per standard care and trial clinicians decided whether to proceed with prescribing the allocated medication. Subsequent decisions as to whether to continue treatment followed standard care guidelines and clinician judgement. Participants were followed up over 12 months, regardless of medication status.

Participants were recruited from six NHS mental health trusts across England. Participants were identified through secondary care clinics or consent for contact initiatives within these trusts, community and online advertisements, and primary care services. Inclusion criteria were: (1) under the care of a general practitioner and/or adult mental health service, (2) current episode of depression meeting *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition criteria for MDD, (3) a Hamilton Depression Rating Scale-17 item score ≥ 14 at screening, (4) aged ≥ 18 years, (5) meet criteria for TRD, defined as failing to adequately respond to at least two therapeutic antidepressant treatment trials in the current episode, (6) current antidepressant treatment at or above a therapeutic dose for ≥ 6 weeks. Exclusion criteria were: (1) a diagnosis of bipolar disorder or current psychosis; (2) adequate use of lithium or quetiapine during the current episode; (3) current use of another atypical antipsychotic, (4) known contraindication to lithium or quetiapine; (5) participation in another Clinical Trial of an Investigational Medicinal Product; (6) insufficient comprehension or attention to engage in trial procedures, (7) pregnancy, trying for pregnancy, or breastfeeding. Participants attended a screening and baseline visit, usually on the same day. Randomisation took place on the same day as the baseline visit. Participants attended visits at the hospital at weeks 8, 26 and 52, and completed weekly assessments via the True Colours app.

The primary outcome measures were depressive symptom severity over 52 weeks, measured weekly using the self-rated Quick Inventory of Depressive Symptomatology (QIDS-SR), and time to all-cause treatment discontinuation of the trial medication. Secondary outcome measures included clinician-rated depression severity (Montgomery–Åsberg Depression Rating Scale, MADRS), response and remission rates (MADRS), health-related quality of life (EuroQol-5 Dimensions, EQ-5D), work and social functioning (Work and Social Adjustment Scale, WSAS), adherence to the trial medication (5-item Medication Adherence Report Scale, MARS-5), weight, blood pressure, clinician-rated global improvement (Clinical Global Impression scale, CGI), side effects (Patient Rated Inventory of Side Effects, PRISE), time to initiation of the trial medication, time to initiation of any new intervention for depression and serious adverse events

(SAEs) between the two treatment arms. The MADRS and CGI were assessed by blind raters. Service use (for the economic analysis) was measured using the Client Service Receipt Inventory.

Primary efficacy analyses were conducted under intention-to-treat (ITT) and per-protocol (PP) assumptions, the latter including only those where the prescription was implemented by clinicians after the pre-prescription safety checks. The QIDS-SR outcome was analysed using a linear mixed model with weekly scores as the dependent variable, and treatment allocation, randomisation stratification variables, time and time by treatment interaction terms as explanatory variables, using an area under the curve (AUC) approach. The time to discontinuation of trial medication outcome was analysed using survival analysis methods. Cox regression models were used to estimate the hazard ratio for discontinuation, with treatment allocation and randomisation stratification variables as independent variables. Restricted mean survival time models were also used, since the Cox regression modelling showed evidence of non-proportional hazards. Time to initiation and time to new intervention for depression were also analysed in this way. Continuous secondary outcome measures were analysed similarly to the QIDS-SR outcome, using linear mixed models, but with data from weeks 8, 26 and 52 as dependent variables. Binary secondary outcomes were modelled using longitudinal logistic mixed models with data from weeks 8, 26 and 52 as dependent variables. Both types of outcome models had treatment allocation, time, time by treatment interaction terms and stratification variables as explanatory variables.

Sensitivity analyses were performed on the two primary outcomes; effects were re-estimated for: (1) participants who had a therapeutic treatment trial of the trial medication, (2) participants who reported themselves adherent to the trial medication, defined as 80% or greater adherence on the MARS-5 during the time they were taking the medication, (3) participants who were both prescribed and reported initiating treatment, (4) scenarios evaluating the effect of departures from the missing at random (MAR) assumption, and (5) subsets defined as being before and after the COVID-19 pandemic started.

The economic analysis compared costs between the two treatment arms over 52 weeks and was conducted under the ITT assumption. The primary analysis was conducted from an NHS and Personal Social Services (PSS) perspective (service use and drug costs), using quality-adjusted life-year (QALY) as the effectiveness outcome. Secondary analyses were conducted to explore (1) costs from a societal perspective (i.e. productivity loss in addition to NHS and PSS costs), and (2) using the QIDS-SR as the effectiveness outcome. Mean difference in cost between arms were estimated from generalised linear models with gamma family and log link with participants' baseline costs and randomisation stratification variables as covariates. Results of the cost effectiveness analysis were reported as incremental cost-effectiveness ratios and incremental net benefit. Sensitivity analyses examined cost-effectiveness when (1) the generic unit cost of the trial drugs were used instead of the cheapest (as in the primary analysis), (2) covariate dependent MAR was assumed for missing data, and (3) the analysis was modified to adjust for the impact of the COVID-19 pandemic.

The sample size was revised to 214 in April 2020 due to challenges with recruitment. With an expected 10% loss at follow-up, a log-rank test for the time to trial treatment discontinuation (50% lithium, 70% quetiapine remaining on treatment) would have 80% power. For the QIDS-SR outcome, simulation and the non-central chi-squared method provided a value of 96.5% power to detect an effect size of 0.38 (a minimum clinically significant difference between treatments), with 40% occasion-wise nonresponse assumed.

Results

Two hundred and twelve participants were randomised, 107 to quetiapine and 105 to lithium. Of those randomised to quetiapine, 95 were prescribed and initiated the medication. Of those allocated to lithium, 86 were prescribed and 84 initiated the medication. Of those who initiated the medication, 38.9% of participants randomised to quetiapine and 50.0% of those randomised to lithium discontinued before 12 months. The main reasons for discontinuation in both arms were side effects and inadequate clinical response.

For the time to discontinuation outcome, 1% of participants in the lithium arm were missing data and none in the quetiapine arm. For the QIDS-SR outcome, 19.6% of participants in the quetiapine arm and 16.2% in the lithium arm

were missing data at week 8, 25.2% in the quetiapine arm and 39.0% in the lithium arm were missing data at week 26, and 28.0% in the quetiapine arm and 35.2% in the lithium arm were missing data at week 52. In the ITT analysis, the area under the quetiapine versus lithium QIDS-SR difference curve (AUC) from the fully adjusted model was −68.36, with a confidence interval (CI) of −129.95 to −6.76, excluding the null value of no difference, indicating lower levels of depression in the quetiapine arm compared to the lithium arm over the 52-week study period ($p = 0.0296$). Median time to discontinuation in the quetiapine arm was 365.0 days (25th–75th percentile 57.0–365.0), and 212.0 days (21.0–365.0) in the lithium arm. Participants in the quetiapine arm had 0.72 times the hazard of discontinuing (95% CI: 0.47 to 1.09, i.e. hazard 28% less in quetiapine arm) compared to those in the lithium arm, but the null value of one/the same hazard in each group, could not be excluded. Primary outcome PP analyses gave similar results.

Regarding secondary outcomes, participants in the quetiapine arm scored 2.98 points lower (95% CI: −5.87 to −0.09, $p = 0.0435$) on the MADRS compared to those in the lithium arm at 52 weeks. Similarly, participants in the quetiapine arm scored 3.64 points lower (95% CI: −6.28 to −0.99) on the WSAS than those in the quetiapine arm at 52 weeks ($p = 0.0071$). There was no difference between arms at 8 weeks. There were no differences in weight, blood pressure, PRISE scores or MARS-5 scores between arms at either time point. At week 8, participants in the quetiapine arm had 1.95 times the odds (95% CI: 0.50 to 7.68) of responding compared to those in the lithium arm. This difference was larger at 52 weeks, with participants in the quetiapine arm having 3.67 times the odds of responding (0.94–14.25, $p = 0.0607$). There was little evidence of a difference between arms in remission or global improvement. Participants in the quetiapine arm had 2.22 times the odds of reaching remission at 8 weeks (0.41–11.95), and 1.38 times the odds at 52 weeks (0.29–6.60) compared to the lithium arm. Participants in the quetiapine arm had 1.23 times the odds (0.35–4.39) and 1.12 times the odds (0.32–3.92) of global improvement as compared to those in the lithium arm. Time to initiation of the trial medication and time to initiation of a new treatment for depression did not significantly differ between the two arms. There were 32 SAEs from 18 participants recorded during the trial, 15 from 7 participants randomised to quetiapine and 17 from 11 participants randomised to lithium. The majority were not related or unlikely related to the trial medication, although one event was possibly related to lithium treatment.

Sensitivity analyses suggested primary outcome effects were similar to the ITT analysis when re-estimated in (1) participants who received a therapeutic treatment trial, (2) participants who initiated the trial medication, (3) participants who were randomised or attended study visits before the COVID-19 pandemic and (4) participants who self-reported as being treatment adherent. In exploring missing data assumptions at 52 weeks only, a 0.2 points or greater worsening on the QIDS-SR in the quetiapine arm only would have been needed to render the difference between arms non-significant; however, assuming this worsening only in the quetiapine arm seems a strong assumption.

Regarding health-related quality of life, there were no differences between arms at baseline. However, at week 8 and week 26, participants in the quetiapine arm had significantly better quality of life than those in the lithium arm. This difference became non-significant at week 52. Mean QALY gain between baseline and 52-week follow-up was 0.540 for the quetiapine arm and 0.468 for the lithium arm. The adjusted difference was 0.074 in favour of quetiapine with a 99.5% chance that quetiapine led to improved QALY. There were no significant differences between arms in total NHS or total societal costs are baseline. Over the 52-week follow-up period, the quetiapine arm had a lower healthcare cost (−£472.32, 95% CI: −£1111.12 to £166.47) and a higher societal cost (162.90, 95% CI: −£1224.13 to 1549.94) compared to the lithium arm, with probabilities of the quetiapine arm being cost saving of 0.94 and 0.45, respectively.

In the NHS and PSS cost-effectiveness analysis, quetiapine was associated with a lower cost and a higher QALY gain, and therefore dominated lithium. At NICE's £20,000 willingness to pay (WTP) threshold per additional unit of QALY, the probability that quetiapine was more cost effective was 0.99. When adopting a societal perspective, quetiapine was associated with a higher cost and a higher QALY, with a probability of quetiapine being more cost effective of 0.91. Analyses also indicated that quetiapine was more cost-effective when using the QIDS-SR as the effectiveness outcome. Quetiapine appeared to be less cost-effective in sensitivity analyses compared to the base case scenario; however, in all sensitivity analyses, quetiapine remained the more cost-effective option, according to the NICE £20,000 WTP threshold for one additional unit of QALY.

Conclusions

Clinical guidelines for the treatment of depression currently recommend lithium or second-generation antipsychotics as first-line augmentation options for those who have not responded to antidepressants alone. However, evidence for these options mainly derives from studies in which lithium was added to tricyclic antidepressants (TCAs) and antipsychotics to selective serotonin reuptake inhibitors/serotonin–norepinephrine reuptake inhibitors. Further, very few studies have directly compared these options head-to-head or over the longer term. This trial aimed to provide evidence for clinicians and patients when choosing between augmentation options for TRD. Overall, our results suggested that quetiapine was superior to lithium augmentation therapy in reducing symptoms of depression and cost-effectiveness. Patients randomised to quetiapine showed a greater reduction in QIDS-SR scores over 12 months compared to those randomised to lithium. This effect was also reflected in some of the secondary outcome measures: compared to the lithium arm, those randomised to quetiapine showed significantly lower MADRS and WSAS scores at week 52, but not week 8, and significantly better EQ-5D scores at week 8, but not week 52. Although those who were randomised to quetiapine showed a longer time to discontinuation than those randomised to lithium, this difference was not statistically significant. Similarly, the direction of effects for several of the other secondary outcomes also favoured quetiapine (i.e. MADRS response and remission, CGI-I), but were not statistically significant.

A limitation of the study was the substantial proportions of missing data for some of the secondary outcome measures, limiting our confidence in these results. A significant strength was the long-term follow-up period of 52 weeks with longitudinal weekly symptom measures, since patients with TRD often show a fluctuating response not captured by less frequent cross-sectional measures. Our results extend previous findings from trials with short-term follow-up periods, suggesting moderate clinically relevant benefit of quetiapine over lithium on depression levels in the longer-term, although there was not strong evidence for a difference in discontinuation. Relatedly, the open-label, pragmatic design of the trial, whereby prescribing was continued by participants' primary or secondary care teams, gives insight into the effectiveness of recommended augmentation therapies for TRD in clinical practice. Additionally, future research should explore predictors of treatment response to establish whether there are additional factors which may inform treatment choice.

Trial registration

This trial is registered as ISRCTN16387615.

Funding

This award was funded by the National Institute for Health and Care Research (NIHR) Health Technology Assessment programme (NIHR award ref: 14/222/02) and is published in full in *Health Technology Assessment*; Vol. 29, No. 12. See the NIHR Funding and Awards website for further award information.

Chapter 1 Introduction

Background and rationale

Major depressive disorder (MDD) is a highly prevalent and disabling illness and requires effective treatment to reduce symptoms and improve quality of life.[1,2] Clinical guidelines recommend the use of antidepressant medication for the treatment of moderate to severe depression.[3] However, between 20% and 50% of patients do not respond to first-line antidepressant treatments, typically referred to as treatment-resistant depression (TRD).[3–5] TRD is associated with a poorer prognosis, higher mortality and a higher frequency of physical comorbidities than non-TRD.[6–8] Further, TRD is associated with higher healthcare utilisation costs than non-TRD.[9] For example, those with TRD have higher rates of inpatient stays and emergency care visits,[7] and a recent review reported greater direct and indirect economic costs for TRD than non-TRD.[10] Adequate treatment of TRD can improve prognosis,[11,12] highlighting the importance of appropriately treating this condition.

Treatment options for patients with TRD include increasing the dose of antidepressant the patient is on, switching antidepressant, or augmenting with a second pharmacological agent.[3] While increasing doses has been associated with increased efficacy for a proportion of patients in some trials,[13,14] meta-analyses have reported no overall benefit to dose escalation for most antidepressants.[15,16] Switching to an alternative antidepressant is recommended in cases where a patient has either made no response or is not tolerating their current medication.[3] However, remission rates to switching to third- or fourth-step antidepressant treatments are in the order of just 10–15%.[4] For patients with TRD where there is a partial response to the antidepressant they are on, and good tolerability, augmentation with another agent is recommended.[3]

The efficacy of combinations of multiple antidepressants has been brought into question by a large randomised controlled trial showing antidepressant combinations to be as effective as antidepressant monotherapy, but with some combinations carrying increased risk of adverse events.[17] Meta-analyses have demonstrated efficacy of lithium and atypical antipsychotics (including quetiapine, aripiprazole, olanzapine and risperidone) as augmentation agents for TRD,[18–20] and clinical guidelines recommend these as treatment options for TRD.[3] However, very few studies have compared these treatment approaches (lithium and atypical antipsychotics) head-to-head. As a result, there is little evidence-based guidance for clinicians when choosing an augmentation option for patients with TRD. The largest study to date compared lithium and quetiapine extended release (XR) over 6 weeks, finding quetiapine to be non-inferior to lithium [< 3 points difference on the Montgomery–Åsberg Depression Rating Scale (MADRS)].[21] Although not statistically significant, quetiapine XR showed a small numerical benefit over lithium in reducing depressive symptoms at day 4, 8 and 22. However, there was no long-term follow-up included in this trial, which is imperative given that TRD is frequently a chronic condition that can fluctuate in severity and/or relapse early after response.[22] Clinical guidelines recommend that those responding to an acute trial of medication remain on effective medication for continuation treatment for at least 9–12 months.[3] Therefore, a 12-month trial duration, with frequent assessment of symptoms, is imperative for assessing efficacy in TRD.

A pragmatic, open-label, multicentre clinical trial to compare the efficacy of lithium versus quetiapine augmentation treatment for TRD was conducted. Lithium was selected as it is the first-choice augmentation treatment for patients with TRD according to the World Federation of Societies of Biological Psychiatry Task Force,[23] and among one of the first-line treatment options recommended by the National Institute of Health and Care Excellence (NICE)[24] and the British Association for Psychopharmacology.[3] Quetiapine was selected as the atypical antipsychotic comparator, as there is good evidence for the efficacy of quetiapine augmentation versus placebo,[18,25] and it is currently the only atypical antipsychotic granted a marketing license for use as an augmentation treatment for TRD (XR formulation). Further, it is the only atypical antipsychotic that has been previously compared head-to-head with lithium under trial conditions, where short-term non-inferiority was established.[21] In the current trial, patients were randomised to receive lithium or quetiapine augmentation therapy in addition to their current antidepressant. Initial prescribing was conducted or overseen by a trial clinician, followed by a fully flexible continuation phase for up to 12 months, as per standard

care. It was hoped that the results of this trial would help shape a modified treatment pathway for TRD, in which one treatment would become a preferential first-line augmentation intervention over the other, or one in which there is known equivalence and factors other than clinical- or cost-effectiveness will determine treatment choice.

Aims and objectives

The aim of this trial was to examine whether it is more clinically and cost-effective to prescribe lithium or quetiapine augmentation therapy for patients with TRD over the course of 12 months. A pragmatic, open-label design was chosen in order to reflect real-world clinical practice as far as possible. Primary endpoints were: (1) time to all-cause treatment discontinuation, that is, the time at which patients stopped taking the trial medication, and (2) longitudinal depressive symptom severity (measured via the self-rated Quick Inventory of Depressive Symptomatology, QIDS-SR[26]), monitored weekly over 52 weeks. Because patients with TRD move between states of remission and illness,[27] we chose to assess depressive symptoms weekly in order to obtain a longitudinal outcome, as a simple cross-sectional outcome measure would likely miss important fluctuations in symptoms. A weekly measure would need to be self-reported for practical reasons, and the QIDS-SR shows very good psychometric properties,[26,28] and has already been used with True Colours in other trials.[29] It was expected that a significant proportion of patients would not benefit from acute phase treatment, and would discontinue the study medication well before the 12-month follow-up. To capture this important aspect of effectiveness or tolerability, time to all cause discontinuation was chosen as the co-primary outcome. It was hypothesised that quetiapine would be superior to lithium in terms of time to trial treatment discontinuation and average depressive symptom severity. Secondary objectives included a comparison of economic costs, clinician-rated depressive severity, response and remission rates, health-related quality of life, work and social functioning, adherence to the trial medication, weight, blood pressure, clinician-rated global improvement, side effects, time to initiation of the trial medication, time to initiation of any new intervention for depression, and serious adverse events (SAEs) between the two treatment arms.

Chapter 2 Trial design and methods

Design

The Lithium versus Quetiapine for Depression (LQD) trial was a phase 4, 12-month parallel group, multicentre, patient randomised, pragmatic, open-label, superiority trial, comparing the clinical and cost-effectiveness of lithium versus quetiapine augmentation treatment to antidepressant medication in TRD. The study was designed to reflect clinical practice as far as possible, within the constraints of a clinical trial. Thus, patients were recruited at the point at which a clinician was, or should be, considering prescribing an augmentation treatment, in line with clinical guidelines. Two parallel arms were randomised 1 : 1 at baseline to the decision to prescribe either lithium or quetiapine. Decisions as to whether to continue treatment continued as per standard care thereafter.

The protocol for this trial has been published previously.[30] Amendments to the protocol throughout the course of the study are outlined in *Table 26*, *Appendix 2*. The trial was registered with the International Standard Randomised Controlled Trial Number registry (reference number: 16387615).

Participants

Recruitment procedures

Participants were recruited from six sites across England: South London and Maudsley National Health Service (NHS) Foundation Trust (SLaM); Cumbria, Northumberland, Tyne and Wear NHS Foundation Trust, Oxford Health NHS Foundation Trust (OHT), Tees, Esk and Wear Valleys NHS Foundation Trust (TEWV), Sussex Partnership NHS Foundation Trust (SPFT), and Avon and Wiltshire Mental Health Partnership NHS Trust (AWP). Potential participants were identified primarily through secondary or tertiary care clinics or via consent for contact initiatives within these trusts, as well as community and online advertisements and Participant Identification Centres including primary care services.

Any participants entering the trial from outside the NHS sites were allocated a trial clinician for the initial prescription. Thereafter, the trial clinician could continue the patient's care themselves or transfer their care to a non-trial secondary or primary care clinician as appropriate. Trial clinicians only prescribed the medication the patient has been randomised to if there were no contraindications on the pre-prescribing safety checks and they believed that it was clinically appropriate. However, all participants, regardless of medication status were followed up over 12 months, unless they withdrew from the study completely.

Inclusion and exclusion criteria

Participants were recruited if they met the following inclusion criteria:

1. Under the care of a general practitioner (GP) and/or adult mental health service.
2. Current episode of depression meeting *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition (DSM-V) criteria for MDD (single or recurrent episode), assessed via the Mini International Neuropsychiatric Interview, Version 7 (MINI v7.0).[31]
3. Hamilton Depression Rating Scale-17 item (HAMD-17)[32] score ≥ 14 at screening.
4. Any gender and aged ≥ 18 years.
5. Meet criteria for TRD, defined as failing to adequately respond to at least two antidepressant therapies in the current episode, prescribed for at least 6 weeks at minimum therapeutic dose, as determined by the Maudsley Prescribing Guidelines[33] and/or *British National Formulary* (BNF).[34] Relapse while on an antidepressant also counts as a failed treatment trial.
6. Current antidepressant treatment has remained unchanged and at, or above a therapeutic dose for ≥ 6 weeks.
7. Provision of written informed consent.

Participants who met any of the following criteria were excluded:

1. Diagnosis of bipolar disorder (type I or II, as per DSM-V criteria) on the MINI v7.0.
2. Diagnosis of current psychosis on the MINI v7.0.
3. Adequate use of lithium or quetiapine during the current episode. An adequate dose of lithium is defined as the patient taking lithium for at least 4 weeks at an adequate dose (leading to a documented plasma concentration of > 0.4 mmol/l) and for quetiapine, prescription in the range of 150–300 mg/day for 4 weeks or longer. Or, if the patient has taken an inadequate dose of lithium or quetiapine in the current episode, the patient and clinician are not willing to re-prescribe/take the medication.
4. Ongoing use of another atypical antipsychotic.
5. Known contraindication to use of either lithium or quetiapine: known hypersensitivity of lithium or quetiapine or any of their excipients; severe renal insufficiency or impairment; untreated hypothyroidism; severe cardiac disease or insufficiency; low sodium levels for example, dehydrated patients or those on low sodium diets; Addison's disease; Brugada syndrome or family history of Brugada syndrome; the rare hereditary inborn errors of metabolism galactosaemia, the Lapp lactase deficiency or glucose-galactose malabsorption; concomitant administration of cytochrome P450 3A4 inhibitors; or congenital QT prolongation.
6. Currently participation in another Clinical Trial of an Investigational Medicinal Product (CTIMP).
7. Insufficient degree of comprehension or attention to be able to engage in trial procedures.
8. Pregnancy, actively trying for pregnancy, or currently breastfeeding.

Sample size

The sample size was changed from 276 to 214 under a variation to contract in April 2020, due to challenges with recruitment. The original sample size description is presented below, with an addendum outlining the updated calculation. The sample size was based on the following assumptions:

Drop out and missing data

We estimated a 20% drop out from treatment by the 8-week follow-up visit as per Bauer and colleagues[21] and 50% treatment discontinuation by 52 weeks. We allowed for discontinuation status at follow-up (co-primary outcome) to be unknown for 10% of the sample. Although previous trials suggest high rates of data return (80%) when using True Colours,[29] we allowed for 40% missingness for each of the post-randomisation QIDS-SR scores (co-primary outcome).

Effect size

The minimum clinically significant difference for outcomes in depression treatment is widely taken to be three points on the HAMD-17.[3] This corresponds to an effect size of 0.38 between treatments. We wished to see a difference of this effect size in the QIDS-SR score sustained over the period of the trial, so estimated the effect as an area under the curve (AUC; a simple average of post-randomisation measures if equally spaced). Over the year of follow-up, we expected a rate of discontinuation of assigned treatment of 50% and have powered for an improvement to 30% (i.e. a 20% reduction).

Power

All power calculations are for two-tailed test and $\alpha = 0.05$. With a sample size of 276 and 10% loss we expected 248.4 at follow-up. Using a log-rank test for the time to discontinuation with 50% (lithium) and 70% (quetiapine) as the proportions remaining on assigned treatment, Stata stpower logrank gives a power of 90%. This determined the lower limit of the sample size. Applying this sample size to the self-report True Colours data we also needed to account for the likely haphazard nature of the missingness at each assessment time. We used simulation and the non-central chi-squared method for calculating power. Over 1000 samples a simple random intercept model covarying for baseline, a time dummy variable and with a single average combined treatment effect gave 99.7% power for an effect size of 0.38 (1 df, 3.84 and 20.86 for the noncentrality parameter), with 40% occasion-wise nonresponse assumed.

Updated sample size recalculation

With a sample size of 214 and 10% loss we expected 192 participants at follow-up. Using the logrank test and parameters described above for the time to discontinuation outcome gave 80% power. Applying this sample size to the self-report True Colours data and using the same methods described above gave 96.5% power for an effect size of 0.38 (1 df, 3.84 and 14.26 for the noncentrality parameter), with 40% occasion-wise nonresponse assumed.

Randomisation and blinding

Participants were randomised 1 : 1 to the decision to prescribe lithium or quetiapine, stratified by geographical region (London, Oxfordshire, North East England, Bristol and Brighton), depression severity (HAMD-17 baseline severity score: moderate 14–18, severe 19–22 or very severe ≥ 23), and TRD severity (failure of two vs. three or more antidepressant treatments in the current episode), using block randomisation with randomly varying block size. Randomisation was conducted by a researcher using a web-based service hosted by King's Clinical Trials Unit (KCTU) on the same day that consent was taken and eligibility confirmed.

This was an unblinded trial, whereby researchers, clinicians and patients were aware of treatment allocation. However, clinician-rated outcome measures [the MADRS[35] and the Clinical Global Impression (CGI) scale[36]] were rated by trained assessors blind to treatment allocation. Participants were reminded not to reveal treatment allocation at the beginning of these assessments. Statisticians were also blinded (see *Clinical effectiveness analyses* for details).

Interventions

Trial clinicians received the following recommendations for titration and dosing, in line with current best practice guidelines:[33,34]

Lithium arm: lithium carbonate, added on to the current antidepressant. Lithium citrate is an acceptable alternative for those who cannot take tablets as it is available in liquid form. Standard BNF dose titration is detailed below for reference: the dose should be adjusted to achieve a serum lithium concentration of 0.4–1.0 mmol/l 12 hours after a dose on days 4–7 of treatment, then every week until dosage has remained constant for 4 weeks aiming for an optimal therapeutic plasma level of 0.6–1.2 mmol/l.[21,33] Blood monitoring should be performed as per Maudsley Prescribing Guidelines.[33]

Quetiapine arm: quetiapine fumarate (XR or immediate release, IR) added on to the current antidepressant, taken once daily before bedtime. Dose was titrated upwards using a standard BNF dose titration protocol of 50 mg on days 1 and 2 and 150 mg on day 3, aiming for a dose of 300 mg/day by week 2 if tolerated. Thereafter, flexible dosing followed in the range 150–300 mg/day according to tolerance.[21] In elderly patients (> 65 years old), the dose titration protocol was modified according to the summary of product characteristics and best practice as follows: 50 mg/day on days 1–3, increasing to 100 mg/day on day 4, 150 mg/day on day 8 and 300 mg/day not before day 22 of treatment if required.

Dosing regimens could be altered in cases where concomitant administration of drugs could interact with quetiapine or lithium. Trial clinicians were also advised to keep patients' existing antidepressant medication treatment(s) at a stable dose within the therapeutic range, as defined in the Maudsley Prescribing Guidelines[33] and BNF.[34]

The initial prescription of trial medication and any essential pre-prescription safety checks were overseen by a trial clinician. As patients were randomised to *the decision* to prescribe either lithium or quetiapine, the clinician could decide whether to prescribe the medication to which the patient had been randomised based on pre-prescribing safety checks and their clinical judgement. Pre-prescribing safety checks are described in *Baseline and randomisation: Week 0 (Time point 0)*. These tests had to be completed prior to prescription unless they had been completed within a sufficiently recent period according to the clinician's judgement.

All patients were followed up over 12 months, regardless of medication status, unless they withdrew from the study completely. Treatment could be discontinued if a participant no longer wished to take the medication, or in the event of a serious drug reaction. Treatment could be extended beyond the study follow-up period at the discretion of the patient's clinician. All concomitant pharmacological and non-pharmacological interventions were recorded, as well as any changes to the trial medication (e.g. changing from one brand of quetiapine to another).

A minimum therapeutic treatment trial of the medication the participant was randomised to, for purposes of analyses, at any point up to and including the 52-week follow-up visit, was defined as:

- A documented lithium serum level between 0.6 and 1.2 mmol/l[21,33] and the participant reporting to have taken lithium for at least 4 weeks by the 52-week study visit (determined via True Colours, treatment initiation and discontinuation forms, and medical records).
- Quetiapine prescribed at ≥ 150 mg/day for at least 4 weeks by the 52-week study visit (determined via True Colours, treatment initiation and discontinuation forms, and medical records).

Primary outcomes

There were two primary outcome measures:

- Time to all-cause trial treatment discontinuation, defined as the difference in days between date of initial prescribing of the drug to which the participant was randomised and the date on which the participant reported the last dose of the same drug was taken.
- Overall burden of depressive symptom severity measured weekly using the QIDS-SR. This was measured using the AUC using all available weekly QIDS-SR measurements.

Secondary outcomes

Secondary outcome measures are as follows:

- Clinician-rated depression severity (MADRS total score), weeks 8 and 52.
- Response rates (proportion with ≥ 50% reduction in baseline MADRS total score) at weeks 8 and 52.
- Remission rates (proportion with MADRS total score ≤ 10) at weeks 8 and 52.
- Health-related quality of life [EuroQol-5 Dimensions (EQ-5D)[37] summary index score] at weeks 8 and 52. This outcome is reported as part of the economic analysis (see *Health economic analyses*).
- Work and social functioning [Work and Social Adjustment Scale (WSAS)[38] total score] measured at weeks 8 and 52.
- Treatment adherence [continuous 5-item Medication Adherence Report Scale (MARS-5)[39] total scores and exploratory cut-offs categorising participants as either adherent or non-adherent] at weeks 8 and 52.
- Weight (kg) from baseline to weeks 8 and 52.
- Diastolic blood pressure (mmHg) at weeks 8 and 52.
- Systolic blood pressure (mmHg) at weeks 8 and 52.
- Global improvement (proportion of participants with a CGI-Improvement score of 'much' or 'very much' improved) at weeks 8 and 52.
- Side effects [Patient Rated Inventory of Side Effects (PRISE)[40] total score] at weeks 8 and 52.
- Time to initiation of trial medication (defined as the date on which the participant reports taking the first dose of the treatment measured from randomisation), as well as the proportion initiating treatment in each arm.
- Time to initiation of any new intervention for depression (defined as the first date on which a participant records taking any new pharmacological or non-pharmacological treatment specifically prescribed for depression, measured from randomisation), as well as the proportion initiating a new intervention in each arm.
- Number of events and number of people with events in each arm summarised as SAEs, serious adverse reactions (SARs), and suspected unexpected serious adverse reactions (SUSARs).

Tertiary and ancillary outcomes

A full list of tertiary and ancillary outcomes is provided in *Appendix 3*. These will not be reported in the current report. One exception is the economic analysis outcome measure, described below.

- Economic cost over 52 weeks [Client Service Receipt Inventory (CSRI)[41] modified for TRD, measured at baseline, 8-, 26- and 52-week visits]. This outcome was used to compare costs between the two arms over 52 weeks.

Trial procedures

See *Figure 1* for a flow chart of trial procedures, and *Table 1* for details of all study measures completed at each visit. Visits took place at screening and baseline, week 8, week 26 and week 52. In addition, weekly monitoring was administered via True Colours. All trial visits took place at an NHS trust or university site, depending on local arrangements. Every effort was made to ensure participants attended a clinical or research facility for their visit; however, in some cases, follow-up visits took place over the phone or via videoconferencing software. In such instances, physical examinations and the Digit Symbol Coding Test (DSCT)[42] could not be completed.

Screening, baseline and randomisation visit

Informed consent, screening, baseline and randomisation could occur on the same or different visits. However, the baseline visit and randomisation had to occur on the same day, defined as Time point 0. This baseline visit could occur anytime within 12 weeks post-informed consent. This was to allow for medical records to be requested or access, in cases where the participant could not confirm eligibility.

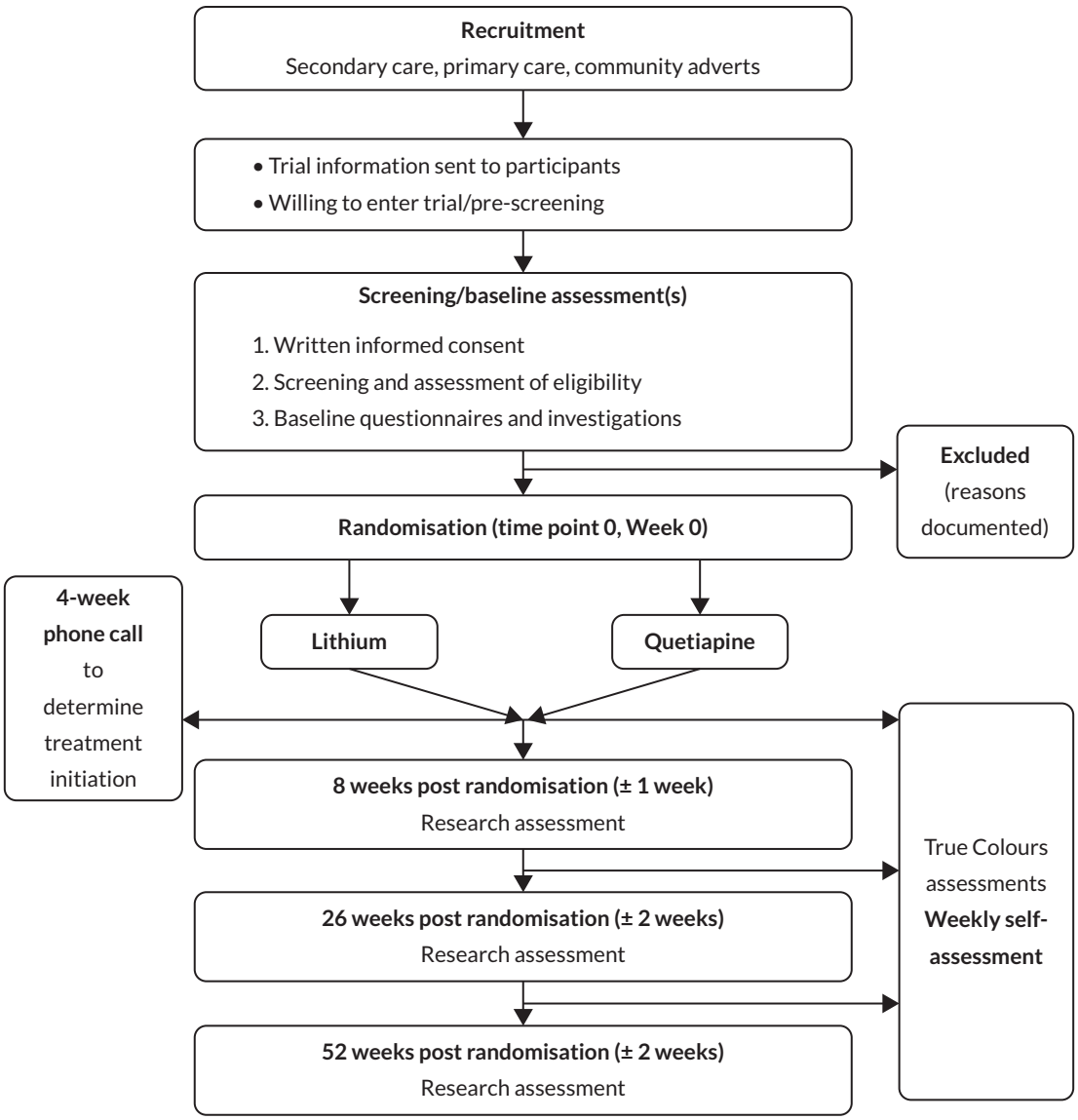

FIGURE 1 Trial flow chart.

TABLE 1 Trial procedures by visit

Time point	Screening	Baseline (week 0)	Follow-up				
			4 (± 1) week call	Weeks 0–52, weekly assessments	Week 8 (± 1 weeks)	Week 26 (± 2 weeks)	Week 52 (± 2 weeks)
Written informed consent	√						
Eligibility assessment	√						
MINI 7.0 to confirm MDD and other Axis 1 disorders	√						
Assessment of depression severity (HAMD-17)	√						
Assessment of medication history in current depressive episode (MTI)	√						
Sociodemographic, psychiatric and medical history (including MSM and medical records)	√						
Assessment of concomitant medication and non-pharmacological therapies	√	√			√	√	√
Randomisation		√					
MÅDRS		√			√	√	√
CGI[a]		√			√	√	√
IDS-C[b]		√			√		
HCL-16[b]		√					
Trial medication initiation check			√				
FIBSER[b]					√	√	√
PRISE					√	√	√
EQ-5D		√			√	√	√
THINC-it[b,c]		√			√	√	√
DSCT[b]		√			√	√	√
QIDS-SR[d]		√		√	√	√	√
WSAS[d]		√		√	√	√	√
Trial medication status[d]				√	√	√	√
ASRM[b]		√			√	√	√
Maudsley-VAS-current[b]		√			√	√	√
Maudsley-VAS-change[b]					√	√	√
GAD-7[b]		√			√	√	√
SAPAS[b]		√			√	√	√
CSRI modified for TRD and employment status[b]		√			√	√	√
TSQM[b]					√	√	√
MARS-5[e]		√			√	√	√
CTQ		√			Or √	Or √	Or √

TABLE 1 Trial procedures by visit (*continued*)

Time point	Screening	Baseline (week 0)	Follow-up: 4 (± 1) week call	Follow-up: Weeks 0–52, weekly assessments	Follow-up: Week 8 (± 1 weeks)	Follow-up: Week 26 (± 2 weeks)	Follow-up: Week 52 (± 2 weeks)
MSM-MTR							√
Physical examination (weight, height, blood pressure, pulse rate, waist circumference)		√			√	√	√
Research blood tests (FBC, U&Es, LFTs, TFT, glucose, lipids, calcium)[b,c]		√			√		√
Lithium and quetiapine serum levels[b,c]					√		√
BioResource sample collection[b,c]		√			√		√
Qualitative interview: patient experience of True Colours system[b,c]					√	Or √	Or √
Qualitative interview: patient views and experience of lithium and quetiapine[b,c]							√ or post-trial completion
Adverse event monitoring			√				
Standard care monitoring (including safety checks and blood monitoring), data collected from medical records if applicable			√				

ASRM, Altman Self-Rating Mania scale; CTQ, Childhood Trauma Questionnaire; EQ-5D, EuroQol-5D health index; FBC, full blood count; FIBSER, Frequency, Intensity, and Burden of Side Effects scale; GAD-7, Generalised Anxiety Disorder Questionnaire; HAMD-17, Hamilton Depression Rating Scale – 17 item; HCL-16, 16 item Hypomanic Checklist; IDS-C, Inventory of Depressive Symptomatology-Clinician Rated; LFTs, liver function tests; Maudsley-VAS-change, Maudsley Visual Analogue Scale – change in depression severity; Maudsley-VAS-current, Maudsley Visual Analogue Scale – current depression severity; MSM-MTR, Maudsley Staging Model of Treatment Resistance, Modified Multi-Therapy Resistance Version; QIDS-SR, Quick Inventory of Depressive Symptomatology-Self Rated; SAPAS, Standard Assessment of Personality – Abbreviated Scale; TFT, thyroid function tests; THINC-it tool for Cognitive Dysfunction in Major Depressive Disorder; TSQM, Treatment Satisfaction Questionnaire for Medication; U&Es, urea, electrolytes and creatinine.
a Only CGI severity taken at baseline, full CGI at follow-ups.
b Tertiary or ancillary outcome measure.
c Optional and/or collected in a subset of participants.
d If True Colours assessments had been completed the same week as a follow-up visit, these were not repeated at the visit.
e Baseline to antidepressant, follow-up to trial medication.

Informed consent process

The consent process was carried out by a trial researcher trained in good clinical practice. Participants were given the opportunity to ask questions and read the participant information sheet to ensure that they understood the study. Participants then provided written consent before any trial procedures as specified in the protocol were performed. The original consent form was kept in the site file, and a copy given to the participant. A further copy was added to the participant's medical notes.

A separate participant information sheet and consent form was provided to participants who wished to be involved with our collaboration with the National Insitute for Health and Care Research (NIHR) BioResource for Mental Health, a national initiative to build up a central BioBank about people's health (REC reference number: 15/SC/0388). These participants provide blood, saliva and/or hair samples. Participation in the BioResource was optional.

Screening

After providing informed consent, the screening procedures commenced to determine eligibility, as described below and summarised in *Table 1*. If at any point the participant was found to be ineligible, screening procedures were stopped and reason for exclusion recorded.

- Age (above 18 years old), confirmed via medical record check.
- Depression severity: HAMD-17 score ≥ 14 and currently meets DSM-V criteria for MDD as determined via the MINI v7.0.
- Treatment resistance: meets criteria of failing to respond adequately to at least two antidepressants given at a therapeutic dose for at least 6 weeks in the current episode of depression, determined via the Maudsley Treatment Inventory (MTI) and the Maudsley Staging Method (MSM).[43] The MTI and MSM are completed by asking the participant about the dose and duration of each treatment they have tried for the current episode of depression, as well as their tolerability, response, and adherence to treatment. Response to the current antidepressant can also be operationally defined using the CGI. Switching treatments due to adverse events counts as a course of treatment if the patient had an adequate trial before switching and remained symptomatic. Information from medical records was used to supplement patient-reported information.
- Medical contraindication to lithium or quetiapine: patient does not meet any of the following, determined via participant self-report and medical record check: known hypersensitivity of lithium or quetiapine or any of their excipients; severe renal insufficiency/impairment; untreated hypothyroidism; severe cardiac disease/insufficiency; low sodium levels for example dehydrated patients or those on low sodium diets; Addison's disease; Brugada syndrome or family history of Brugada syndrome; the rare hereditary inborn errors of metabolism galactosaemia, the Lapp lactase deficiency or glucose-galactose malabsorption; concomitant administration of cytochrome P450 3A4 inhibitors; or congenital QT prolongation.
- Augmentation status: the patient has not taken an adequate dose of lithium or quetiapine in the current episode, and/or no ongoing use of another atypical antipsychotic (determined via MTI and medical records).
- Psychosis: patient does not currently meet DSM-V criteria for current psychotic disorder and/or mood disorder with psychotic features (MINI v7.0).
- Bipolar disorder: patient does not meet DSM-V criteria for bipolar I or II (MINI v7.0).
- Not pregnant, breastfeeding or actively trying to get pregnant (females only), based on verbal report of the participant.
- Not currently participating in another CTIMP, determined via verbal report of the participant and medical records.
- Sufficient degree of comprehension or attention to be able to engage in trial procedures, determined by the trial researcher.
- Under the care of a GP and/or adult mental health services (determined via self-report of the participant).
- Current antidepressant status: currently taking an antidepressant, at a therapeutic dose and for a duration of at least 6 weeks (determined via MTI and medical records).
- Patient does not wish to withdraw consent.

The following sociodemographic information was also collected during screening: date of birth, gender, relationship status, ethnicity, country of birth, first language, education level (years of education and highest level of education), employment (employment status, occupational type, number hours paid work per week, number of hours missed work due to ill health), number of units of alcohol and number of cigarettes smoked per week.

Baseline and randomisation: week 0 (Time point 0)

Eligible participants completed the baseline assessments specified in *Table 1*. Baseline assessments included several self-report questionnaires, researcher-rated assessments, current concomitant pharmacological and non-pharmacological therapy checks, physical examinations, cognitive tests and optional blood tests (research blood tests and BioResource samples). If the baseline visit was more than 7 days after the screening visit, the screening assessments (except demographic information) were conducted again to ensure the participant still met eligibility criteria.

Randomisation was carried out on the same day as the baseline assessment. The researcher informed the trial clinician and the participant which treatment they had been allocated to. The trial clinician then carried out following safety tests before issuing the study prescription, unless they had already been conducted sufficiently recently in the opinion of the clinician. Where study baseline tests were available, these could be used as pre-prescribing safety checks.

- Pre-lithium blood tests (lithium arm only): renal function, thyroid function, full blood count, serum calcium.
- Pre-quetiapine blood tests (quetiapine arm only): full blood count, blood lipids, plasma glucose, liver function tests, creatine phosphokinase and thyroid function tests (recommended but not essential).
- Electrocardiogram (ECG), if clinically indicated, that is in patients with risk factors for, or existing cardiovascular disease (both treatment arms).
- Weight [including body mass index (BMI) and waist circumference if possible] and blood pressure (recommended but not essential).

True Colours

Participants were asked to fill out questionnaires via True Colours (https://oxfordhealth.truecolours.nhs.uk/) each week. The participant received weekly prompts via e-mail or text at a time of their choice. The assessments completed using this system were: the QIDS-SR, WSAS and trial medication status ['Which study medication are you currently taking?' and 'What is your currently prescribed total daily dose (mg/day)?']. If a patient reported that they had not taken the trial medication for 2 consecutive weeks (after previously taking the medication) they were contacted to see whether they had discontinued the medication. If the patient reported discontinuing the medication, the date and reason for discontinuation were recorded. This self-report data was supplemented with information from the patient's medical records and the prescribing clinician where possible.

Week 4 phone call

Participants were called 4 weeks after the baseline/randomisation visit to determine the date at which they initiated the medication. If they had not yet started the medication, the reason for this was recorded, and this question was asked again at follow-up visits.

Follow-up visits: week 8, 26 and 52

Procedures carried out at follow-up visits are detailed in *Table 1*. Follow-up visits at week 8 took place within ± 1 week of the expected visit date. Follow-up visits at 26 and 52 weeks took place within ± 2 weeks of the expected visit date.

At each visit, participants completed several self-report questionnaires, researcher-rated assessments, blind-rated assessments, current concomitant pharmacological and non-pharmacological therapy checks, physical examinations, cognitive tests and optional blood tests (research bloods and BioResource samples, weeks 8 and 52 only). Additionally, two optional qualitative interviews were conducted in a subset of participants at follow-up visits, the results of which have already been published.[44,45]

Data input and validation

Trial data were entered into a secure web-based electronic case report form, InferMed Macro, hosted by KCTU. Data extracts from Macro were requested every 6 months in order to check the data were complete. The following source data validation checks were carried out by King's Clinical Trials Office: 5% of participant eligibility data (selected at random at each site), 50% of participants (selected at random) primary outcome data, and 100% of SAEs.

Patient and public involvement

Patients were involved in the design of the trial during the funding application stage. Five patients with a history of treatment with augmentation therapies (including the trial medications) provided their feedback on the protocol during the preparation of the full outline application. Patient-facing documents, such as the participant information sheet, were reviewed by the NIHR Feasibility and Acceptability Support Team for Researchers, a group of people with experience of mental health problems and their carers who are trained to advise on study documentation and proposals. During participant recruitment, a patient and public involvement (PPI) meeting was held to gain feedback on recruitment methods and how these could be improved. Four people with TRD and study researchers discussed recruitment documents (e.g. letters of invitation) and mode and frequency of communication with potential participants. In addition, two PPI members were on the Trial Steering Committee (TSC), which monitored progress throughout the duration of the trial.

Statistical analyses

Clinical effectiveness analyses

Statistical aspects and analysis followed KCTU statistical Standard Operating Procedures. A statistical analysis plan (SAP) (see *Report Supplementary Material 1*) was developed by the statisticians (KG, RH and ZZ) with input from the Data Monitoring and Ethics Committee (DMEC), Trial Management Group, chief investigator (AC) and trial managers (JKG, LM, RT, HT). Version 1.0 of the SAP was approved and signed by the chief investigator and TSC chair prior to database lock. The trial statistician (ZZ) remained blinded until 6 months before the last participant's final follow-up visit was due. The senior statistician (KG) remained blinded until the primary and secondary outcomes had been analysed, and included in an initial draft of the statistical report for checking. This initial report was prepared with the arms labelled as 'A' and 'B', partially unblinding the senior statistician. After this version was checked, the senior statistician was fully unblinded to allow checking of the final report.

Analyses were carried out by the trial statisticians (ZZ and KG). Stata v17[45] (StataCorp LP, College Station, TX, USA) was used for data description and the main inferential analysis. Variables were summarised overall and for each randomised arm using means and standard deviations (SD), medians and 25th and 75th percentiles (which have been referred to in this document as interquartile range or IQR for reasons of brevity), or frequencies and proportions as appropriate. A subset of baseline variables as outlined in the SAP were summarised. Additionally, the number of co-morbidities (0, 1, 2, 3 or more) as recorded using the MINI v7.0[31] was reported by trial arm. Baseline differences between randomised arms were not statistically tested, as properly performed randomisation should ensure any imbalances over all measured and unmeasured baseline characteristics are due to chance.

In the first instance the primary analyses were conducted under intention to treat (ITT) assumptions, that is, individuals' data were analysed in arms as randomised irrespective of treatment received. The primary outcome analyses were repeated in a per-protocol (PP) subset. This subset was planned to exclude individuals who were not prescribed trial medication, and anyone found later to have been ineligible for the trial. None of the randomised individuals were later found to be ineligible, so the subset consisted of those prescribed trial medication. The significance level for the primary and secondary outcomes was 5% (two-sided). Secondary outcome *p*-values were presented unadjusted and then compared to a Hochberg's improved Bonferroni method threshold.[46] As a false discovery rate (FDR) was not pre-specified, we present this comparison for FDRs of 5%, 10% and 20%.

Missing outcome data

For questionnaire scale measures with missing items, where there was guidance in the primary publication for scoring, this was used. Where this was not the case, the total pro rata score was calculated provided that the level of missingness was no more than 20% (i.e. for a scale with 10 items where no more than 2 items are missing the missing items were replaced with the mean of the non-missing items). Missing continuous baseline variables were imputed as the mean value for the entire complete trial sample.[47] Missing categorical baseline variables were imputed as the modal value.

Analysis of primary outcomes

The QIDS-SR primary outcome was analysed using a linear mixed model with outcomes from the 52 weekly True Colours reports and those collected at the main trial research time points as the dependent variables, and treatment allocation and time as explanatory variables, with time modelled as continuous, and a treatment by time interaction term to allow effects to vary over time. The model was conditioned on baseline QIDS-SR score, baseline variables that predicted missing 52-week QIDS-SR score (ethnicity coded as white/other and number of psychiatric comorbidities as recorded using the MINI), and randomisation stratification variables (TRD severity, depression severity and site). To assess which baseline variables predicted missing 52-week QIDS-SR, a binary variable indicating whether the QIDS-SR was missing at 52 weeks was created. Baseline variables were entered individually into a logistic regression model with the missing 52-week QIDS-SR indicator variable as the dependent variable. Variables with a univariate *p*-value of 0.2 or less were entered into manual forward stepwise logistic regression analysis and retained in models based on likelihood ratio tests. Any variables remaining in a final model with a *p*-value of 0.05 or less were added to the final QIDS-SR mixed model to make the missing at random (MAR) assumption more plausible. Baseline variables assessed were: age, number of psychiatric comorbidities, HAMD-17, MADRS score, HCL score, WSAS score, MARS-5 for antidepressant

adherence score, GAD score, weight, waist circumference, diastolic blood pressure, systolic blood pressure, pulse rate, ethnicity, education, employment status and CGI severity of illness. Random intercepts for individuals were included to account for the repeated measures. A likelihood ratio test indicated random slopes for individuals over time were also warranted, and so the model also allowed for random slopes. The differences between the two treatment arms were estimated for each weekly time point, then the area under the difference curve constructed from these estimates was estimated by integrating (using the trapezoid rule) over the 52 weekly differences. Fitted values incorporating fixed and random effects using Stata's 'predict *variablename*, fitted' command were estimated and plotted against observed values over the 52 weekly time points.

The time to trial medication discontinuation primary outcome was analysed using time-to-event/survival analysis methods. The discontinuation date was defined as the recorded discontinuation date, the 52-week follow-up point date, the withdrawal date if the participant formally withdrew, or the last date on which any contact with the participant was recorded where contact with the participant was lost during follow-up. Individuals who were not prescribed the trial medication, or who were prescribed the trial medication but failed to initiate treatment were assigned a short (i.e. essentially immediate) discontinuation time of 0.5 days. Data were summarised using appropriate summary statistics as described at the beginning of the methods section, and also using Kaplan–Meier methods. Cox regression models were used to estimate the hazard ratio for discontinuation comparing the two treatment arms, with treatment allocation and the randomisation stratification variables as independent variables. Bootstrap methods with 1000 replications were used to obtain hazard ratio confidence intervals (CIs). The assumption of proportional hazards was checked informally via a plot of scaled Schoenfeld residuals for the arm covariate against time. If the proportional hazards assumption held, the plot should show the residuals scattering randomly consistently over time and a line with a zero slope. As there was evidence of non-proportional hazards, we also analysed the data using restricted mean survival time (RMST) models, which were not specified in the SAP, but were settled on in consultation with the Chair of the DMEC, who is a statistician. RMST models are one option in this situation as they do not make the proportional hazards assumption, and may also provide a more interpretable quantity as compared to the hazard ratio. These models were used to estimate the difference between the arms in the average days continuing on trial medication over 12 months of follow-up.

Analysis of secondary outcomes

Repeatedly measured continuous secondary outcomes were analysed using linear mixed models with outcomes from the post-randomisation time points (week 8, 26, where recorded, and 52) as the dependent variables and treatment allocation, time as a categorical variable, and time by treatment interaction terms as explanatory variables. Models were conditioned on the baseline measure of the variable and randomisation stratification variables (TRD severity, depression severity and site). In the mixed models, random intercepts for individuals were included to account for the repeated measures. Binary secondary outcomes were modelled using longitudinal logistic mixed models with the binary variable of interest at the weeks recorded as the dependent variable and the treatment allocation, time as a categorical variable, time by treatment interaction terms, and stratification variables (TRD severity, depression severity and site) as explanatory variables. The time to new intervention and time to treatment initiation outcomes were analysed in a similar manner to time to discontinuation. Time to initiation of treatment was defined as the time in days between randomisation and the date when the participant took the first dose of the trial medication. This analysis was done using the subset of participants who were prescribed the randomised trial medication. Time to initiation of a new intervention was defined as the time in days between randomisation and the date when the participant recorded taking any new pharmacological or non-pharmacological treatment specifically prescribed for depression. The number of SAEs, SARs and SUSARs were summarised overall and by arm as the number of events and number of participants who experienced events.

In analysing the weight measures some potential outlier values were found, so analyses were done both retaining and removing outliers. An outlier was defined as a weight measure that was larger or smaller than three times the SD from the mean at the specific time point.

Most comparisons in the report are for quetiapine versus lithium. To aid interpretation of the hazard and odds ratio results, we have shown comparisons both for quetiapine versus lithium and lithium versus quetiapine.

Sensitivity analyses

The between arm effects for the primary outcomes were re-estimated for (1) the subset of participants who were both prescribed and then went on to initiate the trial medication, (2) the subset of participants that were adherent to trial medication defined as MARS-5 self-report of 80% or greater adherence at the time points when they were taking trial medication, and (3) participants having a therapeutic treatment trial, defined as per *Interventions* of this document and *Section 5.2* of the protocol, using a combination of recorded trial data and clinical expertise.

The MARS-5 sensitivity analysis adherence cut-off of 80% differs from that stated in the SAP (75%). When undertaking the sensitivity analysis, it was noticed that the 75% adherence cut-off was mistakenly transferred to the MARS-5 from the MTI, which recommends 75% of doses taken as an adherence cut-off. However, the MTI was only taken at baseline in this trial, and a cut-off of 75% on the MARS-5 is not supported in the literature. A review of the literature indicated that a total score of 23 or greater on the MARS-5 equates with 80% adherence.[48–50] This is the cut-off that has been used for the sensitivity analysis. Further, the timing of MARS-5 measurement used for the analysis has been determined post hoc due to an inadequately detailed description of this aspect in the SAP. For the purposes of this analysis, participants were considered 80% adherent if they scored 23 or greater on the MARS-5 at the week 8 visit, or if they hadn't initiated by week 8/their week 8 MARS-5 measurement was missing, the earliest complete measurement prior to discontinuation was used instead (i.e. either the week 26 or week 52 measurement).

For the QIDS-SR outcome only, we explored departures from the MAR assumption made when using maximum likelihood estimation in mixed models. We used the *rctmiss* command in Stata to apply a pattern mixture model exploring the effects of assuming a systematic worsening only for the single 52-week time point. We explored a range of delta values from 0.1 to 1 point on the scale (one point on the scale approximately equating to 0.2 of a SD in the QIDS-SR scores at baseline and post randomisation in the LQD trial population). The quetiapine versus lithium difference in QIDS-SR was estimated using a linear regression model with the 52-week measure of QIDS-SR as the dependent variable. Similar to the main primary outcome analysis, treatment allocation, baseline QIDS-SR score, ethnicity, number psychiatric comorbidities, and randomisation stratification variables (TRD severity, depression severity and site) were included as explanatory variables. We evaluated the effect of applying this range of delta values in the quetiapine arm only, the lithium arm only, and in both arms. This was not the analysis specified in the SAP; please see *Appendix 6*, Sensitivity Analysis 4 for more information.

Finally, we performed sensitivity analyses to explore the effect of the COVID-19 pandemic. We first note that for the main analysis we did not delete any data collected after the pandemic started, that is focused on 'world including a pandemic' estimands.[51] For the sensitivity analyses, we provide descriptive statistics for some variables dividing the sample into those randomised before and after 1 February 2020, and for others into outcomes gathered before and after 1 March 2020. We also re-analysed the primary outcomes: (1) completely excluding participants randomised after 1 February 2020 and (2) excluding any outcome measures taken after 1 March 2020. For the discontinuation outcome this equated to setting 1 March 2020 as the censoring date and excluding anyone getting their prescription after 1 March 2020. More detail can be found in *Section 3.4* of the SAP (see *Report Supplementary Material 1*).

Health economic analyses

A health economics analysis plan (see *Report Supplementary Material 2*) was developed by the health economists (NY, HJ), with input from the TSC and chief investigator (AC). The primary aim of the economic analysis was to compare the cost-effectiveness of prescribing quetiapine augmentation therapy with lithium augmentation therapy for patients with TRD. The time horizon of the analysis was 52 weeks. The primary analysis was conducted from an NHS and Personal Social Services (PSS) perspective, as recommended by the National Institute for Health and Care Excellence (NICE),[52] using quality-adjusted life-year (QALY) as the effectiveness outcome. Secondary analyses were conducted to explore (1) costs from a societal perspective (i.e. productivity loss in addition to NHS and PSS costs) and (2) using the QIDS-SR as the effectiveness outcome. The cheapest unit costs of the trial drugs were used. All analyses were conducted using Stata v17 (StataCorp. Stata: Release 17, 2021).

Measurement and valuation of resource data

Details of dosage of the trial medication were recorded weekly throughout the trial. Drug treatment costs were calculated using daily dose information, form and type of drug (modified release vs. IR), most likely prescribed strength

and name of medication using data from the NHS Open Data Portal, and the cost of the generic drugs or cheapest alternative as listed in the Drug Tariff and the BNF.[53]

Data on use of all other services included in the study were collected using the CSRI, modified for TRD. The CSRI measured use of NHS and PSS services, including hospital services and primary/community care services, as well as productivity costs (days off work by patients because of their health condition, and days off work by principal and other informal carers). The CSRI asked participants for the number and duration of contacts with these services and professionals. The CSRI was used to collect information about participants' use of services before the treatment period (for a period of on average 3 months), the period from baseline to 8 weeks (first follow-up), 26 weeks (second follow-up) and 52 weeks (third follow-up) post randomisation. At each of the study visits, service use since the previous interview were recorded; in this way the entire period from 3 months prior to baseline to the final follow-up were covered.

For all NHS/PSS health and social care (including voluntary sector) services, nationally applicable unit costs at 2022–3 prices were employed. These unit costs were taken from the annual Personal Social Services Research Unit (PSSRU) compendium,[54] see *Table 25*, *Appendix 1*. NHS reference costs and national tariffs were used to estimate the cost of inpatient, accident and emergency (A&E) and day hospital attendances where necessary.[55] For unit costs of services not obtained from these two sources, the unit costs were estimated directly based on the salary of the professional delivering it, plus employer on-costs (superannuation and national insurance), overhead costs (administrative managerial, capital), the costs of supervision and any equipment or consumables costs. Costs which were older or more recent than the chosen financial year were inflated using inflation indices reported by the PSSRU.

Cost of productivity losses for those in employment were calculated by combining time off work with average daily earnings, for the participant, or average earning of home care worker for the principal carer.[56]

Data cleaning and missing data

For missing EQ-5D index values at baseline, mean imputation was used to fill in each missing index value independent of the treatment arm. The same approach was used to impute missing data for the total and different categories of costs at baseline. Missing EQ-5D index values at all follow-up points were assumed to be MAR. The missing observations were replaced with a set of imputed values through multiple imputations drawn from posterior predictive distribution given the missing observed data.[57] Missing total cost data at all follow-up points were imputed at the cost level using the same approach outlined for dealing with missing EQ-5D index values at follow-up.

Analysis of service and support use

Resource use by study participants were reported as the mean by arm and as a percentage of the arm who had at least one contact. Differences in the use of services between arms at baseline and over the 52-week follow-up were reported descriptively and not compared statistically to avoid problems associated with multiple testing and because the focus of the economic evaluation is on a quantitative analysis of cost and cost-effectiveness.

Analysis of costs

The economic analysis followed the ITT principle. Costs were calculated using data on the type, number and length of contacts received by each participant. The mean difference in total cost from NHS/PSS and from a societal perspective (NHS/PSS in addition to productivity cost) between the randomised arms over the 52-week follow-up were estimated from generalised linear model with gamma family and log link with participants' baseline costs and randomisation stratification variables as covariates. The associated 95% CIs were estimated using bias-corrected non-parametric bootstrapping. Discounting was not applied as costing data were assessed at 52 weeks.

Analysis of outcomes

Two outcome measures were used in the economic evaluation: QALYs (primary outcome measure) and the QIDS-SR at 52 weeks (secondary outcome measure). Weekly QIDS-SR scores were used to estimate the overall burden of depressive symptom severity. The mean difference in QIDS-SR scores between arms at 52 weeks was estimated from a multilevel linear regression model with the QIDS-SR as the outcome, a treatment by time interaction term, and random intercepts for individuals to account for the correlation between repeated measures. The model was

conditioned on potential baseline confounders and randomisation stratification variables in an adjusted multilevel linear regression model.

Quality-adjusted life-years accrued over the follow-up period were calculated based on the utility scores derived from the EQ-5D-3L using AUC methods, assuming a linear change between any two adjacent time points. The mean difference in QALY over the 52-week follow-up period between the two arms was estimated using ordinary least square regression. Baseline utility score and randomisation stratification variables were used as covariates.

Non-parametric bootstrap methods were used to estimate 95% CIs for both QALYs and QIDS-SR. Discounting was not applied as outcome data were assessed at 52 weeks.

Cost-effectiveness analysis

Results of the cost-effectiveness analysis were reported as incremental cost-effectiveness ratios (ICERs) and incremental net benefit (INB). ICERs were calculated only for any cost–outcome combinations for which there were higher costs and greater benefits for one arm compared with the other, or when there was a trade-off between less effect for less cost. The cost-effectiveness of an intervention was then determined by whether the ICER value fell above or below the decision-maker's willingness to pay (WTP) per QALY, or per point improvement on the QIDS-SR.

Incremental net benefit expresses the adjusted mean difference in benefit in terms of QALYs/QIDS-SR score by converting the adjusted mean difference in total cost between the intervention and the control on to the QALY scale/QIDS-SR scale using the given threshold value. Unlike the ICER, for which interpretation depends on whether the incremental cost and effect are positive or negative, the interpretation of INB is straightforward: the intervention is cost-effective if its INB is positive.

Uncertainty around the cost-effectiveness analysis was explored using cost-effectiveness planes and cost-effectiveness acceptability curves (CEACs) based on the net benefit approach.[58] Cost-effectiveness planes plot the adjusted mean differences in total cost, QALYs and QIDS-SR score. Differences were calculated using the bootstrapped results associated with the regression models described above. The CEAC was derived by calculating the proportion of bootstrapped estimates that were cost-effective across a range of WTP thresholds, to show the probability that the intervention is cost-effective across different threshold values.

Sensitivity analyses

The following sensitivity analyses were carried out to test the robustness of the conclusions under different assumptions and/or different sets of parameters:

1. Changing the cost of the intervention drug from the cheapest drug (used in the base case analysis) to the cost of the generic drug.
2. Testing alternative assumptions of missing data. The analysis of costs and outcomes were rerun assuming that the missing data were covariate dependent missing completely at random (CD-MCAR). All baseline variables that might predict both outcomes/costs and missingness were used in the regression analysis.
3. Testing the potential impact of COVID-19 on resource use and accordingly include whether participants' follow-up overlapped with the COVID-19 pandemic as a binary covariant in missing data imputation, and then re-running the analysis.

Chapter 3 Clinical effectiveness results

Data description

Overall, 212 participants were recruited at the 6 trial sites between December 2016 and July 2021. Follow-up lasted until July 2022. *Figure 2* presents the Consolidated Standards of Reporting Trials flow diagram summarising participant throughput from recruitment, eligibility screening, randomisation, and completion of each follow-up. Exclusions were mostly due to not meeting TRD criteria, or physician concern about enrolling the individual in the study. One hundred and seven participants were randomised to prescription of quetiapine and 105 to prescription of lithium. Prescription

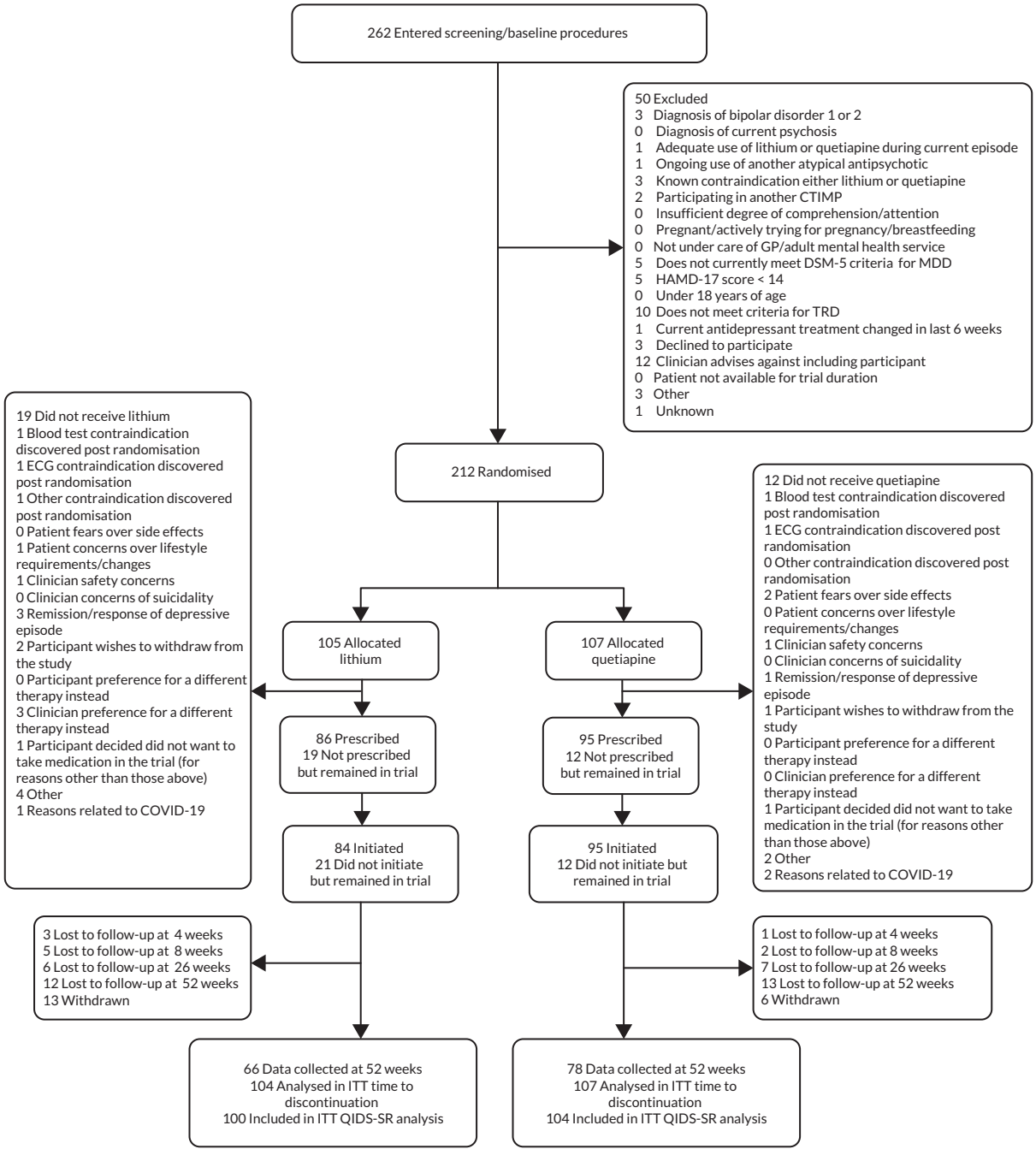

FIGURE 2 Consolidated Standards of Reporting Trials diagram.

and initiation are described in more detail later in the report, but briefly: 95 of the 107 participants randomised to the quetiapine arm were prescribed and initiated the medication, with 12 not being prescribed; and 86 of the 105 participants randomised to the lithium arm were prescribed, with 84 initiating, 19 not being prescribed, and 2 not initiating. Note that those participants who did not receive trial medication or were lost to follow-up generally remained in the study and have been included in the primary analyses, unless there was no information on discontinuation status or no QIDS-SR data over the entire 52 weeks.

Comparability of randomised arms

Baseline sociodemographic and clinical characteristics, as well as randomisation stratification variable descriptives are displayed in *Table 2*. There were no missing sociodemographic data at baseline. The mean age was 42.4 years with a SD of 14.0. Overall, age was balanced between the arms with a mean (SD) of 41.6 (14.6) years in the quetiapine arm and a mean (SD) of 43.2 (13.3) years in the lithium arm. The majority of participants were female (54.2% overall), and this was also the case for those randomised to quetiapine (60.7% female). However, there was more even split in the lithium arm (51.4% male), an apparent chance imbalance in sex between arms. The majority of the sample was from a white background (88.7% overall); this is also reflected in each arm (89.7% in the quetiapine arm and 87.6% in the lithium arm).

TABLE 2 Sociodemographic and clinical characteristics at baseline and randomisation stratification variable descriptive statistics

		Quetiapine	Lithium	Overall
Total	*N*	107	105	212
Age	Mean (SD)	41.6 (14.6)	43.2 (13.3)	42.4 (14.0)
	Median (IQR)	40.0 (29.0–52.0)	43.0 (33.0–53.0)	42.0 (31.0–53.0)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
		n (%)	*n* (%)	*n* (%)
Sex	Male	42 (39.3)	54 (51.4)	96 (45.3)
	Female	65 (60.7)	50 (47.6)	115 (54.2)
	Intersex	0 (0.0)	0 (0.0)	0 (0.0)
	Female to male	0 (0.0)	1 (1.0)	1 (0.5)
	Male to female	0 (0.0)	0 (0.0)	0 (0.0)
	Neither	0 (0.0)	0 (0.0)	0 (0.0)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
Ethnicity	White background	96 (89.7)	92 (87.6)	188 (88.7)
	Mixed/Multiple ethnic background	4 (3.7)	3 (2.9)	7 (3.3)
	Asian background	2 (1.9)	7 (6.7)	9 (4.2)
	Black/African/Caribbean background	2 (1.9)	2 (1.9)	4 (1.9)
	Any other background	2 (1.9)	1 (1.0)	3 (1.4)
	Unrecorded	1 (0.9)	0 (0.0)	1 (0.5)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
Highest completed level of education	Primary education or less (no formal qualifications)	1 (0.9)	6 (5.7)	7 (3.3)

TABLE 2 Sociodemographic and clinical characteristics at baseline and randomisation stratification variable descriptive statistics (*continued*)

		Quetiapine	Lithium	Overall
	Secondary education (GCSE, O Levels)	16 (15.0)	15 (14.3)	31 (14.6)
	College-level education or equivalent (A Level, NVQ, International Baccalaureate, BTEC nationals)	44 (41.1)	37 (35.2)	81 (38.2)
	Degree level education/ Diploma (e.g. BSc, BA)	28 (26.2)	31 (29.5)	59 (27.8)
	Post-graduate degree (e.g. MSc, MA, PhD)	18 (16.8)	16 (15.2)	34 (16.0)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
Current main employment status	Paid employment	54 (50.5)	49 (46.7)	103 (48.6)
	Unemployed	29 (27.1)	43 (41.0)	72 (34.0)
	Student	10 (9.3)	4 (3.8)	14 (6.6)
	Housewife/husband	1 (0.9)	2 (1.9)	3 (1.4)
	Retired	13 (12.1)	7 (6.7)	20 (9.4)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
		n (%)	*n* (%)	*n* (%)
Currently taking a selective serotonin reuptake inhibitor (SSRI)	Yes	51 (47.7)	48 (45.7)	99 (46.7)
	No	55 (51.4)	57 (54.3)	112 (52.8)
	Missing	1 (0.9)	0 (0.0)	1 (0.5)
Currently taking a tricyclic antidepressant (TCA)	Yes	12 (11.2)	9 (8.6)	21 (9.9)
	No	94 (87.9)	96 (91.4)	190 (89.6)
	Missing	1 (0.9)	0 (0.0)	1 (0.5)
Currently taking a monoamine oxidase inhibitor (MAOI)	Yes	0 (0.0)	0 (0.0)	0 (0.0)
	No	106 (99.1)	105 (100.0)	211 (99.5)
	Missing	1 (0.9)	0 (0.0)	1 (0.5)
Currently taking a tetracyclic antidepressant (and related)	Yes	29 (27.1)	33 (31.4)	62 (29.2)
	No	77 (72.0)	72 (68.6)	149 (70.3)
	Missing	1 (0.9)	0 (0.0)	1 (0.5)
Currently taking a serotonin–noradrenaline reuptake inhibitor (SNRI)	Yes	35 (32.7)	39 (37.1)	74 (34.9)
	No	71 (66.4)	66 (62.9)	137 (64.6)
	Missing	1 (0.9)	0 (0.0)	1 (0.5)
Currently taking a serotonin modulator and stimulator	Yes	4 (3.7)	4 (3.8)	8 (3.8)

continued

TABLE 2 Sociodemographic and clinical characteristics at baseline and randomisation stratification variable descriptive statistics (*continued*)

		Quetiapine	Lithium	Overall
	No	102 (95.3)	101 (96.2)	203 (95.8)
	Missing	1 (0.9)	0 (0.0)	1 (0.5)
Currently taking a noradrenaline reuptake inhibitor (NRI)	Yes	0 (0.0)	0 (0.0)	0 (0.0)
	No	106 (99.1)	105 (100.0)	211 (99.5)
	Missing	1 (0.9)	0 (0.0)	1 (0.5)
Currently taking a noradrenaline and dopamine reuptake inhibitor	Yes	0 (0.0)	0 (0.0)	0 (0.0)
	No	106 (99.1)	105 (100.0)	211 (99.5)
	Missing	1 (0.9)	0 (0.0)	1 (0.5)
Currently taking a melatonin agonist	Yes	1 (0.9)	1 (1.0)	2 (0.9)
	No	105 (98.1)	104 (99.0)	209 (98.6)
	Missing	1 (0.9)	0 (0.0)	1 (0.5)
Currently taking 'Other' class of antidepressant	Yes	1 (0.9)	1 (1.0)	2 (0.9)
	No	105 (98.1)	104 (99.0)	209 (98.6)
	Missing	1 (0.9)	0 (0.0)	1 (0.5)
Received psychological therapy for depression in current episode	Yes	71 (66.4)	71 (67.6)	142 (67.0)
	No	34 (31.8)	34 (32.4)	68 (32.1)
	Missing, n(%)	2 (1.9)	0 (0.0)	2 (0.9)
Number of comorbid psychiatric diagnoses	Mean (SD)	1.8 (1.6)	1.9 (1.5)	1.8 (1.6)
	Median (IQR)	1.0 (0.0–3.0)	2.0 (1.0–3.0)	1.5 (1.0–3.0)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
HAMD-17 score	Mean (SD)	21.2 (5.0)	21.9 (5.4)	21.5 (5.2)
	Median (IQR)	20.0 (17.0–24.0)	21.0 (18.0–25.0)	20.0 (18.0–25.0)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
Weight (kg)	Mean (SD)	88.4 (22.5)	88.3 (22.1)	88.4 (22.2)
	Median (IQR)	86.8 (73.4–98.6)	88.6 (71.0–102.5)	87.6 (72.0–101.8)
	Missing, *n* (%)	14 (13.1)	6 (5.7)	20 (9.4)
Waist circumference (cm)	Mean (SD)	99.3 (17.1)	101.4 (22.0)	100.3 (19.7)
	Median (IQR)	99.0 (86.0–110.5)	99.0 (88.0–113.3)	99.0 (87.5–112.0)
	Missing, *n* (%)	28 (26.2)	26 (24.8)	54 (25.5)
Systolic blood pressure (mmHg)	Mean (SD)	131.1 (16.8)	130.7 (17.2)	130.9 (16.9)
	Median (IQR)	130.0 (118.5–141.0)	129.0 (120.0–143.0)	129.0 (120.0–142.0)
	Missing, *n* (%)	15 (14.0)	14 (13.3)	29 (13.7)

TABLE 2 Sociodemographic and clinical characteristics at baseline and randomisation stratification variable descriptive statistics (*continued*)

		Quetiapine	Lithium	Overall
Diastolic blood pressure (mmHg)	Mean (SD)	77.8 (10.9)	79.7 (11.4)	78.7 (11.1)
	Median (IQR)	79.0 (70.0–84.0)	80.0 (72.0–86.0)	79.5 (71.0–85.0)
	Missing, *n* (%)	16 (15.0)	14 (13.3)	30 (14.2)
Pulse rate (bpm)	Mean (SD)	76.5 (15.2)	74.3 (12.1)	75.4 (13.8)
	Median (IQR)	73.5 (65.0–85.0)	74.0 (66.0–82.0)	74.0 (65.0–82.0)
	Missing, *n* (%)	17 (15.9)	16 (15.2)	33 (15.6)
MARS-5 score	Mean (SD)	23.5 (1.9)	23.2 (2.1)	23.4 (2.0)
	Median (IQR)	24.0 (23.0–25.0)	24.0 (22.0–25.0)	24.0 (23.0–25.0)
	Missing, *n* (%)	3 (2.8)	5 (4.8)	8 (3.8)
QIDS-SR score	Mean (SD)	17.2 (4.4)	18.0 (4.1)	17.6 (4.2)
	Median (IQR)	17.0 (15.0–21.0)	18.0 (15.0–21.0)	18.0 (15.0–21.0)
	Missing, *n* (%)	0 (0.0)	1 (1.0)	1 (0.5)
WSAS score	Mean (SD)	27.2 (7.1)	27.9 (7.8)	27.5 (7.5)
	Median (IQR)	28.0 (23.0–31.0)	28.4 (22.5–34.0)	28.0 (23.0–33.0)
	Missing, *n* (%)	1 (0.9)	3 (2.9)	4 (1.9)
HCL-16 score	Mean (SD)	6.6 (3.7)	7.0 (3.9)	6.8 (3.8)
	Median (IQR)	7.0 (4.0–10.0)	7.5 (4.0–10.0)	7.0 (4.0–10.0)
	Missing, *n* (%)	4 (3.7)	7 (6.7)	11 (5.2)
MADRS score	Mean (SD)	30.8 (6.4)	31.5 (7.2)	31.1 (6.8)
	Median (IQR)	31.0 (27.0–35.0)	30.5 (26.0–36.0)	31.0 (26.0–36.0)
	Missing, *n* (%)	0 (0.0)	1 (1.0)	1 (0.5)
GAD-7	Mean (SD)	12.4 (5.6)	12.6 (5.7)	12.5 (5.6)
	Median (IQR)	13.0 (8.0–17.0)	13.0 (8.0–18.0)	13.0 (8.0–17.0)
	Missing, *n* (%)	0 (0.0)	2 (1.9)	2 (0.9)
		n (%)	*n* (%)	*n* (%)
Severity of mental illness	Normal, not at all ill	1 (0.9)	0 (0.0)	1 (0.5)
	Borderline mentally ill	0 (0.0)	0 (0.0)	0 (0.0)
	Mildly ill	10 (9.3)	9 (8.6)	19 (9.0)
	Moderately ill	55 (51.4)	55 (52.4)	110 (51.9)
	Markedly ill	31 (29.0)	31 (29.5)	62 (29.2)
	Severely ill	10 (9.3)	10 (9.5)	20 (9.4)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
Randomisation stratification variable descriptive statistics				
	n (%)	*n* (%)	*n* (%)	
Study site	Bristol	4 (3.7)	5 (4.8)	9 (4.2)

continued

TABLE 2 Sociodemographic and clinical characteristics at baseline and randomisation stratification variable descriptive statistics (*continued*)

		Quetiapine	Lithium	Overall
	Brighton	7 (6.5)	7 (6.7)	14 (6.6)
	London	42 (39.3)	41 (39.0)	83 (39.2)
	Newcastle	26 (24.3)	27 (25.7)	53 (25.0)
	Oxford	28 (26.2)	25 (23.8)	53 (25.0)
TRD severity	2 previous antidepressants	43 (40.2)	42 (40.0)	85 (40.1)
	More than 2 previous antidepressants	64 (59.8)	63 (60.0)	127 (59.9)
Depression severity	Moderate	36 (33.6)	33 (31.4)	69 (32.5)
	Severe	32 (29.9)	33 (31.4)	65 (30.7)
	Very severe	39 (36.4)	39 (37.1)	78 (36.8)

Overall, 38.2% of participants had a college-level education or equivalent as the highest completed level of education, and this was similar within each arm (41.1% in the quetiapine arm; 35.2% in the lithium arm). Overall, 48.6% were in paid employment, and again, this is similar between arms (50.5% in the quetiapine arm; 46.7% in the lithium arm). However, there was an apparent imbalance between arms for participants who were unemployed. Overall, 34.0% of participants were unemployed. However, there was a greater proportion of participants unemployed in the lithium arm (41.0%) compared to the quetiapine arm (27.1%).

Regarding clinical characteristics, there were no data missing for number of comorbidities (as recorded using the MINI v7.0), HAMD-17 score (an inclusion criteria measure) and the severity of mental illness. All other measures had various degrees of missing data. Waist circumference had the highest proportion of missing data (25.5% overall with 26.2% in the quetiapine arm, and 24.8% in the lithium arm). Pulse rate, diastolic blood pressure and systolic blood pressure were all missing similar proportions of data, (15.6%, 14.2% and 13.7% overall, respectively). Weight was missing in 9.4% of the sample overall, but there was larger proportion of participants in the quetiapine arm missing weight (13.1%) than in the lithium arm (5.7%). All other measures had < 6% missing data overall with similar proportions missing between arms.

The mean (SD) QIDS-SR score was 17.6 (4.2) overall and appeared to be relatively balanced between arms [17.2 (4.4) in the quetiapine arm and 18.0 (4.1) in the lithium arm] at baseline. All other clinical measures were relatively balanced between arms. We note one participant was recorded as having a 'normal' severity of mental illness at baseline.

The randomisation process was successful in achieving balance between the randomised arms on the pre-specified factors. Most participants were recruited from the London site (39.2%), and the majority had more than two previous antidepressants (59.9% overall). There was a relatively even split across depression severity categories.

Treatment fidelity and adherence

Treatment fidelity and adherence to allocated treatment is summarised in *Table 3*. The trial medication was prescribed for 88.8% (95 participants) of those randomised to quetiapine and 81.9% (86 participants) to those randomised to lithium. Of those 95 prescribed and randomised to quetiapine, all 95 started taking the medication, and 37 of these (38.9%) discontinued treatment before 12 months. Of those 86 prescribed and randomised to lithium, 84 started taking the medication. Of the 84 that initiated, 42 (50.0%) discontinued treatment before 12 months, with discontinuation date missing for 1 participant. Mean (IQR) dose of quetiapine was 194.8 mg (127.8) and for lithium was 681.3 mg (421.3).

TABLE 3 Adherence to allocated treatment and treatment fidelity descriptive statistics

		Quetiapine	Lithium	Overall
The trial medication was prescribed	*N*	107	105	212
	n (%) of those randomised	95 (88.8)	86 (81.9)	181 (85.4)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
The participant started taking the medication	*N*	95	86	181
	n (%) of those prescribed	95 (100.0)	84 (97.7)	179 (98.9)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
Participant was not prescribed treatment	*N*	107	105	212
	n (%) of those randomised	12 (11.2)	19 (18.1)	31 (14.6)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
The treatment was discontinued before 12 months	*N*	95	84	178
	n (%) of those initiated	37 (38.9)	42 (50.0)	79 (44.4)
	Missing, *n* (%)	0 (0.0)	1 (1.2)	1 (0.6)
MARS-5 score at baseline	*N*	104	100	204
	Mean (SD)	23.5 (1.9)	23.2 (2.1)	23.4 (2.0)
	Median (IQR)	24.0 (23.0–25.0)	24.0 (22.0–25.0)	24.0 (23.0–25.0)
	Missing, *n* (%)	3 (2.8)	5 (4.8)	8 (3.8)
MARS-5 score at week 8	*N*	75	71	146
	Mean (SD)	24.3 (1.6)	24.3 (1.6)	24.3 (1.6)
	Median (IQR)	25.0 (24.0–25.0)	25.0 (24.0–25.0)	25.0 (24.0–25.0)
	Missing, *n* (%)	32 (29.9)	34 (32.4)	66 (31.1)
MARS-5 score at week 26	*N*	67	61	128
	Mean (SD)	24.0 (1.7)	24.1 (1.3)	24.0 (1.5)
	Median (IQR)	25.0 (24.0–25.0)	24.0 (24.0–25.0)	25.0 (24.0–25.0)
	Missing, *n* (%)	40 (37.4)	44 (41.9)	84 (39.6)
MARS-5 score at week 52	*N*	60	47	107
	Mean (SD)	23.9 (2.5)	24.1 (1.3)	24.0 (2.1)
	Median (IQR)	25.0 (24.0–25.0)	25.0 (24.0–25.0)	25.0 (24.0–25.0)
	Missing, *n* (%)	47 (43.9)	58 (55.2)	105 (49.5)

The MARS-5 measured medication adherence. At baseline, this referred to adherence to patients' current antidepressant, whereas at follow-up visits the MARS-5 measured adherence to the trial medication. The mean and median for both arms were very similar at each time point, and across all post-randomisation time points. There was a slightly lower proportion of missing MARS-5 scores in the quetiapine arm compared to those in the lithium arm (2.8% compared to 4.8% at baseline; 29.9% compared to 32.4% at week 8; 37.4% compared to 41.9% at week 26; and 43.9% compared to 55.2% at week 52).

Missing outcome data and participant withdrawal

Table 4 summarises missing outcome data and withdrawal of participants from the trial. There was a higher rate of missing data in the lithium arm compared to the quetiapine arm across outcome measures. At week 26, 25.2% of those randomised to quetiapine were missing QIDS-SR scores compared to 39.0% randomised to lithium. At week 52, 28.0% of those in the quetiapine arm were missing QIDS-SR scores compared to 35.2% in the lithium arm. As noted previously, there was one participant in the lithium arm for whom time to discontinuation of the trial medication could

TABLE 4 Missing data and participant withdrawal descriptive statistics

	Quetiapine	Lithium	Overall
N	107	105	212
Primary outcomes:	***n (%)***	***n (%)***	***n (%)***
Missing QIDS-SR score at baseline	0 (0.0)	1 (1.0)	1 (0.5)
Missing QIDS-SR score at week 8	21 (19.6)	17 (16.2)	38 (17.9)
Missing QIDS-SR score at week 26	27 (25.2)	41 (39.0)	68 (32.1)
Missing QIDS-SR score at week 52	30 (28.0)	37 (35.2)	67 (31.6)
Missing time to discontinuation/follow-up	0 (0.0)	1 (1.0)	1 (0.5)
Secondary outcomes			
Missing MADRS score at baseline	0 (0.0)	1 (1.0)	1 (0.5)
Missing MADRS score at week 8	19 (17.8)	26 (24.8)	45 (21.2)
Missing MADRS score at week 26	27 (25.2)	40 (38.1)	67 (31.6)
Missing MADRS score at week 52	38 (35.5)	51 (48.6)	89 (42.0)
Missing WSAS score at baseline	1 (0.9)	3 (2.9)	4 (1.9)
Missing WSAS score at week 8	23 (21.5)	17 (16.2)	40 (18.9)
Missing WSAS score at week 26	27 (25.2)	40 (38.1)	67 (31.6)
Missing WSAS score at week 52	30 (28.0)	40 (38.1)	70 (33.0)
Missing weight (kg) at baseline	14 (13.1)	6 (5.7)	20 (9.4)
Missing weight (kg) at week 8	27 (25.2)	31 (29.5)	58 (27.4)
Missing weight (kg) at week 26	45 (42.1)	56 (53.3)	101 (47.6)
Missing weight (kg) at week 52	62 (57.9)	70 (66.7)	132 (62.3)
Missing diastolic blood pressure (mmHg) at baseline	16 (15.0)	14 (13.3)	30 (14.2)
Missing diastolic blood pressure (mmHg) at week 8	31 (29.0)	36 (34.3)	67 (31.6)
Missing diastolic blood pressure (mmHg) at week 26	44 (41.1)	59 (56.2)	103 (48.6)
Missing diastolic blood pressure (mmHg) at week 52	60 (56.1)	69 (65.7)	129 (60.8)

TABLE 4 Missing data and participant withdrawal descriptive statistics (*continued*)

		Quetiapine	Lithium	Overall
Missing systolic blood pressure (mmHg) at baseline		15 (14.0)	14 (13.3)	29 (13.7)
Missing systolic blood pressure (mmHg) at week 8		31 (29.0)	36 (34.3)	67 (31.6)
Missing systolic blood pressure (mmHg) at week 26		44 (41.1)	59 (56.2)	103 (48.6)
Missing systolic blood pressure (mmHg) at week 52		60 (56.1)	69 (65.7)	129 (60.8)
Missing PRISE total score at week 8		25 (23.4)	27 (25.7)	52 (24.5)
Missing PRISE total score at week 26		43 (40.2)	47 (44.8)	90 (42.5)
Missing PRISE total score at week 52		50 (46.7)	56 (53.3)	106 (50.0)
Withdrawals		***n (%)***	***n (%)***	***n (%)***
Withdrawn		6 (5.6)	13 (12.4)	19 (9.0)
Main reason for withdrawal (*n*% out total withdrawn)	Participant no longer willing to participate in trial	5 (83.3)	10 (76.9)	15 (78.9)
	Lost to follow-up	0 (0.0)	0 (0.0)	0 (0.0)
	Death	0 (0.0)	1 (7.7)	1 (5.3)
	Other	1 (16.7)	1 (7.7)	2 (10.5)
	Unknown	0 (0.0)	1 (7.7)	1 (5.3)
	Total	6 (5.6)	13 (12.4)	19 (9.0)

not be calculated as their discontinuation date was missing. Considering secondary outcomes at week 52, 35.5% of participants in the quetiapine arm were missing MADRS scores compared to 48.6% in the lithium arm. Twenty-eight per cent of participants in the quetiapine arm were missing WSAS scores compared to 38.1% in the lithium arm. Overall, 57.9% of participants in the quetiapine arm were missing weight measures at 52 weeks compared to 66.7% in the lithium arm. Overall, 56.1% in the quetiapine arm compared to 65.7% in the lithium arm were missing diastolic and systolic blood pressure.

In total, 19 (9.0%) participants actively withdrew from the trial, 6 (5.6%) participants from the quetiapine arm, and 13 (12.4%) in the lithium arm. For both arms the main reason cited was that the participant was no longer willing to participate.

Serious adverse event reporting

Table 5 summarises SAEs reported during the trial. There were 32 SAEs for 18 participants. Fifteen SAEs were for 7 participants randomised to the quetiapine arm, where 46.7% of SAEs were deemed unlikely to be related to, and 53.3% were not related to quetiapine treatment. Seventeen SAEs were from 11 participants randomised to the lithium arm, where one event (equating to 5.9%) was possibly related to lithium treatment and was therefore a SAR, 41.2% were unlikely to be related, and 52.9% were deemed not related to lithium. There were no SUSARs.

Descriptive statistics for outcome measures

Primary outcomes

Self-rated quick inventory of depressive symptomatology scores are displayed in *Table 6*. QIDS-SR scores decreased over the trial follow-up period, from a mean (SD) of 14.7 (5.6), overall, at week 8–13.1 (6.2), overall, at week 52. This decrease in raw scores was larger in the quetiapine arm than in the lithium arm, with a decrease from 14.4 (5.5) at

TABLE 5 Serious adverse event reporting

		Quetiapine	Lithium	Overall
Number of SAEs	*N*	15	17	32
Related to lithium or quetiapine add on therapy, *n* (%)	Definitely related	0 (0.0)	0 (0.0)	0 (0.0)
	Likely related	0 (0.0)	0 (0.0)	0 (0.0)
	Possibly related	0 (0.0)	1 (5.9)	1 (3.1)
	Unlikely related	7 (46.7)	7 (41.2)	14 (43.8)
	Not related	8 (53.3)	9 (52.9)	17 (53.1)
Number of participants with SAEs	*N*	7	11	18

TABLE 6 Self-rated Quick Inventory of Depressive Symptomatology scores descriptive statistics

		Quetiapine, *N* = 107	Lithium, *N* = 105	Overall, *N* = 212
QIDS-SR score at week 8	Mean (SD)	14.4 (5.5)	15.1 (5.7)	14.7 (5.6)
	Median (IQR)	14.5 (10.0–18.0)	15.5 (11.0–19.5)	15.0 (11.0–19.0)
QIDS-SR score at week 26	Mean (SD)	13.0 (5.8)	14.1 (6.1)	13.5 (5.9)
	Median (IQR)	12.5 (8.0–17.0)	14.0 (10.0–19.0)	13.0 (9.0–19.0)
QIDS-SR score at week 52	Mean (SD)	12.1 (6.2)	14.1 (6.1)	13.1 (6.2)
	Median (IQR)	12.0 (8.0–16.0)	14.5 (10.0–19.0)	13.0 (9.0–18.0)

week 8−12.1 (6.2) at week 52 in the quetiapine arm compared to 15.1 (5.7) at week 8−14.1 (6.1) at week 52 in the lithium arm. At each time point, participants in the quetiapine arm had a lower average QIDS-SR score than those in the lithium arm.

Figure 3 shows the unadjusted mean QIDS-SR scores, and 95% CIs, at study visits (baseline, week 8, week 26 and week 52), by trial arm. *Figure 4* shows the unadjusted mean QIDS-SR scores, and 95% CIs, collected at study visits and each week via True Colours, by trial arm. The scores were initially similar in the two arms, but over the follow-up the QIDS-SR scores decreased more in the quetiapine arm, leading to greater separation in the QIDS-SR profiles of the two arms over time.

Discontinuation data are displayed in *Table 7*. The median (IQR) time to trial treatment discontinuation in days was shorter in the lithium arm, 212.0 (21.0–365.0), compared to the quetiapine arm, 365.0 (57.0–365.0). The time to prescription was similar in the two arms, with the time to treatment initiation in days shorter in the quetiapine arm median (IQR) 16.5 (6.0–33.0), compared to 20.0 (8.0–34.0) in the lithium arm, likely due to the potentially longer process involved in ensuring individuals are suitable for lithium treatment.

In the quetiapine arm, 95 participants were prescribed the treatment, the main reasons to not prescribe treatment included: patient fears over side effects, reasons related to COVID-19, and other. In the lithium arm, 86 participants were prescribed the treatment, the main reasons to not prescribe treatment were: remission/response of depressive episode, clinician preference for a different therapy instead, and other. All those prescribed treatment in the quetiapine arm initiated treatment; however, of the 86 prescribed lithium, only 84 initiated treatment. Reasons for not commencing trial medication were not collected.

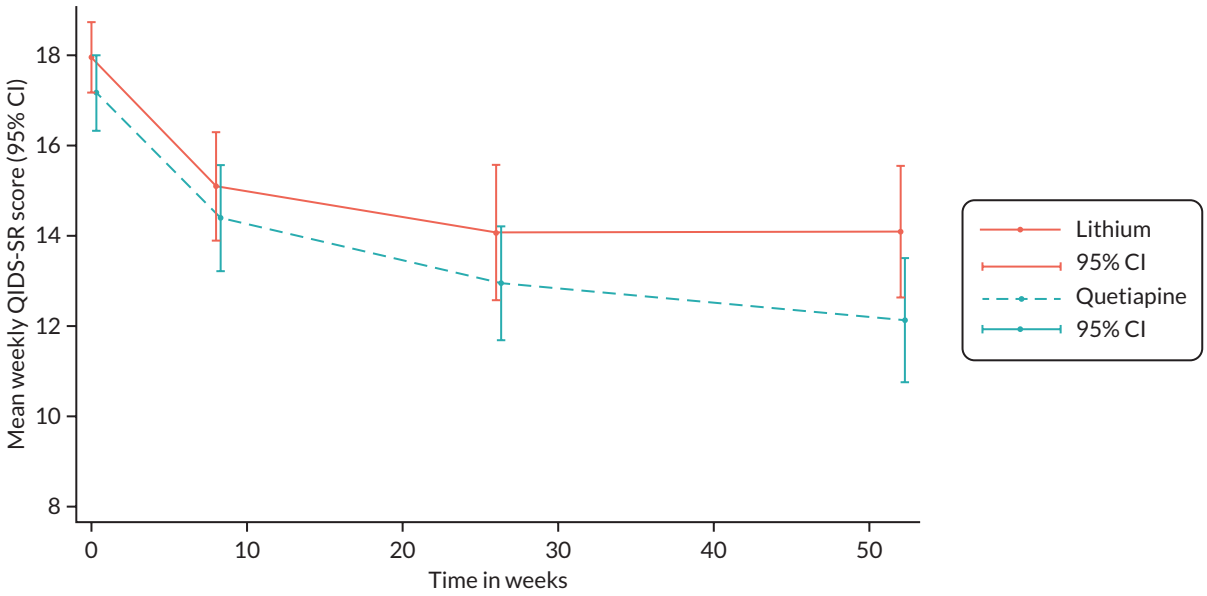

FIGURE 3 Mean QIDS-SR scores at baseline and follow-up visits (unadjusted).

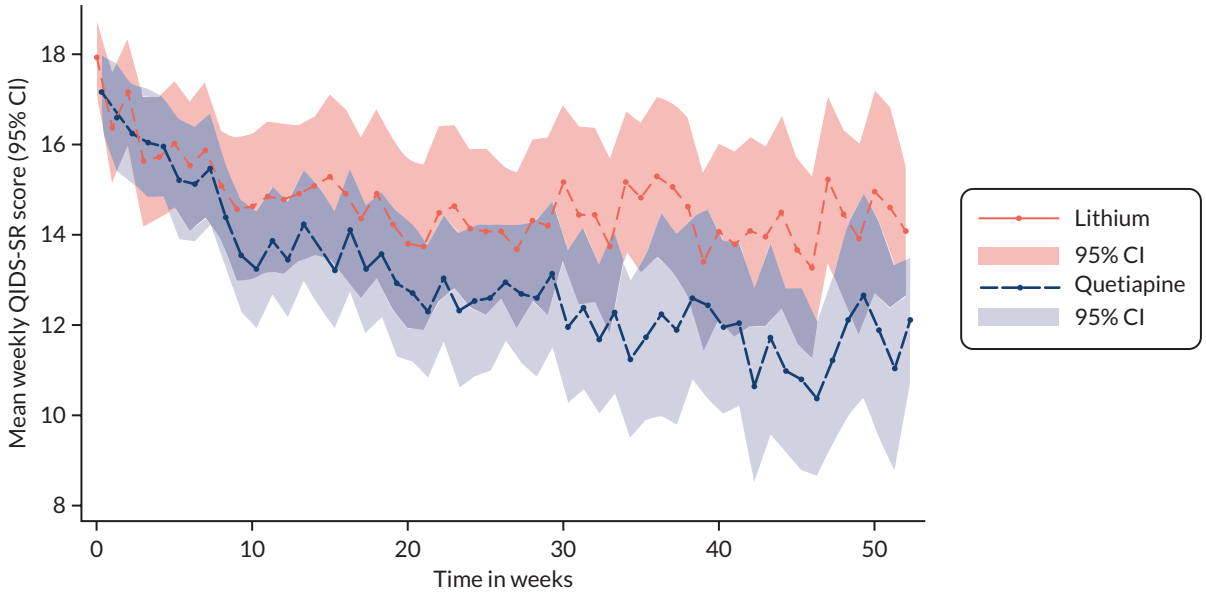

FIGURE 4 Mean weekly QIDS-SR scores (unadjusted).

TABLE 7 Trial medication discontinuation descriptive statistics

Outcomes		Quetiapine, *N* = 107	Lithium, *N* = 105	Overall, *N* = 212
Time to discontinuation (days)	Mean (SD)	228.1 (158.1)	201.8 (154.0)	215.1 (156.3)
	Median (IQR)	365.0 (57.0–365.0)	212.0 (21.0–365.0)	266.0 (47.0–365.0)
Time to prescription (days)	Mean (SD)	23.7 (35.4)	22.3 (23.8)	23.1 (30.3)
	Median (IQR)	14.5 (4.0–30.0)	15.0 (7.0–29.0)	15.0 (6.0–30.0)

continued

TABLE 7 Trial medication discontinuation descriptive statistics (*continued*)

Outcomes		Quetiapine, *N* = 107	Lithium, *N* = 105	Overall, *N* = 212
Time to initiation (days)	Mean (SD)	26.8 (36.1)	25.7 (24.8)	26.3 (31.2)
	Median (IQR)	16.5 (6.0–33.0)	20.0 (8.0–34.0)	19.5 (7.0–33.0)
		n (%)	*n* (%)	*n* (%)
Prescribed treatment		95 (88.8)	86 (81.9)	181 (85.4)
Not prescribed		12 (11.2)	19 (18.1)	31 (14.6)
Reasons why treatment was not prescribed	Blood test contraindication discovered post randomisation	1 (8.3)	1 (5.3)	2 (6.5)
	ECG contraindication discovered post randomisation	1 (8.3)	1 (5.3)	2 (6.5)
	Other contraindication discovered post randomisation	0 (0.0)	1 (5.3)	1 (3.2)
	Patient fears over side effects	2 (16.7)	0 (0.0)	2 (6.5)
	Patient concerns over lifestyle requirements/changes	0 (0.0)	1 (5.3)	1 (3.2)
	Clinician safety concerns	1 (8.3)	1 (5.3)	2 (6.5)
	Clinician concerns of suicidality	0 (0.0)	0 (0.0)	0 (0.0)
	Remission/response of depressive episode	1 (8.3)	3 (15.8)	4 (12.9)
	Participant wishes to withdraw from the study	1 (8.3)	2 (10.5)	3 (9.7)
	Participant preference for a different therapy instead	0 (0.0)	0 (0.0)	0 (0.0)
	Clinician preference for a different therapy instead	0 (0.0)	3 (15.8)	3 (9.7)
	Participant decided did not want to take medication in the trial (for reasons other than those above)	1 (8.3)	1 (5.3)	2 (6.5)
	Other	2 (16.7)	4 (21.1)	6 (19.4)
	Reasons related to COVID-19	2 (16.7)	1 (5.3)	3 (9.7)
Initiated treatment		95 (88.8)	84 (80.0)	179 (84.4)
Discontinued treatment	No	58 (54.2)	41 (39.0)	99 (46.7)
	Yes	37 (34.6)	42 (40.0)	79 (37.3)
	Unknown	0 (0.0)	1 (1.0)	1 (0.5)
	Did not initiate treatment	12 (11.2)	21 (20.0)	33 (15.6)
Reason for discontinuation	Side effects (adverse event/serious adverse event)	17 (45.9)	14 (33.3)	31 (39.2)
	Clinical worsening	0 (0.0)	1 (2.4)	1 (1.3)
	Poor adherence	1 (2.7)	3 (7.1)	4 (5.1)
	Remission/response	0 (0.0)	0 (0.0)	0 (0.0)
	Death	0 (0.0)	0 (0.0)	0 (0.0)
	Inadequate clinical response	8 (21.6)	11 (26.2)	19 (24.1)
	Patient forgot (consistently > 2 weeks)	0 (0.0)	0 (0.0)	0 (0.0)
	Patient dissatisfaction with medication (not covered with other options)	3 (8.1)	3 (7.1)	6 (7.6)
	Other medical reasons (not relating to depression/side effects)	0 (0.0)	1 (2.4)	1 (1.3)
	Other	4 (10.8)	6 (14.3)	10 (12.7)
	Unknown	4 (10.8)	3 (7.1)	7 (8.9)

In the quetiapine arm, 37 participants discontinued treatment (34.6% of those randomised, 38.9% of those initiated), the main reasons given for discontinuation included side effects and inadequate clinical response. In the lithium arm, 42 participants discontinued treatment (40.0% of those randomised, 50% of those initiated), the main reasons given for discontinuation were also side effects and inadequate clinical response.

Secondary outcomes

Secondary outcomes are reported by arm and overall in *Table 8*. Mean MADRS scores decreased over the follow-up period for both arms, with the participants in the quetiapine arm scoring lower on average than those in the lithium arm, with a mean (SD) of 23.4 (9.8) compared to 25.2 (9.3) at week 8, 21.3 (10.5) compared to 23.8 (9.9) at week 26; and 20.2 (9.7) compared to 22.4 (9.2) at week 52. Unadjusted mean MADRS scores and 95% CIs by trial arm, at baseline and follow-up visits are shown in *Figure 5*.

There was a decrease in WSAS scores over time, with the participants in the quetiapine arm having lower scores than those in the lithium arm, with a mean (sd) WSAS score of 26.8 (8.8) compared to 27.3 (10.0) at week 8, 25.5 (10.3) compared to 26.7 (8.7) at week 26; and 22.2 (10.9) compared to 25.9 (10.4) at week 52. Unadjusted mean WSAS scores with 95% CIs by trial arm are shown in *Figure 6* (study visit data) and *Figure 7* (study visit and weekly True Colours data).

Weight was similar across all follow-up periods in the quetiapine arm. However, there was an increase in weight in the lithium arm over time, see *Figure 8*. Note that there were 2 outliers in the quetiapine arm and 0 in the lithium arm at baseline, and 1 outlier in the quetiapine arm and 0 in the lithium arm at weeks 8, 26 and 52.

Unadjusted mean systolic and diastolic blood pressure, with 95% CIs at baseline and follow-up visits are shown in *Figures 9* and *10*, respectively. Mean systolic and diastolic blood pressure showed little change over the follow-up time period and did not vary appreciably between arms.

Patient rated inventory of side effects mean score decreased slightly over the follow-up periods in the quetiapine arm [6.1 (3.0) at week 8, 6.0 (3.3) at week 26; and 5.2 (3.1) at week 52] but had little change in the lithium arm [6.6 (3.4) at week 8, 6.6 (3.6) at week 26; and 6.2 (4.2) at week 52]. Unadjusted mean PRISE scores, with 95% CIs, at baseline and follow-up visits, are shown in *Figure 11*.

On the Global Improvement scale, there was one participant in the quetiapine arm who was clinically deemed much worse at week 8 and one participant in the lithium arm clinically deemed much worse at week 52. Overall, there was a possible shift towards improvement in both arms over the course of follow-up.

Overall, there were 37 (17.5%) responders at the 52-week time point, 25 (23.4%) in the quetiapine arm and 12 (11.4%) in the lithium arm. There were fewer remitters at 52 weeks, 21 (9.9%) overall, with 12 (11.2%) in the quetiapine arm and 9 (8.6%) in the lithium arm. There appeared to be a larger proportion of responders in the quetiapine arm at the later time points, with the proportion increasing from 8 to 26 weeks and then decreasing again at 52 weeks for participants in the lithium arm. The proportion of remitters was broadly similar within the two arms across all the time points. There were higher proportions of both responders and remitters in the quetiapine arm.

Inferential analysis

Primary outcomes

The following analyses are ITT. As shown in *Table 9*, the model adjusted by baseline QIDS-SR score and the stratification factors yielded a significant negative quetiapine versus lithium area of −69.32 (95% CI: −131.31 to −7.32) points over 52 weeks, indicating significantly worse depression over the period in the lithium arm (p = 0.0284). The full model that had also been adjusted by baseline variables that predicted missing 52-week QIDS-SR score yielded a significant negative quetiapine versus lithium area of −68.36 (95% CI: −129.95 to −6.76), again indicating significantly worse depression in the lithium arm (p = 0.0296).

TABLE 8 Secondary outcome measure descriptive statistics by arm

Outcomes		Week 8			Week 26			Week 52		
		Quetiapine, *N* = 107	Lithium, *N* = 105	Overall, *N* = 212	Quetiapine, *N* = 107	Lithium, *N* = 105	Overall, *N* = 212	Quetiapine, *N* = 107	Lithium, *N* = 105	Overall, *N* = 212
MADRS score	Mean (SD)	23.4 (9.8)	25.2 (9.3)	24.3 (9.6)	21.3 (10.5)	23.8 (9.9)	22.4 (10.3)	20.2 (9.7)	22.4 (9.2)	21.1 (9.5)
	Median (IQR)	24.0 (16.0–31.0)	25.0 (20.0–32.0)	24.0 (17.0–31.0)	21.5 (14.0–29.5)	24.0 (15.0–32.0)	23.0 (15.0–31.0)	20.0 (12.0–28.0)	23.0 (17.0–28.0)	21.0 (14.0–28.0)
WSAS score	Mean (SD)	26.8 (8.8)	27.3 (10.0)	27.1 (9.4)	25.5 (10.3)	26.7 (8.7)	26.0 (9.6)	22.2 (10.9)	25.9 (10.4)	23.9 (10.8)
	Median (IQR)	28.0 (21.6–32.5)	30.0 (22.8–35.0)	29.0 (22.3–34.0)	29.0 (17.8–34.0)	29.0 (21.0–33.0)	29.0 (20.0–34.0)	23.8 (15.0–31.0)	28.8 (18.0–35.0)	26.1 (16.3–33.0)
Weight (kg)	Mean (SD)	87.3 (21.4)	87.1 (21.5)	87.2 (21.3)	86.1 (20.7)	88.0 (22.0)	86.9 (21.2)	87.3 (23.3)	90.7 (20.0)	88.8 (21.8)
	Median (IQR)	85.2 (71.3–98.6)	86.2 (72.8–101.1)	85.5 (72.0–99.4)	83.7 (67.5–99.0)	89.0 (72.3–100.0)	84.7 (71.0–100.0)	83.7 (72.0–100.2)	91.0 (75.6–100.6)	87.6 (72.9–100.6)
Systolic blood pressure (mmHg)	Mean (SD)	129.3 (17.3)	128.8 (15.5)	129.0 (16.4)	125.1 (19.1)	129.5 (14.2)	127.0 (17.3)	127.6 (20.6)	130.5 (15.7)	128.8 (18.6)
	Median (IQR)	126.0 (119.5–139.5)	127.0 (116.0–140.0)	127.0 (117.0–140.0)	124.0 (115.0–133.0)	128.0 (120.0–136.0)	126.0 (117.0–136.0)	126.0 (115.0–139.0)	132.0 (119.5–139.5)	130.0 (118.0–139.0)
Diastolic blood pressure (mmHg)	Mean (SD)	76.6 (10.0)	78.9 (11.1)	77.7 (10.6)	78.9 (15.9)	79.0 (11.3)	78.9 (14.1)	80.7 (13.2)	77.6 (11.4)	79.4 (12.5)
	Median (IQR)	76.0 (70.0–83.0)	78.0 (72.0–86.0)	77.0 (71.0–84.0)	78.0 (70.0–84.0)	80.0 (70.0–87.0)	78.0 (70.0–85.0)	78.0 (72.0–90.0)	78.0 (71.5–85.5)	78.0 (72.0–89.0)
PRISE total score	Mean (SD)	6.1 (3.0)	6.6 (3.4)	6.3 (3.2)	6.0 (3.3)	6.6 (3.6)	6.3 (3.4)	5.2 (3.1)	6.2 (4.2)	5.7 (3.6)
	Median (IQR)	6.0 (3.4–8.0)	6.5 (4.0–9.0)	6.0 (4.0–8.0)	6.0 (4.0–8.0)	6.0 (4.0–9.0)	6.0 (4.0–8.0)	5.0 (3.0–7.0)	6.0 (3.0–8.0)	5.0 (3.0–8.0)
		n (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)
Responder		16 (15.0)	10 (9.5)	26 (12.3)	24 (22.4)	15 (14.3)	39 (18.4)	25 (23.4)	12 (11.4)	37 (17.5)
Remission		11 (10.3)	6 (5.7)	17 (8.0)	13 (12.1)	7 (6.7)	20 (9.4)	12 (11.2)	9 (8.6)	21 (9.9)
Global improvement	Very much improved	5 (4.7)	4 (3.8)	9 (4.2)	6 (5.6)	0 (0.0)	6 (2.8)	5 (4.7)	4 (3.8)	9 (4.2)

continued

TABLE 8 Secondary outcome measure descriptive statistics by arm (*continued*)

Outcomes		Week 8			Week 26			Week 52		
		Quetiapine, *N* = 107	Lithium, *N* = 105	Overall, *N* = 212	Quetiapine, *N* = 107	Lithium, *N* = 105	Overall, *N* = 212	Quetiapine, *N* = 107	Lithium, *N* = 105	Overall, *N* = 212
	Much improved	11 (10.3)	9 (8.6)	20 (9.4)	17 (15.9)	18 (17.1)	35 (16.5)	23 (21.5)	16 (15.2)	39 (18.4)
	Minimally improved	34 (31.8)	23 (21.9)	57 (26.9)	26 (24.3)	19 (18.1)	45 (21.2)	21 (19.6)	13 (12.4)	34 (16.0)
	No change	24 (22.4)	30 (28.6)	54 (25.5)	20 (18.7)	14 (13.3)	34 (16.0)	16 (15.0)	13 (12.4)	29 (13.7)
	Minimally worse	8 (7.5)	7 (6.7)	15 (7.1)	2 (1.9)	4 (3.8)	6 (2.8)	2 (1.9)	1 (1.0)	3 (1.4)
	Much worse	1 (0.9)	0 (0.0)	1 (0.5)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (1.0)	1 (0.5)
	Missing	24 (22.4)	32 (30.5)	56 (26.4)	36 (33.6)	49 (46.7)	85 (40.1)	40 (37.4)	57 (54.3)	97 (45.8)

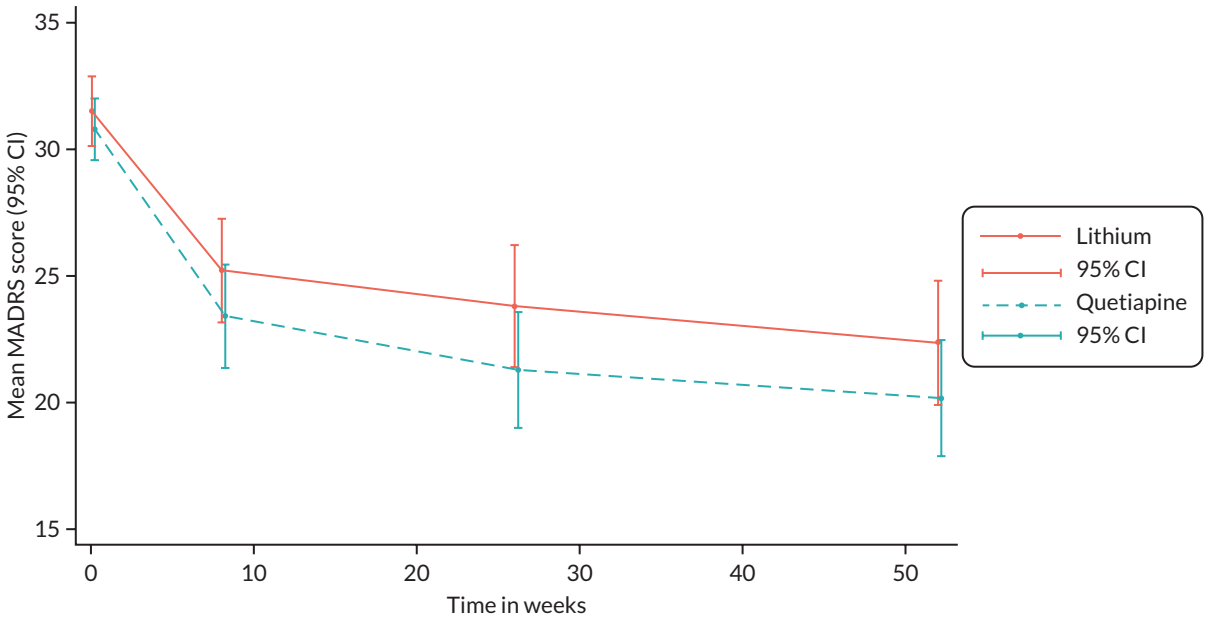

FIGURE 5 Mean MADRS scores at baseline and follow-up visits (unadjusted).

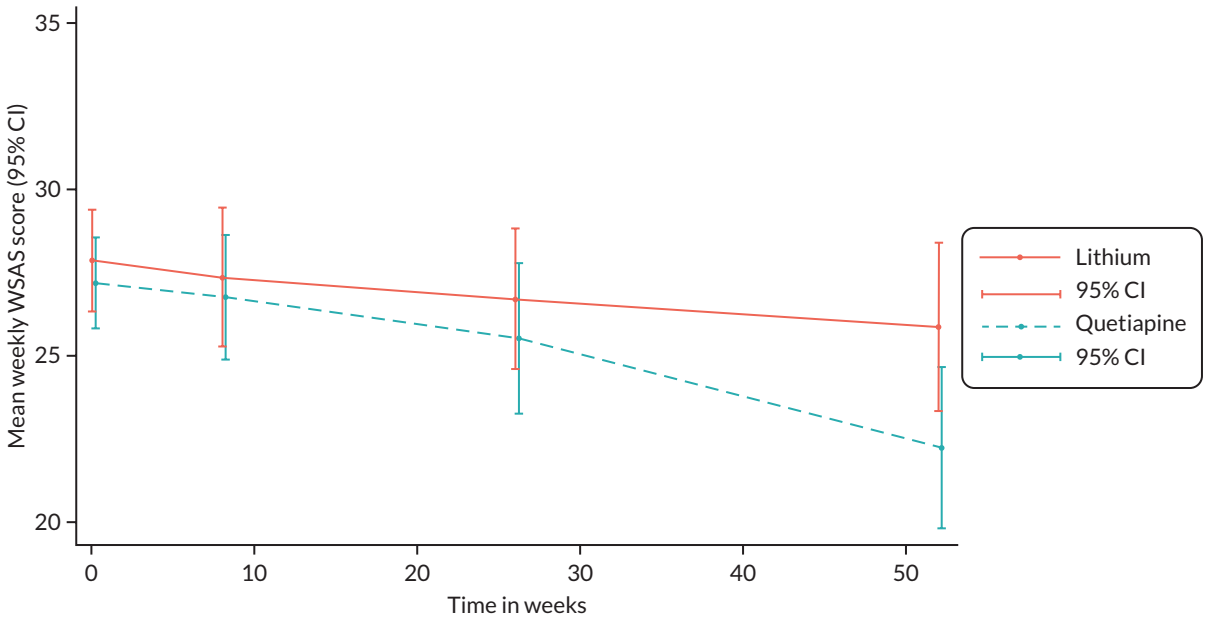

FIGURE 6 Mean WSAS scores at baseline and follow-up visits (unadjusted).

Figure 12 shows the weekly observed mean QIDS-SR scores by arm, against the predicted QIDS-SR scores as predicted by the fully adjusted ITT model (adjusted on baseline QIDS-SR score, TRD severity, depression severity, site, ethnicity and number of comorbidities).

As shown in *Table 10*, the median (IQR) days to trial medication discontinuation for participants in the quetiapine arm was 365.0 (57.0–365.0) which was longer than those in the lithium arm, 212.0 (21.0–365.0). Participants in the quetiapine arm had 0.72 times the hazard of discontinuing (95% CI: 0.47 to 1.09) compared to those in the lithium arm, which was not statistically significant (p = 0.1196), see *Table 11*.

The Kaplan–Meier plot in *Figure 13* as well as the Schoenfeld residual plot in *Figure 27*, *Appendix 4* shows a pattern to the residuals (a line with a non-zero slope), indicating the proportional-hazards assumption was violated. To address

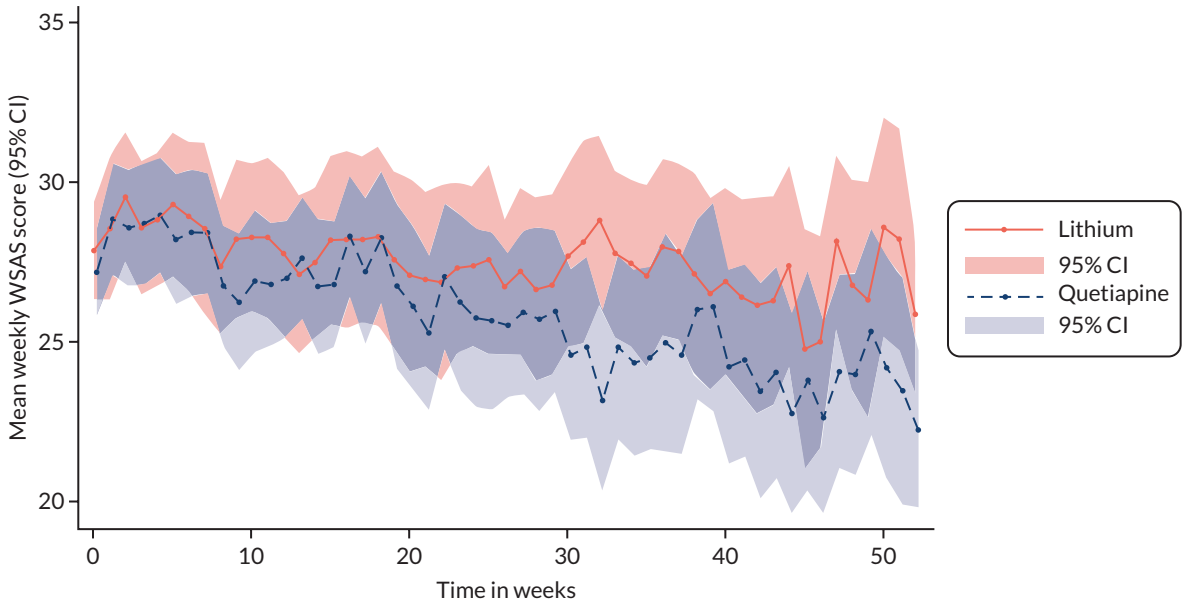

FIGURE 7 Mean weekly WSAS scores (unadjusted).

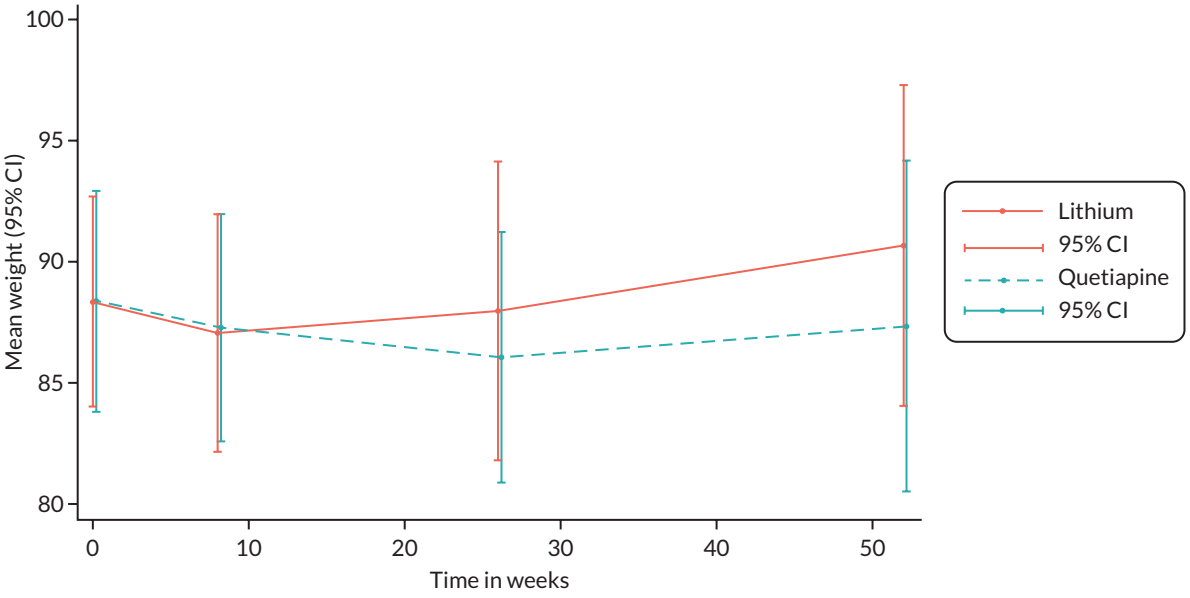

FIGURE 8 Mean weekly weight scores (unadjusted), including outliers.

this issue, the RMST difference is also presented. Over the 52-week study period, participants in the quetiapine arm remained on the trial medication for 21.70 more days on average than those in the lithium arm (95% CI: −16.03 to 59.44), which was also not a statistically significant difference (p = 0.2597).

Per-protocol analyses

Per-protocol analyses on the primary outcomes are reported in full in *Appendix 5* and summarised briefly here. For participants who were prescribed the trial medication, there was a significant quetiapine versus lithium area of −72.66 (95% CI: −137.03 to −8.29, p = 0.0269), similar to the ITT estimate. Median (IQR) days to discontinuation was 365.0 (100.0–365.0) in the quetiapine arm and 301.0 (140.0–365.0) in the lithium arm. The inferential Cox regression (p = 0.3866) and RMST results (p = 0.9068) showed no significant differences between the arms in time to discontinuation, consistent with the ITT result.

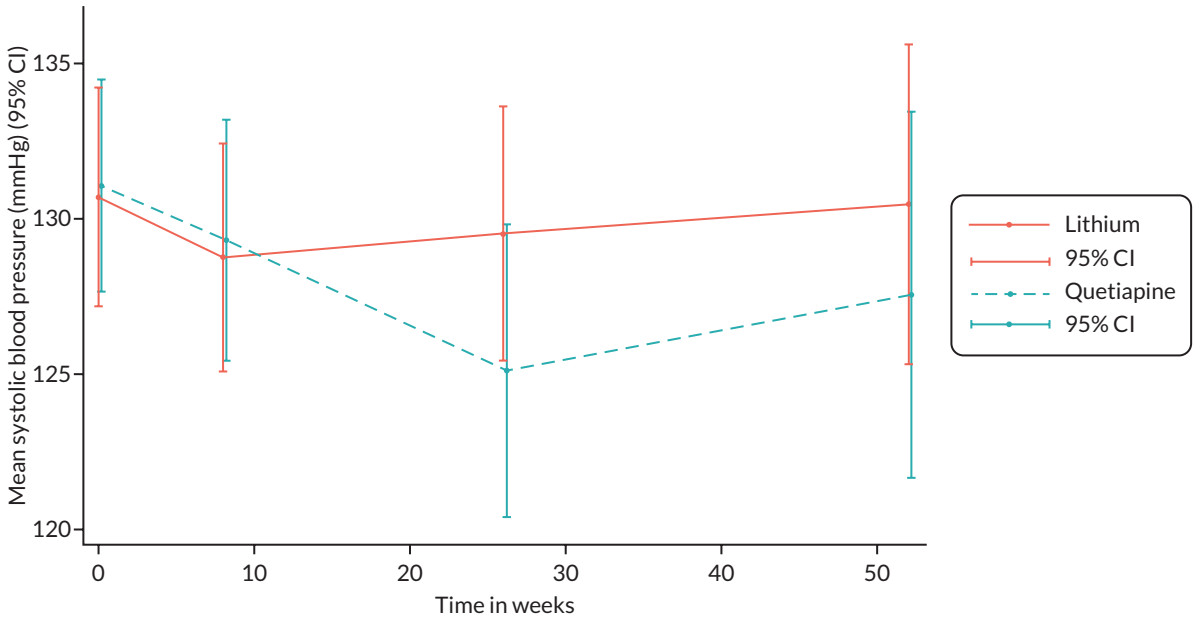

FIGURE 9 Mean systolic blood pressure (mmHg) (unadjusted).

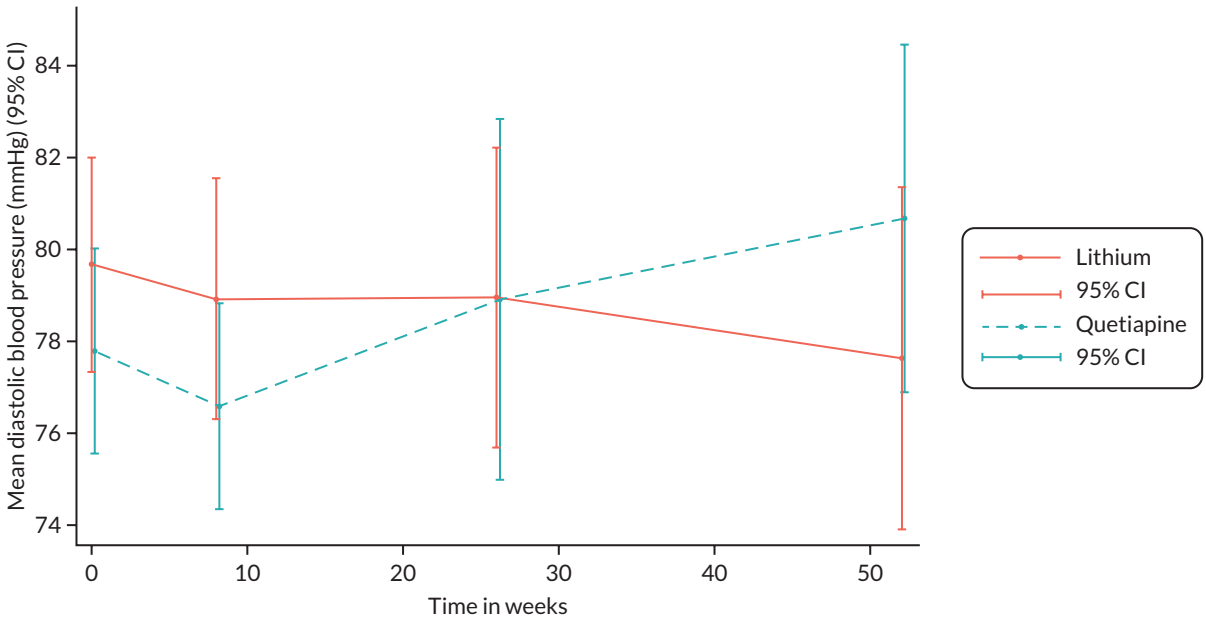

FIGURE 10 Mean diastolic blood pressure (mmHg) (unadjusted).

Sensitivity analyses

Sensitivity analyses on the primary outcomes are reported in full in *Appendix 6* and summarised briefly here.

Re-estimating effects for participants who received a therapeutic treatment trial resulted in a quetiapine versus lithium area under the difference curve of −83.11 points over 52 weeks (95% CI: −161.86 to −4.36), which was statistically significant (p = 0.0386), and similar to the ITT estimate. The median (IQR) days to discontinuation in the quetiapine arm in the therapeutic trial population was 365.0 (240.0–365.0) and in the lithium arm 365.0 (202.0–365.0). The inferential Cox regression (p = 0.3789) and RMST results (p = 0.7332) showed no significant differences between the arms in time to discontinuation, consistent with the ITT result.

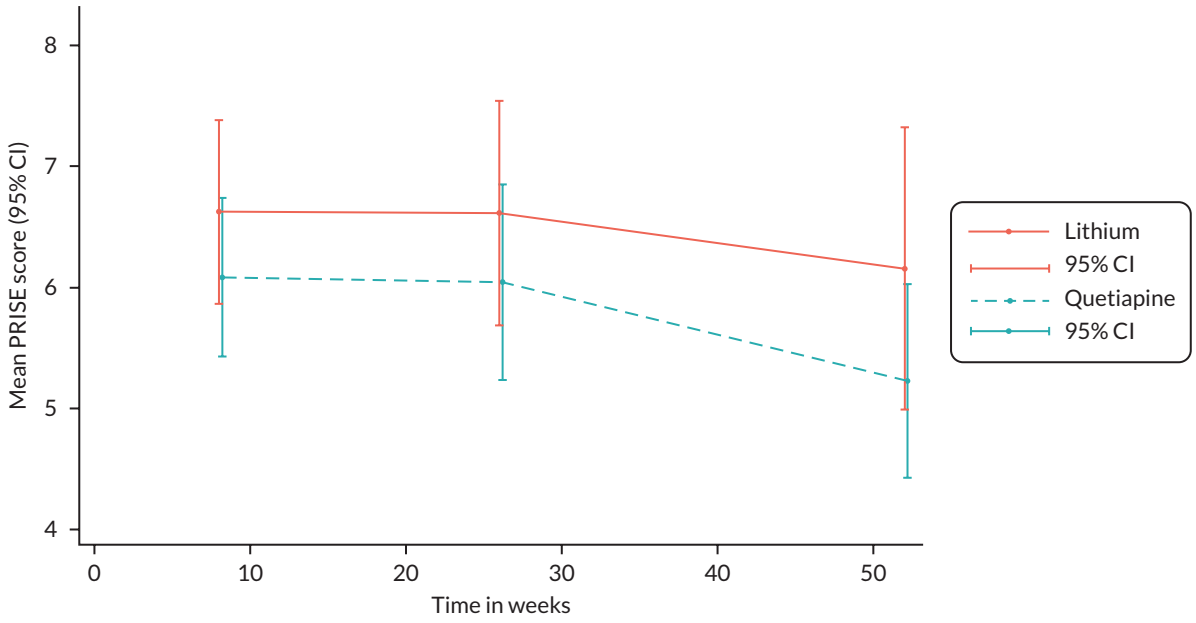

FIGURE 11 Mean PRISE scores (unadjusted).

TABLE 9 Between arm ITT area under the QIDS-SR difference curve estimates

	Quetiapine vs. lithium difference in AUC	95% CI	*p*-value
Weekly QIDS-SR adjusted by stratification factors[a]	−69.32	(−131.31 to −7.32)	0.0284
Weekly QIDS-SR adjusted by stratification factors[a] and missingness predictors[b]	−68.36	(−129.95 to −6.76)	0.0296

a Stratification factors (TRD severity, depression severity and site).
b Missing predictors number of comorbidities and ethnicity.

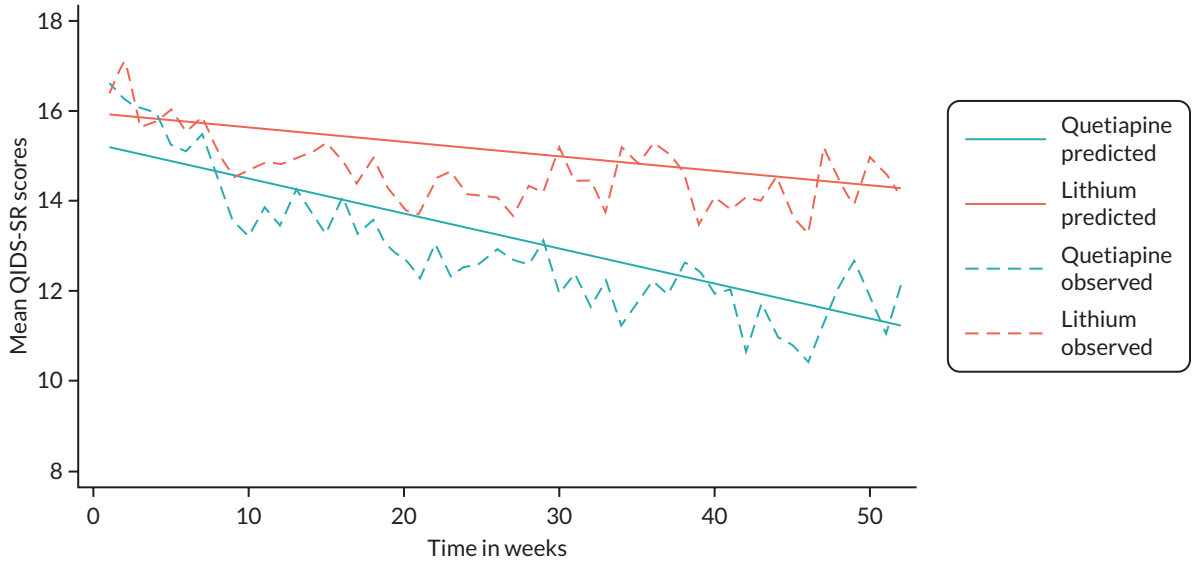

FIGURE 12 Weekly QIDS-SR scores, by arm, observed and model predicted (ITT).

TABLE 10 Time to discontinuation ITT descriptive statistics

		Quetiapine, *N* = 107	Lithium, *N* = 104
Time to discontinuation	Mean days (95% CI)	228.07 (197.77 to 258.36)	201.75 (171.80 to 231.71)
	Median days (IQR)	365.0 (57.0–365.0)	212.0 (21.0–365.0)

TABLE 11 Between arm time to discontinuation difference comparisons (ITT)

Time to discontinuation[a]	Hazard ratio (95% CI)	*p*-value	RMST difference (95% CI)	*p*-value
Quetiapine vs. lithium	0.72 (0.47 to 1.09)	0.1196	21.70 (−16.03 to 59.44)	0.2597
Lithium vs. quetiapine	1.39 (0.92 to 2.11)	0.1196		

a Adjusted by stratification factors (TRD severity, depression severity and site).

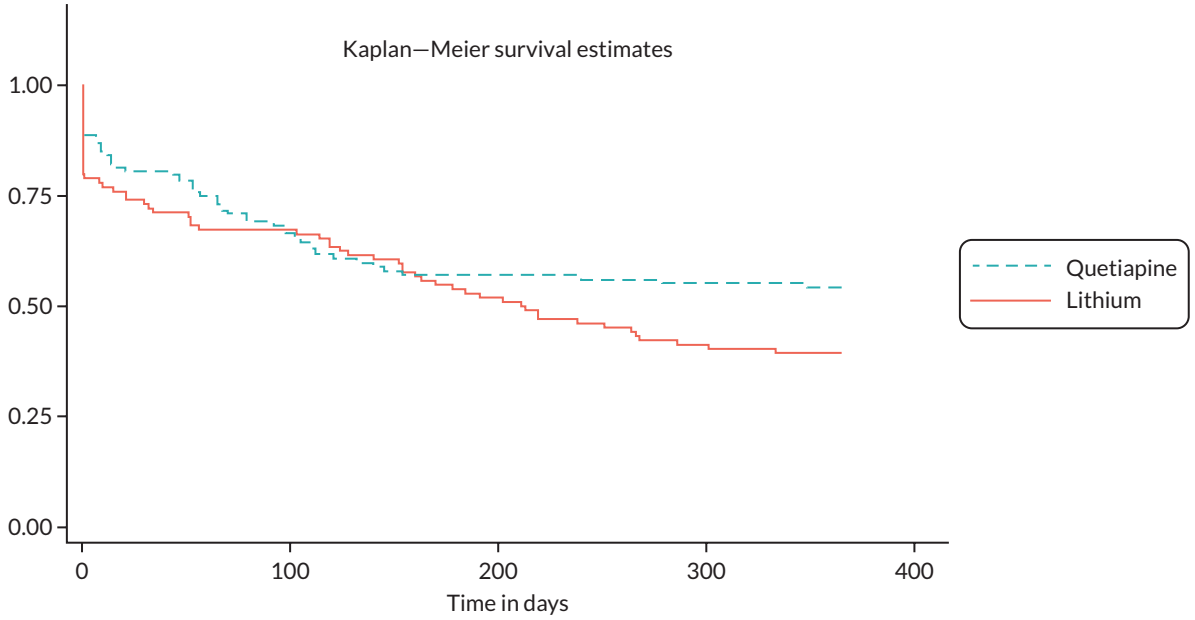

FIGURE 13 Time to discontinuation Kaplan–Meier curve (ITT) by arm.

Re-estimating effects for participants who were adherent to the trial medication resulted in quetiapine versus lithium area under the difference curve of −71.13 (95% CI: −145.18 to 2.91), which while not strictly statistically significant ($p = 0.0597$), was broadly consistent with the ITT estimate. Median (IQR) days to discontinuation in the quetiapine arm was 365.0 (154.0–365.0) and in the lithium arm 365.0 (166.5–365.0) for those who were adherent to the trial mediation. The inferential Cox regression ($p = 0.1384$) and RMST results ($p = 0.6227$) showed no significant differences between the arms in time to discontinuation, consistent with the ITT result.

Re-estimating effects for participants who initiated the trial medication resulted in a quetiapine versus lithium area under the difference curve of −72.51 (95% CI: −137.21 to −7.80, $p = 0.0281$), which was consistent with the ITT result. The median (IQR) days to discontinuation for participants in the quetiapine arm was 365.0 (100.0–365.0), compared to 333.0 (152.0–365.0) for those in the lithium arm in the population who initiated the trial medication. The inferential Cox regression ($p = 0.4655$) and RMST results ($p = 0.8793$) showed no significant differences between arms in time to discontinuation, consistent with the ITT result.

Exploring deviations from the MAR assumption, participants with data at week 52 in the quetiapine arm (n = 77, 72%) scored 1.79 points lower on the QIDS-SR than those in the lithium arm (n = 68, 65%) (95% CI: −3.57 to −0.02, p = 0.0473). This effect is robust to deviations from the MAR assumption applied in the main analysis: the effect becomes non-significant when worsening of 0.2 points or greater on the QIDS-SR is assumed, but only if this is assumed to occur solely in the quetiapine arm. The effect is robust to delta values across the range if these are assumed to apply to the lithium arm only, or to both arms.

Re-estimating effects on participants who were randomised before 1 February 2020 (onset of the COVID-19 pandemic) yielded a quetiapine versus lithium area under the difference curve of −73.07 (95% CI: −137.43 to −8.71), which was statistically significant (p = 0.0261), similar to the ITT result. Median (IQR) days to discontinuation for participants randomised before 1 February 2020 was 365.0 (66.0–365.0) in the quetiapine arm and 219.0 (25.5–365.0) in the lithium arm. The inferential Cox regression (p = 0.1413) and RMST results (p = 0.3153) showed no significant differences between arms in time to discontinuation, consistent with the ITT result. Similar results were found when effects were re-estimated excluding data collected after 1 March 2020 (onset of COVID-19 measures).

Secondary outcomes

Table 12, *Figures 14* and *15* show the results for secondary outcomes at week 8 and 52, presented as adjusted mean differences and 95% CIs. Regarding the MADRS, at 8 weeks participants in the quetiapine arm had a non-significant 1.50 points lower/better MADRS score on average than those in the lithium arm. At 52 weeks, the score in the quetiapine arm was 2.98 points lower, a statistically significant difference (p = 0.0435). On the WSAS, where higher scores indicate more functional impairment, there was little difference between arms at 8 weeks. The difference between arms was significant at 52 weeks, however, with those in the quetiapine arm scoring 3.64 points lower than those in the lithium arm (p = 0.0071). There were no significant differences between arms in self-reported medication adherence at either time point, as measured by the MARS-5.

TABLE 12 Between arm differences for continuous secondary outcomes

Secondary outcomes	Week	Quetiapine vs. lithium adjusted[a] mean difference (95% CI)	p-value
MADRS score	8	−1.50 (−4.06 to 1.07)	0.2521
	52	−2.98 (−5.87 to −0.09)	0.0435
WSAS score	8	−0.74 (−3.23 to 1.75)	0.5596
	52	−3.64 (−6.28 to −0.99)	0.0071
MARS-5 score	8	−0.03 (−0.55 to 0.48)	0.8967
	52	−0.44 (−1.04 to 0.17)	0.1563
Weight (kg)	8	0.04 (−2.29 to 2.37)	0.9746
	52	1.33 (−1.82 to 4.48)	0.4075
Weight (kg), with outliers removed	8	0.03 (−1.86 to 1.93)	0.9717
	52	0.83 (−1.50 to 3.17)	0.4839
Diastolic blood pressure (mmHg)	8	−0.58 (−4.00 to 2.85)	0.7415
	52	3.55 (−0.82 to 7.91)	0.1113
Systolic blood pressure (mmHg)	8	0.30 (−4.17 to 4.77)	0.8952
	52	−1.94 (−7.89 to 4.00)	0.5217
PRISE total score	8	−0.48 (−1.47 to 0.51)	0.3380
	52	−0.89 (−2.03 to 0.25)	0.1265

a Adjusted by stratification factors (TRD severity, depression severity and site).

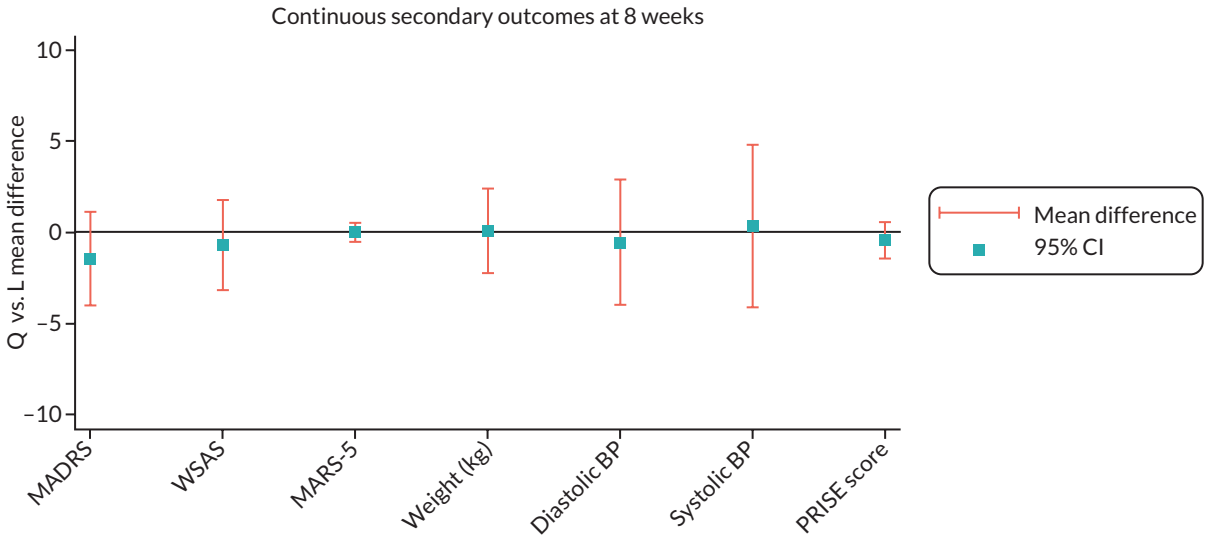

FIGURE 14 Adjusted between arm differences for continuous secondary outcomes at week 8. BP, blood pressure; L, lithium; Q, quetiapine.

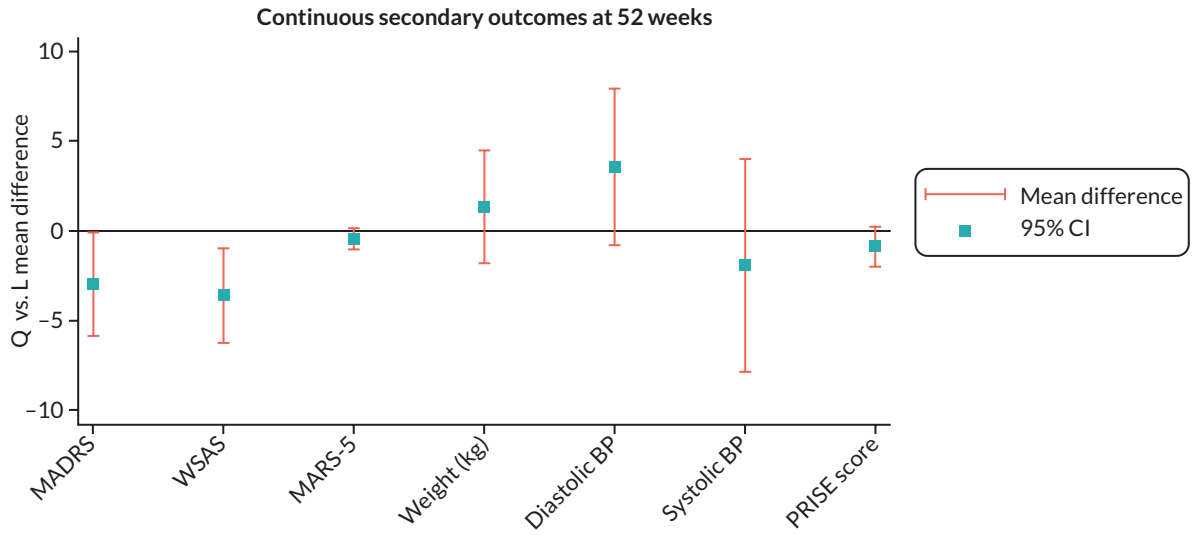

FIGURE 15 Adjusted between arm differences for continuous secondary outcomes at week 52. BP, blood pressure; L, lithium; Q, quetiapine.

Weight was analysed with the outliers included and excluded; please refer back to *Secondary outcomes* for the number of outliers in each arm, and see *Figures 45* and *46* in *Appendix 7* for the arm difference estimates at 8 and 52 weeks. There was little evidence of differences between the arms in weight at either time point in either of these analyses.

There appeared to be little difference in diastolic or systolic blood pressure between the two arms at either time point. The PRISE was used to assess the presence of side effects, a higher score indicates worse side effect. There was little evidence of a difference between the arms in self-reported side effects at either follow-up time point.

The results of responder, remitter and global improvement comparisons are shown in *Table 13* and *Appendix 7* (see *Figures 46–48*). At week 8, participants in the quetiapine arm had 1.95 times the odds (95% CI: 0.50 to 7.68) of responding compared to those in the lithium arm, which was not a statistically significant result. At 52 weeks, this was a larger difference of 3.67 times the odds (0.94–14.25), which was significant at the 10% but not 5% level.

TABLE 13 Between arm differences for binary secondary outcomes

Secondary outcomes	Week	Quetiapine vs. lithium, odds ratio (95% CI)[a]	Lithium vs. quetiapine odds ratio (95% CI)[a]	*p*-value
Responders – defined as a 50% reduction in the baseline MADRS total score	8	1.95 (0.50 to 7.68)	0.51 (0.13 to 2.01)	0.3370
	52	3.67 (0.94 to 14.25)	0.27 (0.07 to 1.06)	0.0607
Remission – defined as a MADRS total score of no > 10 points	8	2.22 (0.41 to 11.95)	0.45 (0.08 to 2.42)	0.3515
	52	1.38 (0.29 to 6.60)	0.72 (0.15 to 3.46)	0.6856
Global improvement – proportion of participants 'much' or 'very much' improved using CGI	8	1.23 (0.35 to 4.39)	0.81 (0.23 to 2.89)	0.7467
	52	1.12 (0.32 to 3.92)	0.89 (0.25 to 3.10)	0.8540

a adjusted by stratification factors (TRD severity, depression severity and site).

In terms of remission, participants in the quetiapine arm had 2.22 times the odds at 8 weeks (0.41–11.95), and 1.38 times the odds at 52 weeks (0.29–6.60) as compared to the lithium arm, with neither of these comparisons being statistically significant.

Finally, participants in the quetiapine arm had 1.23 times the odds (0.35–4.39) and 1.12 times the odds (0.32–3.92) of global improvement (much or very much improved) as compared to those in the lithium arm, again these comparisons were not statistically significant.

Time to initiation of the trial medication by arm is shown in *Table 14*. The median (IQR) days to initiation for participants in the quetiapine arm was 16.5 (6.0–33.0) which was similar to that in the lithium arm, 20.0 (8.0–34.0). Participants in the quetiapine arm had 1.09 times the hazard of initiating (95% CI: 0.74 to 1.62) as compared to those in the lithium arm, which was not statistically significant, see *Table 15*.

The Kaplan–Meier plot in *Figure 16* shows time to initiation between treatment arms over the 52-week study period. The Shoenfeld residuals plot (see *Figure 28*, *Appendix 4*) shows a pattern to the residuals/a line with a non-zero slope, indicating the proportional-hazards assumption was violated. Therefore, the RMST difference is also presented. Over the 52-week follow-up period, participants in the quetiapine arm initiated the trial medication 0.44 days quicker on average (95% CI: −7.86 to 6.98), which was not a statistically significant difference ($p = 0.9069$).

TABLE 14 Time to initiation of the trial medication descriptive statistics

		Quetiapine, *N* = 94	Lithium, *N* = 84
Time to initiation	Mean days (95% CI)	26.83 (19.44 to 34.22)	25.71 (20.33 to 31.10)
	Median days (IQR)	16.5 (6.0–33.0)	20.0 (8.0–34.0)

TABLE 15 Between arm comparisons of time to initiation of trial medication

Time to initiation[a]	Hazard ratio (95% CI)	*p*-value	RMST difference (95% CI)	*p*-value
Quetiapine vs. lithium	1.09 (0.74 to 1.62)	0.6502	−0.44 (−7.86 to 6.98)	0.9069
Lithium vs. quetiapine	0.91 (0.62 to 1.35)	0.6502		

a Adjusted by stratification factors (TRD severity, depression severity and site).

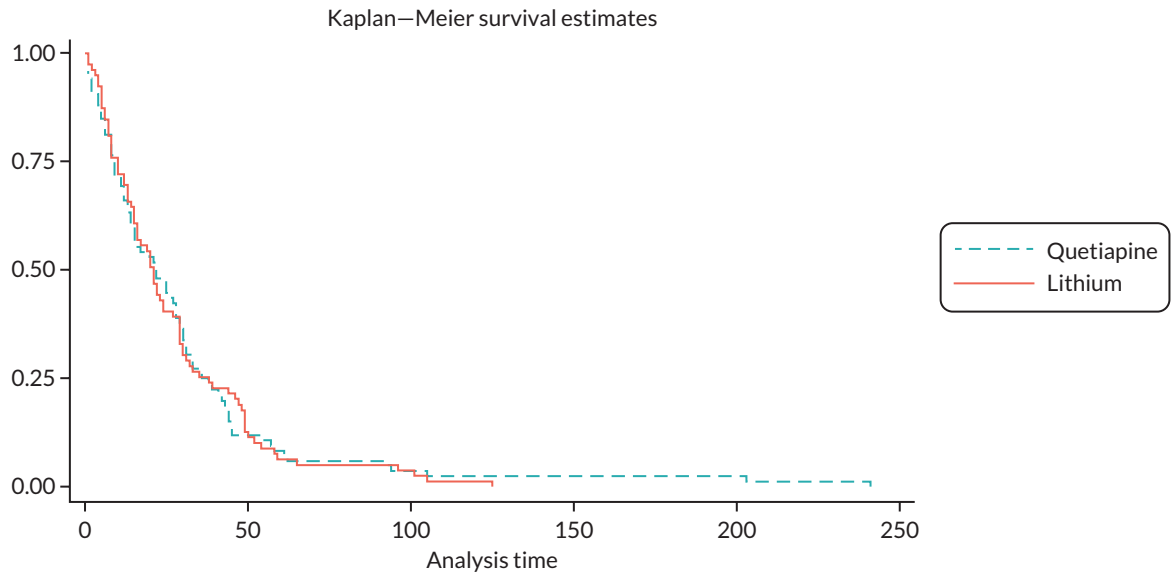

FIGURE 16 Time to initiation of trial medication Kaplan–Meier curve by arm.

The median (IQR) days to initiation of any new intervention for depression for participants in the quetiapine arm was 115.0 (48.5–163.0), which was similar to that for those in the lithium arm, 118.0 (46.0–195.0), see *Table 16*. The Kaplan–Meier plot in *Figure 17* shows time to initiation of any new intervention for depression by treatment arm over the 52-week study period. As shown in *Table 17*, participants in the quetiapine arm had 1.34 times the hazard of initiating a new treatment (95% CI: 0.73 to 2.44) as compared to those in the lithium arm, which was not a statistically significant difference. The Schoenfeld plot (see *Figure 29*, *Appendix 4*) shows a pattern to the residuals/a line with a non-zero slope, indicating the proportional-hazards assumption was violated. The RMST difference over the 52-week follow-up period showed that participants in the quetiapine arm initiated a new intervention for depression 6.24 days earlier on average (95% CI: −42.88 to 30.40), which was also not a statistically significant difference (p = 0.7384).

For all secondary outcome p-values, critical values were calculated using Hochberg's improved Bonferroni method.[46] The FDR was not prespecified, therefore FDRs of 5%, 10% and 20% were used. When the FDR was 5% or 10%, there was no instance where the p-value was lower than its Benjamini–Hochberg critical value, that is, no between arm secondary outcome p-values remained significant. However, when the FDR was set at 20%, the between arm WSAS comparison at 52 weeks remained significant (p = 0.0071 compared to critical value p = 0.0077).

TABLE 16 Time to initiation of any new intervention for depression descriptive statistics

		Quetiapine, *N* = 40	Lithium, *N* = 39
Time to initiation of any new intervention for depression	mean days (95% CI)	118.78 (92.06 to 145.49)	126.92 (96.77 to 157.08)
	median days (IQR)	115.0 (48.5–163.0)	118.0 (46.0–195.0)

TABLE 17 Between arm time to initiation of any new intervention for depression comparisons

Time to initiation of any new intervention for depression[a]	Hazard ratio (95% CI)	*p*-value	RMST difference (95% CI)	*p*-value
Quetiapine vs. lithium	1.34 (0.73 to 2.44)	0.3409	−6.24 (−42.88 to 30.40)	0.7384
Lithium vs. quetiapine	0.75 (0.41 to 1.36)	0.3409		

a Adjusted by stratification factors (TRD severity, depression severity and site).

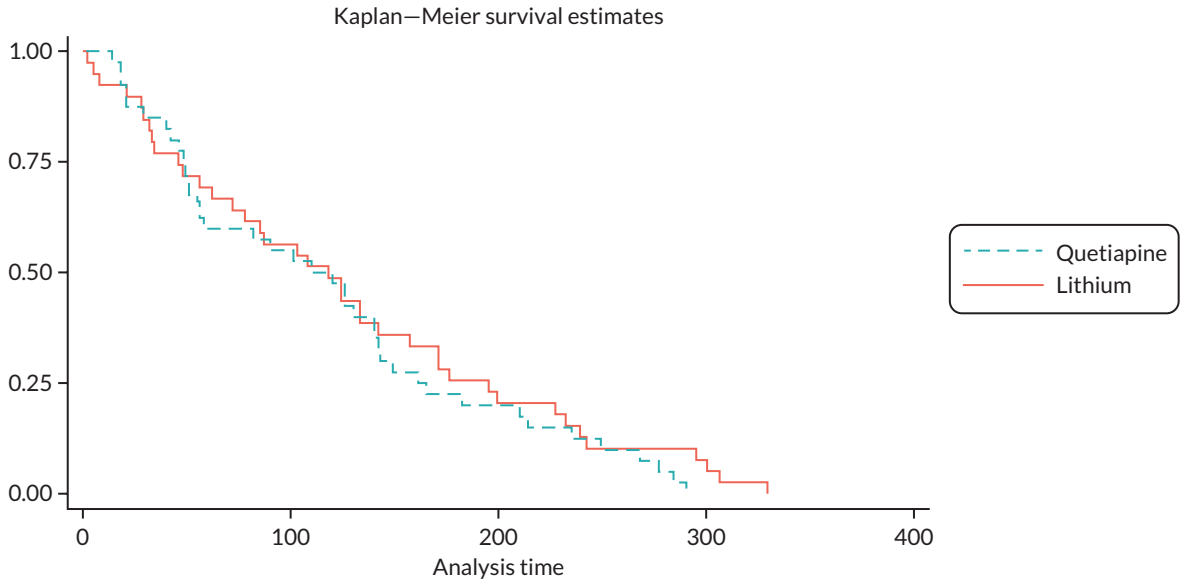

FIGURE 17 Time to initiation of any new intervention for depression Kaplan–Meier curve by arm.

Chapter 4 Health economic results

Use and cost of the intervention

Of 105 participants randomised to lithium, 86 (81.9%) were prescribed the medication and 84 (77.8%) started the medication. Eighty-five (81.0%) received lithium carbonate while one participant (1.0%) received lithium citrate. Of 107 participants randomised to quetiapine, 95 (88.8%) were prescribed quetiapine and subsequently started taking the medication. Fifty-nine participants (55.1%) received quetiapine XR while 53 participants (32.7%) received quetiapine IR. Costs of the most prescribed formulations (branded and the generic equivalents) are summarised in *Table 18*. Generic and cheapest branded options were similar, apart from quetiapine XR, where the cheapest branded option (Sondate XL 50 mg tablets) was much cheaper than the generic equivalent.

Service use

The number (%) of participants with CSRI service-use data at baseline and each follow-up were: 102 (97.1%) in the lithium arm and 104 (97.2%) in the quetiapine arm at baseline, 91 (86.7%) in the lithium arm and 93 (86.9%) in the quetiapine arm at week 8, 69 (65.7%) in the lithium arm and 81 (75.7%) in the quetiapine arm at week 26, and 63 (60.0%) in the lithium arm and 73 (68.2%) in the at week 52.

Table 19 summarises service use by treatment arm before the baseline assessment and over the 52-week follow-up period. In the 3 months before the baseline assessment, the most frequently used type of service for both arms were day care services (e.g. drop-in centres, acute day hospital services, one-to-one support, and group support provided outside of hospital), used by 73.3% of participants in the lithium arm and 74.8% of those in the quetiapine arm, with a mean number of contacts of 57.27 and 56.66, respectively. GP appointments and outpatient services were also frequently used in both arms. Although not tested for statistical significance, the lithium arm appeared to have a higher number of admissions to A&E and a higher number of contacts with community nurses and dentists; while the quetiapine arm had a higher number of contacts with counsellors, physiotherapists, mental health professionals and other doctors. There were few noticeable differences between the two arms at baseline for other type of services, including productivity losses for health reasons.

During the 52-week follow-up, day care services remained the most frequently used service for both arms, followed by GP appointments and outpatient services. Those in the lithium arm had more admissions to A&E, and a higher number of contacts with community nurses, district nurses and other professionals, while the quetiapine arm had a higher number of contacts with physiotherapists and community psychologists. Patients in the lithium arm lost 9.55 working days while patients in the quetiapine arm lost 16.13 working days.

Total cost by treatment arm

See *Table 25*, *Appendix 1* for unit costs. Baseline healthcare and societal cost by treatment arm is reported in *Table 20*. The quetiapine arm had a higher baseline cost of NHS Community-based services by £319.70 (95% CI: £17.56 to 621.85) compared to the lithium arm. However, no significant difference was observed in the total NHS cost or societal cost between arms.

Total healthcare and societal costs per participant over the 52-week follow-up period are reported in *Table 21*. The quetiapine arm had a lower healthcare cost (−£472.32, 95% CI: −£1111.12 to £166.47) and a higher societal cost (£162.90, 95% CI: −£1224.13 to 1549.94) compared to the lithium arm, with probabilities of the quetiapine arm being cost saving of 0.94 and 0.45, respectively.

TABLE 18 Medication costs according to the NHS open data portal

Medication type	Most prescribed formulation	Generic cost, £	Cheapest branded cost, £
Lithium carbonate	Priadel 400 mg	8.50	8.50
Lithium citrate	Priadel 520 mg/5 ml	6.73	6.73
Quetiapine (XR)	Quetiapine 50 mg modified release tablet	67.66	11.99
Quetiapine (IR)	Quetiapine 25 mg tablet	1.15	1.15

EuroQol-5 Dimensions scores and quality-adjusted life-years by treatment arm

EuroQol-5 Dimensions scores at baseline and each follow-up are detailed in *Table 22* and *Figure 18*. At baseline, mean EQ-5D utility scores were similar across arms. From the 8-week follow-up, mean utility scores in the quetiapine arm became significantly higher than the lithium arm; however, the difference became non-significant at 52 weeks. Mean QALY gain between baseline and 52-week follow-up was 0.540 for the quetiapine arm and 0.468 for the lithium arm. The adjusted difference was 0.074 in favour of quetiapine with a 99.5% chance that quetiapine leads to improved QALY.

Cost-effectiveness of the interventions

Primary analysis

The base case cost-effectiveness results from the NHS and PSS perspective are reported in *Table 23*. Quetiapine was associated with a lower cost and a higher QALY gain, and therefore dominates lithium. The INB of quetiapine was 0.094 QALYs per patient.

The cost-effectiveness plane showing NHS and PSS cost and QALY differences at 52-week follow-up is shown in *Figure 19*. The most likely scenario with 92.2% of the re-samples was that quetiapine arm resulted in lower costs and a high QALY gain compared with lithium.

The CEAC for the 52-week follow-up period from a healthcare perspective is shown in *Figure 20*. If the NHS was not willing to attach any monetary value to QALY gains, the likelihood for quetiapine to become the most cost-effective option is 0.92. As a QALY is valued at higher levels, this likelihood increases.

Secondary analyses

Adopting a societal perspective

The cost-effectiveness results from a societal perspective are reported in *Table 24*. Compared with lithium, quetiapine was associated with a higher cost and a higher QALY. Since the ICER of quetiapine (£4,657.78 per QALY) was lower than the NICE WTP threshold of £20,000 per additional unit of cost, quetiapine was deemed to be cost-effective. The INB of quetiapine was 0.055 QALYs per patient.

The cost-effectiveness plane of the societal cost and QALY differences at 52-week follow-up is shown in *Figure 21*. 57.1% of the re-samples showed higher costs and a higher QALY gain for the quetiapine arm while 42.4% showed lower costs and a greater QALY gain. Only 0.4% showed higher costs and a lower QALY gain for quetiapine arm and 0.1% showed lower costs and a lower QALY gain.

The CEAC for the 52-week follow-up period from a societal perspective is shown in *Figure 22*. The probability that quetiapine was the cost-effective option is 0.43 at a WTP threshold of £0 per QALY. At the NICE WTP threshold of £20,000–30,000, the likelihood for quetiapine to become the most cost-effective option increased to 0.91.

TABLE 19 Number (%) of patients using healthcare services and mean (SD) number of contacts before baseline assessment and during the 52-week follow-up period

	3-month to baseline				52-week follow-up period			
	Lithium (*n* = 105)		Quetiapine (*n* = 107)		Lithium (*n* = 105)		Quetiapine (*n* = 107)	
Service	*N* (%)	Mean (SD) contacts[a]	*N* (%)	Mean (SD) contacts[a]	*N* (%)	Mean (SD) contacts[a]	*N* (%)	Mean (SD) contacts[a]
Hospital services								
Mental health admission	5 (4.76)	0.05 (0.22)	0 (0)	0 (0.00)	5 (4.8)	0.11 (0.65)	0 (0.0)	0.00 (0.00)
Non-mental health admission	5 (4.76)	0.05 (0.22)	2 (1.9)	0.05 (0.35)	4 (3.8)	0.45 (3.92)	6 (5.6)	0.12 (0.67)
Outpatient	26 (24.8)	1.09 (4.12)	34 (31.8)	1.16 (3.12)	53 (50.5)	3.35 (6.03)	49 (45.8)	2.78 (5.05)
A&E	15 (14.3)	0.32 (0.89)	10 (9.3)	0.16 (0.53)	26 (24.8)	0.77 (1.91)	19 (17.8)	0.31 (0.74)
Day case	4 (3.8)	0.07 (0.39)	4 (3.7)	0.05 (0.26)	6 (5.7)	0.12 (0.50)	10 (9.3)	0.22 (0.83)
NHS Community-based services								
GP	75 (71.4)	2.21 (2.55)	82 (76.6)	2.6 (2.85)	79 (75.2)	4.62 (6.15)	89 (83.2)	5.27 (6.28)
Practice nurse	17 (16.2)	0.53 (2.00)	16 (14.9)	0.3 (0.78)	32 (30.5)	1.12 (2.53)	34 (31.8)	0.81 (1.98)
District nurse	2 (1.9)	0.12 (1.03)	2 (1.9)	0.02 (0.14)	5 (4.8)	0.38 (2.21)	4 (3.7)	0.07 (0.44)
Community nurse	24 (22.9)	0.87 (3.25)	9 (8.4)	0.14 (0.49)	24 (22.9)	3.53 (13.62)	11 (10.3)	0.59 (3.30)
Occupational therapist	2 (1.9)	0.08 (0.64)	4 (3.7)	0.16 (1.14)	3 (2.9)	0.07 (0.47)	4 (3.7)	0.12 (0.74)
Community psychiatrist	29 (27.6)	0.41 (0.7)	16 (14.9)	0.39 (1.25)	41 (39.0)	1.97 (3.74)	43 (40.2)	2.02 (4.03)
Community psychologist	11 (10.5)	0.51 (1.86)	9 (8.4)	0.42 (1.92)	10 (9.5)	0.61 (3.39)	20 (18.7)	1.71 (5.99)
Dentist	20 (19.1)	0.27 (0.6)	11 (10.3)	0.15 (0.46)	34 (32.4)	0.65 (1.10)	28 (26.2)	0.71 (1.94)
Optician	7 (6.7)	0.07 (0.26)	6 (5.6)	0.08 (0.34)	19 (18.1)	0.29 (0.65)	14 (13.1)	0.21 (0.59)
Counsellor	8 (7.6)	0.47 (1.84)	12 (11.2)	0.86 (2.76)	6 (5.7)	0.89 (4.21)	10 (9.3)	0.93 (3.67)
Physiotherapist	2 (1.9)	0.07 (0.61)	5 (4.7)	0.24 (1.18)	7 (6.7)	0.40 (2.04)	12 (11.2)	0.56 (1.95)
Other doctors	9 (8.6)	0.16 (0.65)	11 (10.3)	0.8 (6.03)	11 (10.5)	0.47 (2.73)	12 (11.2)	0.32 (1.07)
Alternative therapist	0 (0)	0 (0.0)	1 (0.9)	0.06 (0.6)	1 (1.0)	0.06 (0.62)	2 (1.9)	0.10 (0.84)

TABLE 19 Number (%) of patients using healthcare services and mean (SD) number of contacts before baseline assessment and during the 52-week follow-up period *(continued)*

	3-month to baseline				52-week follow-up period			
	Lithium (*n* = 105)		Quetiapine (*n* = 107)		Lithium (*n* = 105)		Quetiapine (*n* = 107)	
Service	***N* (%)**	**Mean (SD) contacts[a]**	***N* (%)**	**Mean (SD) contacts[a]**	***N* (%)**	**Mean (SD) contacts[a]**	***N* (%)**	**Mean (SD) contacts[a]**
Mental health professional	3 (2.9)	0.06 (0.37)	4 (3.7)	0.21 (1.12)	6 (5.7)	0.34 (2.09)	5 (4.7)	0.21 (1.32)
Other professionals	7 (6.7)	0.42 (1.9)	13 (12.1)	0.49 (1.71)	24 (22.9)	2.86 (11.93)	21 (19.6)	1.24 (4.77)
NHS line	2 (1.9)	0.02 (0.14)	3 (2.8)	0.05 (0.29)	2 (1.9)	0.06 (0.52)	1 (0.9)	0.01 (0.10)
Non-NHS Community-based services								
Social worker	7 (6.7)	0.13 (0.57)	1 (0.9)	0.01 (0.1)	3 (2.9)	0.19 (1.65)	3 (2.8)	0.03 (0.17)
Day care	77 (73.3)	56.66 (36.39)	80 (74.8)	57.27 (36)	94 (89.5)	161.62 (69.30)	98 (91.6)	173.81 (76.49)
Informal care	4 (3.8)	5.14[b] (31.67)	1 (0.93)	4.44[b] (44.4)	2 (1.9)	4.65[b] (37.91)	0 (0.0)	0.00[b] (0.00)
Employment								
Loss of productivity	48 (46)	7.80[c] (19.38)	52 (48.6)	8.02[c] (19.32)	45 (42.9)	9.55[c] (23.94)	54 (50.5)	16.13[c] (41.88)
Medication								
Concomitant medication[d]	–	–	–	–	105 (100)	3.98 (3.31)	106 (99)	4 (4)

a Number of contacts is for the whole sample.
b Number of hours per week carer took off work to care for the participant.
c Number of days off work for health reasons in the last 3 months.
d Not assessed in period prior to baseline. Covers medications for the following systems: Cardiovascular, Hepatic, Endocrine, Neurological, Psychological, and Dermatological.

TABLE 20 Total costs (£) per participant at baseline

Cost item	Treatment arm, mean cost (SD)		Difference[a]	95% CI
	Lithium (*n* = 105)	Quetiapine (*n* = 107)		
Hospital services	610.72 (2810.00)	281.14 (821.25)	−460.48	−992.94 to 71.96
NHS Community-based services	338.08 (303.45)	878.92 (3619.81)	319.70	17.56 to 621.85
Non-NHS Community-based services	146.66 (790.85)	129.85 (1114.54)	−25.99	−273.77 to 221.79
Loss of productivity	940.19 (2360.62)	971.05 (2353.10)	−157.57	−867.15 to 551.99
Total NHS[b]	948.81 (2893.06)	1160.06 (3714.42)	56.93	−534.69 to 648.55
Total societal[c]	2035.66 (3714.68)	2260.96 (4423.71)	−150.22	−1153.17 to 852.72

a Using a generalised linear model with gamma family and log link and adjusted for arm, EQ-5D baseline index value, QIDS baseline score, site, failure of antidepressant treatment and depression severity.
b Total NHS cost consists of hospital cost plus NHS Community costs.
c Total societal consists of total NHS cost plus costs of lost productivity and non-NHS Community-based services.

TABLE 21 Total costs (£) per participant over the 52-week follow-up period

Cost item	Treatment arm, mean cost (SD)		Adjusted difference[a]	95% CI	Probability quetiapine is cost saving (%)
	Lithium (*n* = 95)	Quetiapine (*n* = 99)			
Trial drug[b]	211.21 (1379.57)	239.29 (158.28)	215.78	−20.37 to 451.93	–
Concomitant medication	426.78 (468.75)	388.25 (371.08)	−28.34	−156.9 to 100.21	–
Hospital services	905.94 (1478.91)	780.80 (1531.49)	−11.40	−646.48 to 623.67	–
NHS Community-based services	958.24 (1425.69)	987.72 (1070.10)	84.53	−271.79 to 440.87	–
Non-NHS Community-based services	162.07 (1032.65)	166.21 (723.76)	−8.02	−265.95 to 249.9	–
Loss of productivity	1181.39 (2953.85)	1990.60 (5168.56)	902.67	−385.62 to 2190.97	–
Total NHS cost[c]	3151.46 (2556.76)	2706.77 (2277.89)	−472.32	−1111.12 to 166.47[d]	94.1
Total Societal cost[c]	4366.97 (4140.29)	4702.33 (5849.77)	162.90	−1224.13 to 1549.94[d]	44.7

a Using a generalised linear model with gamma family and log link and adjusted for by arm, EQ-5D baseline index value, baseline cost of the same item, QIDS baseline score, site and failure of antidepressant treatment.
b Using the cheapest unit cost of the trial drug.
c Missing values in the total costs are multiple imputed assuming MAR.
d Bias corrected using bootstrap regression.

Using Self-rated Quick Inventory of Depressive Symptomatology scores as the effectiveness outcome

The cost-effectiveness plane showing NHS and PSS cost and QIDS-SR differences at 52-week follow-up is shown in *Figure 23*. Overall, 72.8% of the re-samples showed lower costs and an improvement in QIDS-SR scores for the quetiapine arm while 20.0% showed lower costs and a worsening in QIDS-SR score. Six per cent showed higher costs and an improvement in QIDS-SR score for quetiapine arm and only 1.2% showed higher costs and a worsening in QIDS-SR score.

TABLE 22 EuroQol-5 Dimensions utility scores and QALYs by treatment arm over the 52-week follow-up period

	Treatment arm						
	Lithium		Quetiapine				
	n	Mean score (SD)	*n*	Mean score (SD)	Difference	95% CI	Probability quetiapine improving QALY (%)
Baseline	101	0.381 (0.310)	106	0.407 (0.313)	0.001	−0.075 to 0.076	–
8 weeks	87	0.456 (0.324)	94	0.530 (0.288)	0.075	0.003 to 0.146	–
26 weeks	68	0.480 (0.338)	86	0.545 (0.315)	0.088	0.011 to 0.165	–
52 weeks	63	0.477 (0.340)	72	0.553 (0.346)	0.072	−0.009 to 0.155	–
All-period QALY	46	0.468 (0.275)	63	0.540 (0.281)	0.074[a]	0.018 to 0.131[b]	99.5

a Missing values are multiple imputed and the estimated difference is adjusted for EQ-5D baseline index value, QIDS baseline score, site, and failure of antidepressant treatment.
b Bias corrected using bootstrap regression.

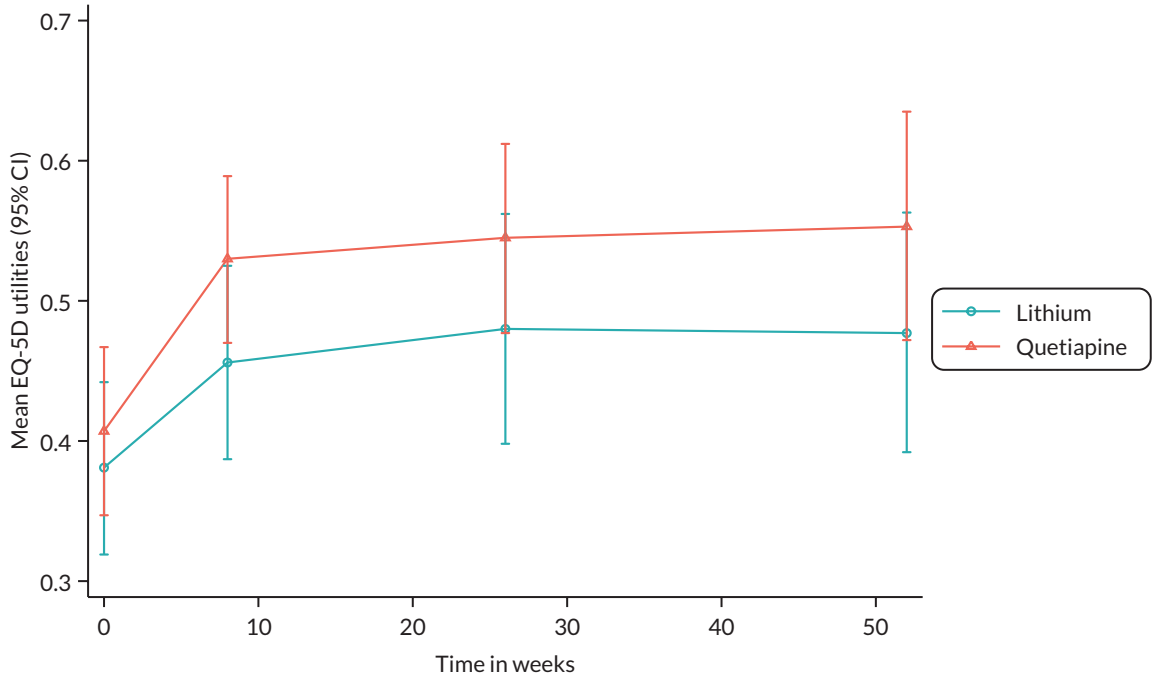

FIGURE 18 Mean EQ-5D utility score (unadjusted) with 95% CI.

TABLE 23 Base case cost-effectiveness results from the NHS and PSS perspective

Intervention	NHS and PSS cost (£)	QALY	Incremental cost[a]	Incremental QALY[b]	ICER	INB
Lithium	3151.46	0.468	–	–	Dominated	–
Quetiapine	2706.77	0.540	−472.32	0.074	Dominating	0.097

a Missing value are multiple imputed and the estimated difference is adjusted for arm, EQ-5D baseline index value, baseline cost of the same item, QIDS baseline score, site and failure of antidepressant treatment.
b Missing values are multiple imputed and the estimated difference is adjusted for EQ-5D baseline index value, QIDS baseline score, site and failure of antidepressant treatment.

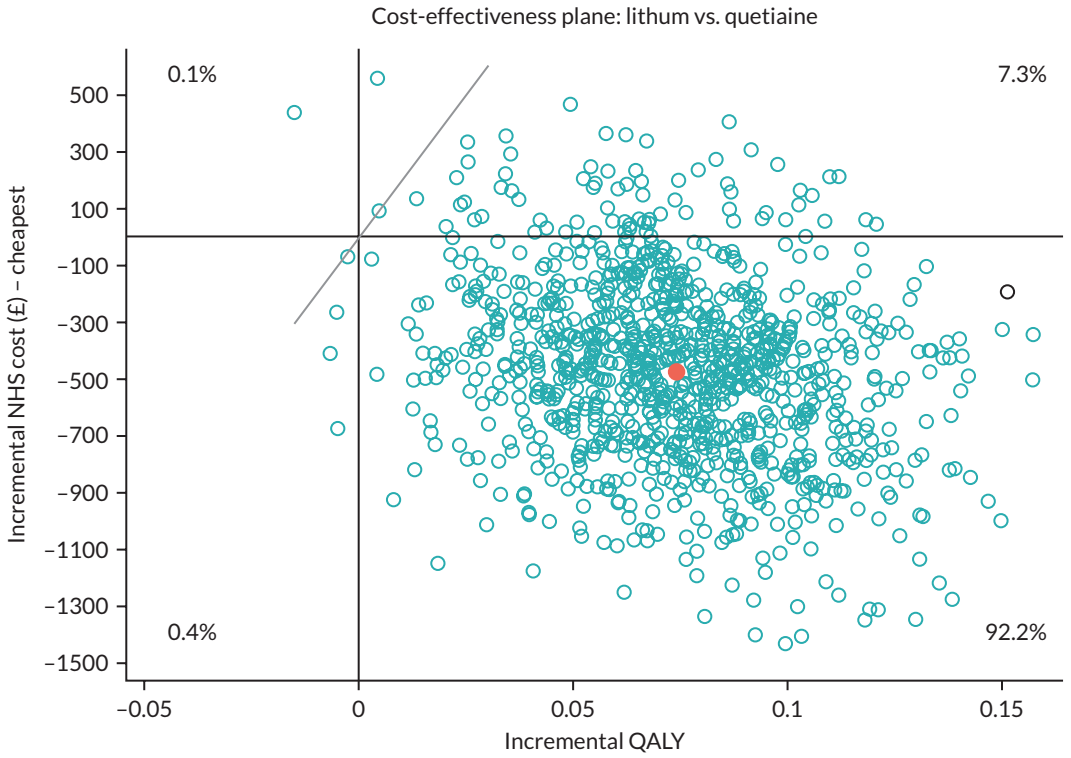

FIGURE 19 Cost-effectiveness plane of NHS and PSS cost and QALY differences.

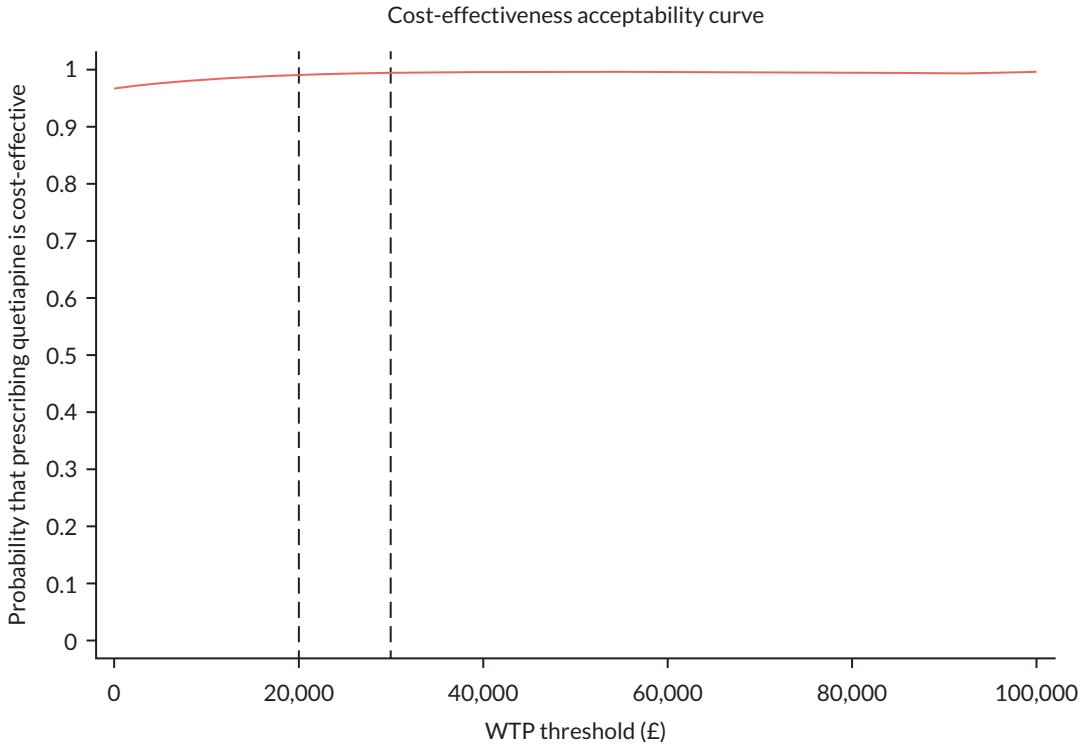

FIGURE 20 Cost-effectiveness acceptability curve showing the probability that quetiapine is the cost-effective option at different WTP threshold for improvement in QALYs – NHS and PSS perspective.

TABLE 24 Base case cost-effectiveness results from a societal perspective

Intervention	Societal cost (£)	QALY	Incremental cost	Incremental QALY	ICER	INB
Lithium	4366.97	0.468	–	–	–	–
Quetiapine	4702.33	0.540	335.36	0.072	4657.78	0.055

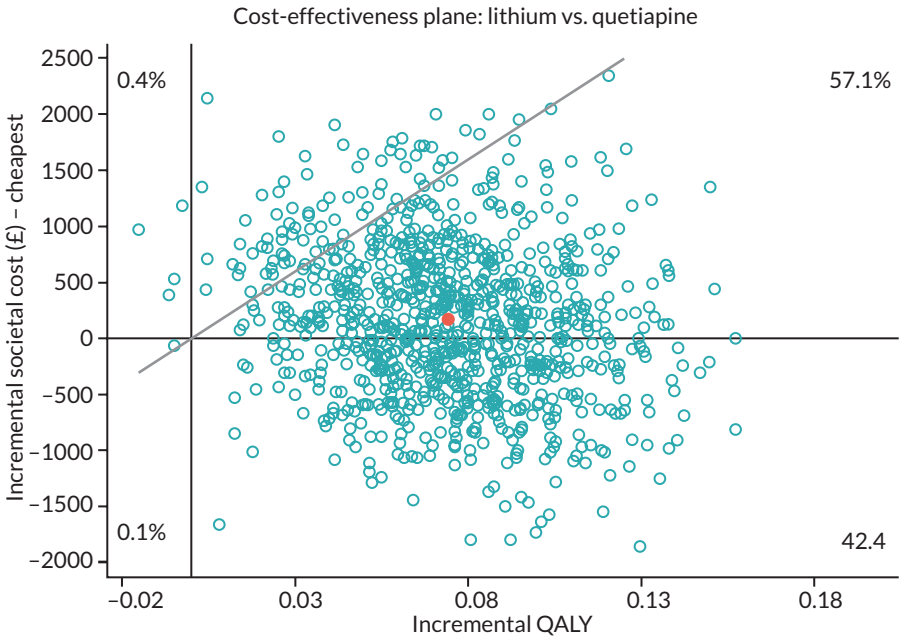

FIGURE 21 Cost-effectiveness plane of societal cost and QALY differences.

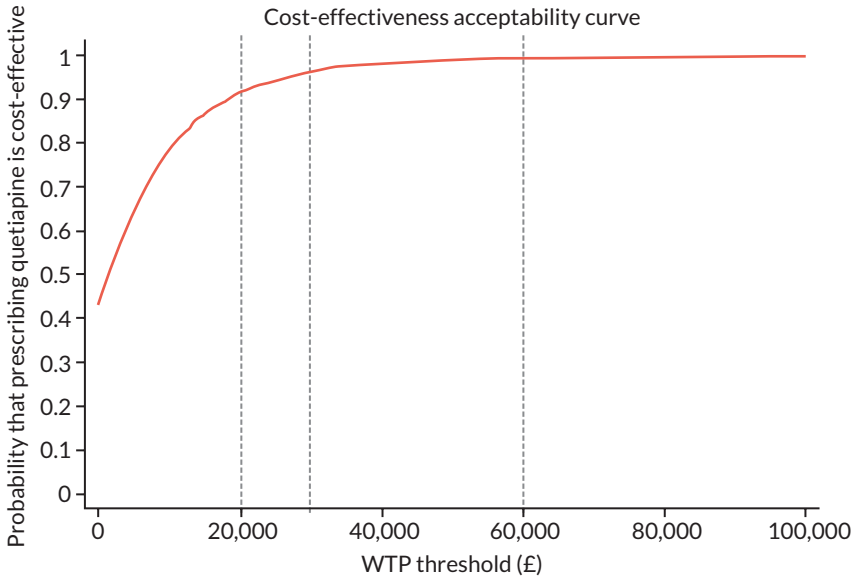

FIGURE 22 Cost-effectiveness acceptability curves showing the probability that quetiapine was the most cost-effective option at different WTP threshold for improvement in QALYs – societal perspective.

The CEAC for the 52-week follow-up period from the NHS and PSS perspective is shown in *Figure 24*. The probability that quetiapine was the cost-effective option is 0.92 at a WTP threshold of £0 per unit improvement in QIDS-SR score. As a unit improvement is valued at higher levels, the probability relatively decreases, as lithium becomes more cost-effective for the iterations where quetiapine is less effective and costing less (lithium more effective but costing more), but stays around 0.8 in favour of quetiapine.

Adopting a societal perspective with Self-rated Quick Inventory of Depressive Symptomatology scores as the effectiveness outcome

The cost-effectiveness plane showing societal cost and QIDS-SR differences at 52-week follow-up is shown in *Figure 25*. In total, 45.8% of the re-samples showed higher costs and an improvement in QIDS-SR score for the

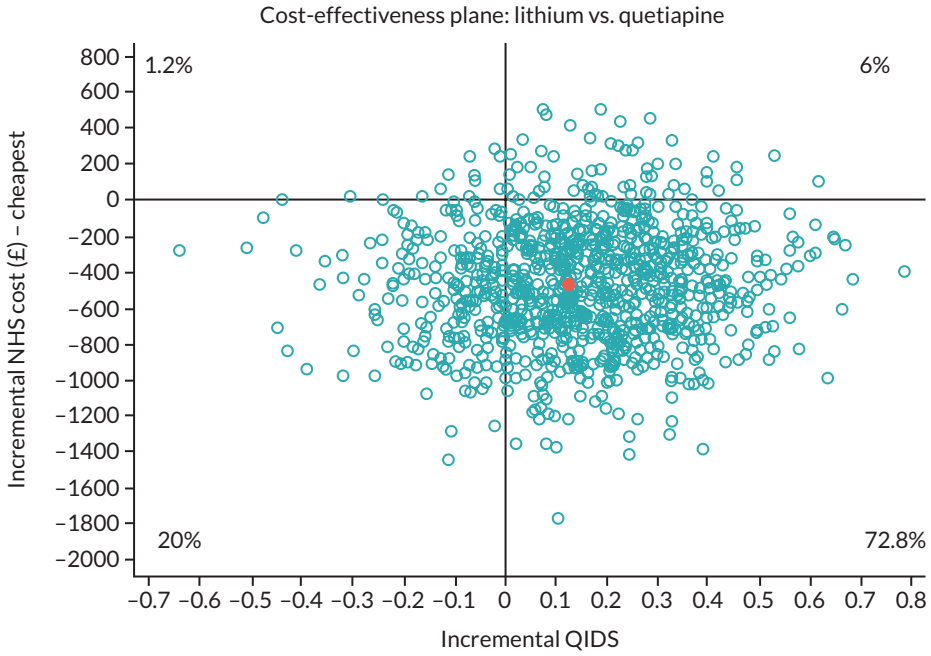

FIGURE 23 Cost-effectiveness plane of NHS and PSS cost and QIDS-SR differences.

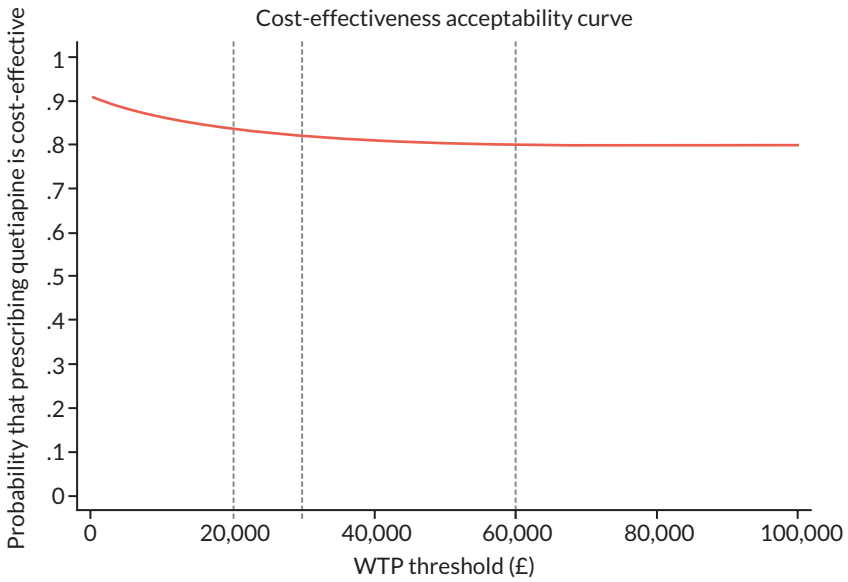

FIGURE 24 Cost-effectiveness acceptability curve showing the probability that quetiapine was the cost-effective option at different WTP thresholds for improvement in QIDS-SR score – NHS and PSS perspective.

quetiapine arm while 33.0% showed lower costs and an improvement in QIDS-SR score. Overall, 11.7% showed higher costs and a worsening in QIDS-SR score for quetiapine arm and 9.5% showed lower costs and a worsening in QIDS-SR score.

The CEAC for the 52-week follow-up period from the societal perspective is shown in *Figure 26*. The probability that quetiapine is the cost-effective option increases with the increased WTP threshold, however, it does not exceed 0.8 even when the decision maker is willing to pay £100,000 per point of improvement in the QIDS-SR score.

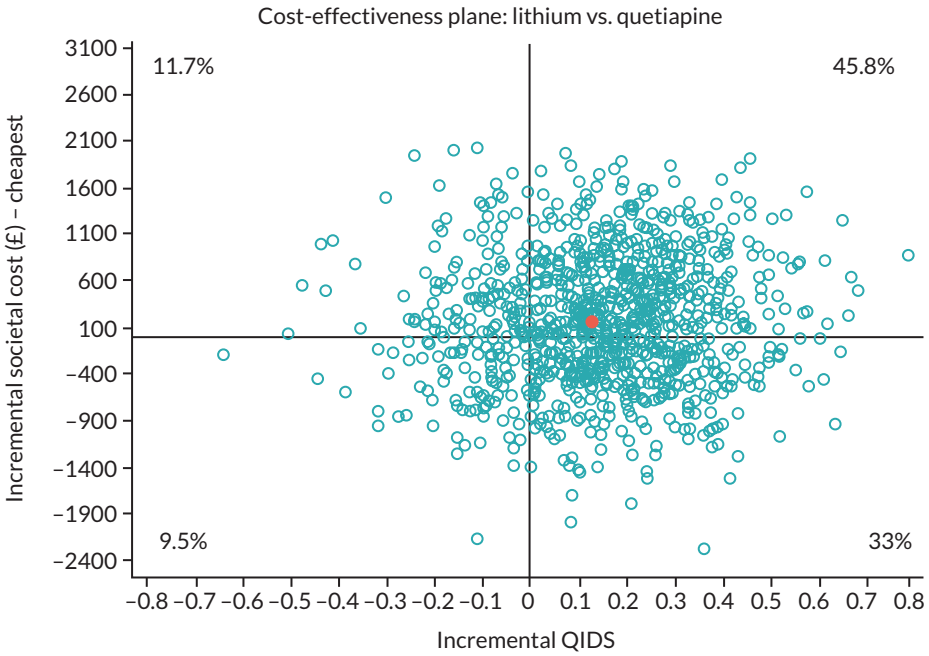

FIGURE 25 Cost-effectiveness plane of societal cost and QIDS-SR differences.

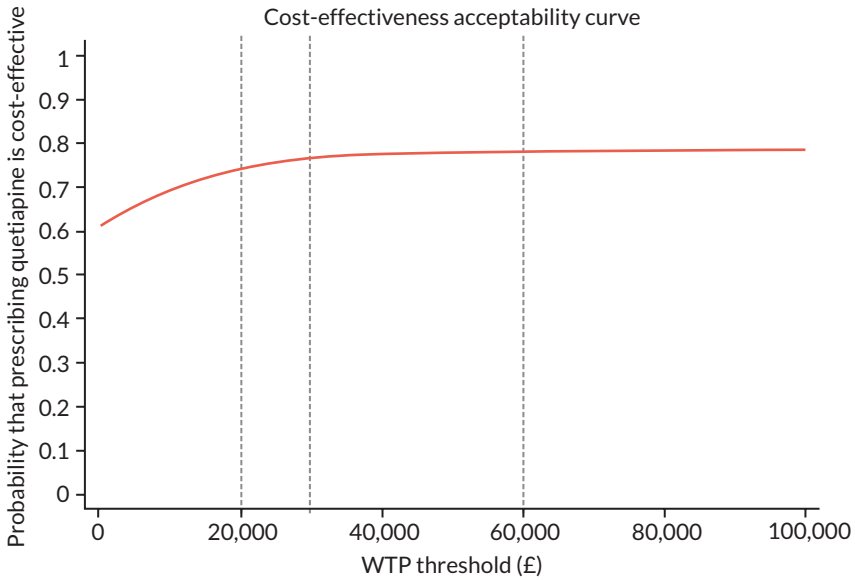

FIGURE 26 Cost-effectiveness acceptability curve showing the probability that quetiapine was the most cost-effective option at different WTP thresholds for an improvement in QIDS-SR score – societal perspective.

Sensitivity analyses

The results of the sensitivity analyses are reported in full in *Appendix 8*, and briefly summarised here. Quetiapine appeared to be less cost-effective in the following scenarios compared with the base case scenario:

- the generic unit cost of trial drugs was used (cheapest unit cost was used in the base case scenario)
- CD-MCAR was assumed for missing data (MAR was assumed for the base case scenario)
- whether participants' follow-up overlapped with occurrence of the COIVD-19 pandemic, and therefore, the analysis was modified to address the impact of COVID-19 (impact of the COVID-19 pandemic was not considered in the base case scenario).

However, in all scenarios tested, quetiapine remained the more cost-effective option, according to the NICE £20,000 WTP threshold for one additional unit of QALY.

Chapter 5 Discussion

Summary of findings and comparison with previous studies

The aim of this trial was to compare the clinical and cost-effectiveness of lithium versus quetiapine augmentation therapy for patients with TRD over 12 months. At baseline, most participants in our sample had three or more unsuccessful antidepressant treatment trials, and were moderately or markedly ill, with substantial functional impairment. Overall, our results suggested that quetiapine was superior to lithium augmentation therapy in terms of both clinical effectiveness on the depression symptom outcome, and cost-effectiveness. In agreement with our hypothesis, patients randomised to quetiapine showed a greater reduction in self-reported depression symptom severity over 12 months compared to those randomised to lithium, measured via the QIDS-SR. Although those who were randomised to quetiapine showed a longer time to discontinuation (median = 365 days) than those randomised to lithium (median = 212 days), this difference was not statistically significant. Quetiapine was more cost-effective than lithium from an NHS and PSS perspective (service use and drug costs), and a societal perspective (productivity loss in addition to NHS and PSS costs). With regard to secondary outcomes, those randomised to quetiapine showed lower clinician-rated depression (MADRS) and self-reported functional impairment (WSAS) at week 52, but not week 8, and better health-related quality of life at week 8, but not week 52, compared to those in the lithium arm. The direction of effects for several of the other secondary outcomes also favoured quetiapine (i.e. MADRS response and remission, CGI-I), but were small. Side effects, as measured by the PRISE, were similar across arms.

Over the course of 12 months, participants randomised to quetiapine experienced a greater reduction in depressive symptoms compared to those randomised to lithium. This was true for both ITT and PP populations, and for both self-reported (measured via the QIDS-SR) and clinician-rated depression severity (measured via the MADRS), though the latter at week 52 only. A previous trial comparing these augmentation options head-to-head found that quetiapine XR was non-inferior to lithium at week 6, defined as < 3 points difference on the MADRS (least squares mean change of −1.64).[21] In the present trial, the adjusted mean difference in MADRS scores was −1.50 at week 8, very similar to the result reported by Bauer *et al.* At week 52, the adjusted mean difference increased to −2.98, suggesting a benefit of quetiapine over lithium over a longer follow-up period. A difference of 2 or more points on the MADRS would be considered clinically relevant.[59]

Depending on time point, response rates in the present trial ranged from 15.0% to 23.4% for quetiapine, and 9.5–14.3% for lithium. Remission rates were between 10.3% and 12.1% for quetiapine, and 5.7–8.6% for lithium. These are considerably lower than those reported by Bauer *et al.*[21] (response 52.4% for quetiapine, 46.2% for lithium; remission 31.9% for quetiapine, 27.1% for lithium), likely due differences in illness severity and trial design. On the first of these, it is important to note that Bauer *et al.* recruited those with MDD who were significantly less treatment-resistant than our population, having exhibited an inadequate response to one or two antidepressant trials. In contrast, participants in the present trial were more severely treatment-resistant, with at least two, but more often three or more failed antidepressant treatment trials. In this way, the current trial is perhaps more comparable to levels 3 and 4 of the STAR*D trial, whereby participants received a different antidepressant or augmentation after two or three unsuccessful antidepressant trials respectively. Remission rates during levels 3 and 4 were between 10% and 20%,[60] more similar to those found in the present trial. In relation to trial design, in Bauer *et al.* participants were randomised directly to augmentation treatment. Crucially, our trial was pragmatic and designed to reflect real world clinical scenarios. Thus, participants were randomised to the *decision* to prescribe lithium or quetiapine, as would occur in the clinic, and initiation of the treatment after safety checks, and continuation of the medication thereafter, proceeded as per standard care. Thus, some participants were never prescribed or initiated the medication they were randomised to (e.g. if the treating clinician thought it inappropriate after safety checks had been undertaken) and participants could discontinue the medication without being excluded from the trial. We consider that inability to start treatment due to failed safety measures or other occurrences while safety measures are undertaken, and less optimal treatment regimens not subject to rigid clinical trial direction, are part-and-parcel of real-world augmentation treatment. As is often seen, such real-world treatment is reflected in lower response rates than those seen in more traditional clinical trials. It is also worth

reiterating that the response and remission rates are based on a single cross-sectional measure of outcome at 52 weeks, and likely fail to capture the longitudinal nature of TRD including fluctuation in symptom levels over time, hence our choice of weekly QIDS-SR measures over 52 weeks as the primary outcome.

Although time to discontinuation of the trial medication did not significantly differ between the two trial arms, visual inspection of Kaplan–Meier discontinuation curves shows a more gradual decline in the quetiapine arm than the lithium arm in the first half of the trial period, followed by an almost flat slope in the second half. This might suggest that after an initial group of discontinuations, there is a subsample of participants who perceive benefits from quetiapine and stay on the medication long term. In contrast, there is a sharper drop early in the trial period in the lithium arm, followed by a gradual but continuous decline in the slope until the end of the trial, suggesting a steadier rate of discontinuations over 12 months and possibly poorer tolerability or limited efficacy. Examining potential patient characteristics that predict discontinuation of the trial medications (e.g. physical health parameters or comorbidities) may reveal important factors to consider when prescribing these medications in clinical practice.

Overall, 38.9% of participants who initiated quetiapine and 50.0% of those who initiated lithium discontinued their trial medication within 12 months. Very few trials have included a long-term follow-up period making comparison with previous trials difficult; however, these figures are similar to those reported in cohort studies of lithium treatment.[61] In both arms, side effects were reported as the most common reason for discontinuation (45.9% of those who initiated and discontinued quetiapine, and 33.3% of those who initiated and discontinued lithium). The specific side effects leading to discontinuation were not recorded and PRISE scores, blood pressure and weight measurements were similar between arms at follow-up visits, making it difficult to determine which side effects may have prevented adherence to the trial medications. However, results from our previously published ancillary analysis examining patient views of the trial medications suggested that sedation, weight gain, and increased appetite were common for those taking quetiapine, and tremor and dry mouth were common for those taking lithium.[45] These side effects have been reported in previous trials of quetiapine and lithium augmentation therapy.[21,62,63]

Primary outcome results remained similar to those in the ITT population when effects were re-estimated for (1) participants who initiated the medication (similar to the PP analysis), (2) subsets of participants who were randomised before the beginning of the COVID-19 pandemic (1 February 2020), or whose outcome measures were collected before pandemic measures were implemented (1 March 2020), (3) participants who had a therapeutic treatment trial of the trial medication, and (4) participants who self-reported as treatment adherent (scoring ≥ 23 on the MARS-5). While the ITT analysis suggested that quetiapine was more effective than lithium in reducing depressive symptoms in clinical practice (when pragmatically some individuals will not be able to take them, or will need to discontinue), our results for those who took a therapeutic treatment trial suggest that quetiapine may also be more effective than lithium when taken for a minimum therapeutic trial of adequate dose and duration. Similar to the ITT analysis, the difference in time to discontinuation did not differ between the groups in participants who took a therapeutic treatment trial. However, it is interesting to note that the median time to discontinuation became longer in the lithium group (365.0 days as opposed to 212.0 days in the ITT analysis), matching the quetiapine group. This might suggest that those who are able to tolerate lithium and reach therapeutic serum levels can benefit from the medication in the long term. We also showed that the findings were robust to violation of the MAR assumption made by the models; we would have to assume people without week 52 data had worse QIDS-SR scores in the quetiapine arm only, which seems unlikely.

Although direct comparison of pre- and post-COVID participants is limited by the small sample size recruited after the onset of the pandemic, it is promising to note that (1) primary outcome results remained similar when those excluding those randomised post-pandemic onset, and (2) there did not appear to be difficulties with prescription and initiation of the trial medications during the pandemic. There was some concern that participants may experience difficulties obtaining their prescriptions due to lockdown and social distancing measures, leading to delays in initiating or even early discontinuation. This did not appear to be the case: rates of trial medication initiation were somewhat higher in those finishing the study after onset of the pandemic, and rates of discontinuation lower.

Regarding the cost-effectiveness of the trial medications, the economic analysis suggested that compared to lithium, quetiapine was associated with a lower cost to the NHS and PSS and better effectiveness, both in terms of QALY gains and improvement in QIDS-SR scores. At NICE's £20,000 WTP threshold per additional unit of QALY, the probability

that quetiapine was more cost effective was 0.99. When the generic drug costs were used instead of the cheapest, the probability that quetiapine was the most cost-effective option was reduced to 0.92 for the same WTP threshold. While the use of generic drug cost is the usual recommended approach, the scenario where branded versions are significantly cheaper, as is the current case for quetiapine, is unusual, and it would be more common that in practice the cheapest version will be used where possible.

Similarly, when examining societal costs, which included absence from work by participants and their carers in addition to NHS and PSS costs, quetiapine was again found to be the more cost-effective option, with a probability of 0.91. Although the probability of quetiapine being more cost-effective was somewhat reduced in comparison to the primary analysis, these results were robust to sensitivity analyses testing alternative assumptions regarding missing data and testing the impact of the COVID-19 pandemic.

While this trial and others suggest augmentation therapies are effective treatment options for some patients, very few actually receive them.[64,65] Although there is limited research on prescribing of lithium and quetiapine for unipolar depression, survey data from primary care physicians suggest that confidence in prescribing these medications for bipolar depression is low,[66] and it is usually recommended that they at least be initiated in secondary care.[24] Prescribing of lithium in primary care may be particularly challenging due to the blood monitoring required. Nevertheless, as only a minority of patients with TRD gain access to secondary care services, and there are often long waiting lists for such services, methods to increase access to augmentation therapies in primary care are needed. This could include more primary care based psychiatric services, together with education and training around prescribing and monitoring of augmentation therapies for primary care physicians; such initiatives may improve outcomes in patients with TRD who unfortunately at present are often not easily able to access augmentation therapies.

Strengths and limitations

A significant strength of the trial was the long-term follow-up period of 52 weeks. A recent meta-analysis of augmentation therapies for TRD found that of 28 studies, only two included a follow-up of 52 weeks or more, and the median duration of follow-up for all included studies was 6 weeks.[12] TRD is frequently a chronic condition,[22] and clinical guidelines recommend that those responding to an acute trial of medication continue the medication for continuation treatment for at least 9–12 months.[3] Our long-term follow-up period therefore gives a better picture of effectiveness in clinical practice. Another strength of the trial is the pragmatic design. Because the trial was designed to reflect clinical practice as far as possible, our results may be more generalisable to real-world settings than strictly controlled, non-pragmatic trials. A potential downside to this design is the lack of blinding. Patients, clinicians and researcher assistants were unblinded. In order to mitigate possible effects on clinician rated outcomes, blind raters administered the MADRS and CGI. Furthermore, trial statisticians remained blinded as described in the methods section. Nevertheless, expectancies regarding the trial medications, either from participants themselves or study staff, could conceivably have influenced the results of the trial. Notwithstanding this, neither treatment was new and both are already known to be effective augmentation therapies for TRD with little if any prior evidence of superiority of one over the other. There are some differences in monitoring, in that lithium involves more regular blood tests and initial contact with clinicians to adjust dosage. This might have led to some beneficial effects of enhanced non-pharmacological/placebo effects in the lithium group, or alternatively might have been seen as aversive or burdensome.

It should be noted at baseline that there appeared to be some chance imbalances between groups; for example, more participants in the lithium arm were unemployed compared to the quetiapine arm. There is some evidence to suggest that unemployment is associated with poorer outcomes in those with depression,[67] and therefore this imbalance may have affected outcomes in the current trial. Similarly, more participants in the lithium arm had contact with community nurses, psychiatrists, and dentists prior to baseline than in the quetiapine arm, while more participants in the quetiapine arm had contact with counsellors, physiotherapists, and other professionals. These may suggest differences in illness or treatment in the 3 months before baseline; nevertheless, comparability of the groups across psychiatric symptom measures at baseline suggests similar illness severity between arms.

Another limitation is the slower than expected rate of recruitment that occurred during the trial, which lead to a variation to contract to reduce the target sample size, and the resulting reduction in power to 80% for the discontinuation outcome, and 96.5% for the QIDS-SR outcome. A strength was our lower than expected rates of missing data for the primary outcomes. Only one participant (0.5%) was missing time to discontinuation data, lower than we had planned for in our power calculations (10%), and between 17.9% and 32.1% were missing QIDS-SR scores at follow-up visits, lower than planned in our power calculations (40% missing). Thus, the analysis was likely overpowered to detect differences in the primary outcomes.

However, there were substantial proportions of missing data for some of the secondary outcome measures. This is partly due to the pragmatic nature of the trial. Participants were not withdrawn from the trial if they missed follow-up visits, as primary outcome data could be obtained from True Colours or medical records. However, secondary outcome data could not be obtained in this way. The COVID-19 pandemic also contributed to missing outcome data: follow-up visits could not take place in person during the pandemic and instead took place virtually, so physical parameters, such as blood pressure and weight could not be measured. There is also a pattern of slightly more missing data for the QIDS-SR and secondary outcome measures in the lithium arm at week 26 and week 52 follow-up points, which is less apparent at week 8. This accompanies a slightly higher rate of trial medication discontinuation in the lithium arm compared to the quetiapine arm, and thus may represent less engagement in trial follow-up visits by patients in the lithium arm, introducing a possible detection bias. Alternatively, more engagement with the trial in the quetiapine group compared to the lithium group may have had differential effects on patients' depression.

The size of the difference between treatment arms on the QIDS-SR is also worthy of discussion. The primary outcome analysis used an AUC approach, using all available weekly QIDS-SR data for each participant, meaning that our results present a longitudinal view and reflect an overall reduction in depressive symptoms over the whole 12-month follow-up period, rather than a significant difference *at* 12 months only. The cumulative impact of this reduction in depression severity over 12 months may be more meaningful to patients, particularly since it is a frequent occurrence for patients with TRD to show fluctuations in symptom levels.[27] The results suggest a gradually increasing difference between quetiapine and lithium over time, and longer-term studies would be needed to see if that difference continued to expand with longer durations of treatment. It is also important to consider that this trial compared two well-established treatments for TRD added on to another recognised antidepressant treatment. Thus, we would expect smaller differences than an efficacy trial comparing, for example, lithium or quetiapine to placebo. Nevertheless, on a cross-sectional comparison at 52 weeks in the ITT analysis there was a 2-point difference in the QIDS-SR between treatments, which we would regard as a clinically significant difference in the context of two active treatments in a relatively severe sample of patients with TRD.

The lack of a 'usual care' or 'watchful waiting' condition is another limitation, especially given the modest response and remission rates found here. However, it has been suggested that although TRD is indeed often treatable, in some instances treatment should also focus on optimising symptom control and maximising functioning, rather than simply complete remission of depressive symptoms.[68] Our results suggested that quetiapine was also superior to lithium in this regard, with significantly less functional impairment at week 52 and better health-related quality of life at week 8.

Initial prescriptions were overseen by a trial clinician for all patients. Prescribing could then be handed over to patients' own clinicians where possible and at the discretion and judgement of the trial clinician. A limitation is that the impact of use of trial versus patients' own clinicians on outcome measures was not assessed. It is possible that some outcomes, such as treatment discontinuation, may have been influenced by to the use of trial versus patients' own clinicians. However, clinical guidelines recommend augmentation therapies are initiated in specialist mental health settings,[24] so this aspect of the trial design may have been reflective of clinical practice.

Lastly, due to resource constraints, we were unable to test the impact of using alternative medication on health-related quality of life for patients' caregivers. The direction of bias resulting from this is unclear. However, given the robustness of the base case conclusion (quetiapine being more cost-effective than lithium) under extensive sensitivity analyses, we believe that adding health-related quality of life for patients' caregivers into the analysis is unlikely to change the base case conclusion.

Patient and public involvement

Patient and public involvement shaped the design of the trial. While the research question was commissioned, all five PPI members who gave feedback on our proposal during the application stage were supportive of the research question and its importance. Considering the proposed primary outcome measures, we had concerns about the feasibility and intrusiveness of the regular QIDS-SR ratings via True Colours. PPI members were very supportive of regular QIDS-SR assessments and suggested that the frequency at which participants provided their QIDS-SR ratings via True Colours should be weekly, as they felt that this would give a better indication of long-term symptom fluctuations and that participants may benefit from more frequent assessments. Thus, PPI members' input shaped the design of one of the primary outcome measures. PPI members also recommended the addition of a follow-up visit at week 26, which was also incorporated into the study plan during the application stage.

One method of recruitment used during the trial involved GP mailouts; letters were sent to patients who appeared to be eligible for the trial according to their medical records. During the trial, we noted that we were receiving fewer responses to GP mailouts than expected. A PPI meeting was set up to gain feedback on the GP mailout content. As well as simplifying some of the language used, the PPI group recommended that the mailout letter should be clearer on certain details, such as who was running the study, whether a psychiatrist would be involved, and the trial medications. These recommendations were incorporated in an amendment to the GP mailouts. From these examples, it is apparent that PPI provided valuable insight into the design and running of the trial, and more could be gained from more frequent PPI. Thus, we plan to work with patients to give their feedback on the content and routes for dissemination of the results of the trial.

Equality, diversity and inclusion

This was a pragmatic trial, designed to reflect standard clinical practice as far as possible, within the constraints of a clinical trial. Participants were recruited at the point at which their clinician should have been considering augmentation therapy, and prescribing of the trial medication could be managed by the patients' primary or secondary care team, as per usual NHS practice. However, it must be noted that our trial sites were mostly large, well-resourced, urban research facilities, and this may limit applicability of our findings to elsewhere in the UK. Nonetheless, we used several different recruitment methods to reach patients with TRD. Although referral to secondary care is recommended for patients who have not responded to two or more antidepressants,[3] the majority of those with TRD are not referred to secondary care.[64,65] In addition to recruitment from secondary care services, we therefore used primary care recruitment methods, and in particular, participant identification based on non-response to two or more antidepressants, in an attempt to reach a representative sample of those with TRD.

Considering the demographic characteristics of our sample, our trial may have benefitted from a concerted effort to reach more people from non-white ethnic backgrounds. Comparing our sample to census data for England, our sample lacked representation from Asian and black ethnic groups. This was especially apparent when considering ethnicity by site. For example, only 2.4% participants recruited at SLaM were black and 8.4% were Asian. In London, 13.5% of the population are black and 20.7% are Asian.[69] Input from ethnically diverse PPI members during trial design and delivery may be an important way of improving recruitment of ethnically representative samples.

Clinical implications

Clinical guidelines for the treatment of depression currently recommend lithium or second-generation antipsychotics as first-line augmentation options for those who have not responded to antidepressant treatment alone.[24] However, evidence for these options mainly derives from studies in which lithium was added to TCAs, and antipsychotics to selective serotonin reuptake inhibitor (SSRI)/serotonin–norepinephrine reuptake inhibitors.[3] Further, very few studies have directly compared these options head-to-head, or followed patients beyond acute phase treatment. This trial aimed to provide evidence for clinicians and patients when choosing between augmentation options for TRD. Our results suggest that in routine clinical practice quetiapine may be more beneficial than lithium in reducing depressive

symptoms and improving functioning in people with TRD. Our results also suggest quetiapine to be more cost-effective, which has implications for NHS treatment recommendations. It must be noted that while other atypical antipsychotics also show some efficacy in the treatment of TRD, the lack of good head to head comparisons of atypical antipsychotics in TRD, together with varying mechanisms of action, side effect profiles, and tolerability, mean we cannot generalise our results to other atypical antipsychotics.

An important direction for future research will be to examine predictors of treatment response to establish whether there are additional factors which may guide treatment choice. We plan to conduct an ancillary analysis on the trial data to examine whether sociodemographic factors, personality, severity of depression and treatment resistance, depression chronicity and subtype, psychiatric comorbidity, and type of antidepressant predict treatment response to quetiapine or lithium. Predictors of discontinuation may also be an interesting complementary avenue to explore. Understanding which patients are more likely to discontinue these medications early would potentially reduce the number of unsuccessful treatment trials.

Importantly, the open-label, pragmatic design of the trial, whereby prescribing was continued by participants' primary or secondary care teams, gives insight into the effectiveness of recommended augmentation therapies in clinical practice over a relatively long follow-up period. Although meta-analyses have suggested superiority of both lithium and quetiapine over placebo in treating TRD,[12,70] evidence derives from trials with short-term follow-up periods and non-pragmatic designs. These trials generally report higher response and remission rates than those found in the present trial. Although some of this difference may relate to the more severely ill patients recruited to the current study, it is apparent that more effective treatments are required to improve functioning in those with TRD. It is also notable that the use of regular clinical assessments as part of measurement-based care may be another aspect of clinical trial practice that benefits patients in real world settings.[71] The use of tools such as True Colours is welcomed by patients[44] and may be a practical means to implement measurement-based care.

Additional information

CRediT contribution statement

Jess Kerr-Gaffney (https://orcid.org/0000-0001-5834-8307): Investigation, Project administration, Writing – original draft.

Zohra Zenasni (https://orcid.org/0000-0002-2335-1378): Formal analysis, Writing – reviewing and editing.

Kimberley Goldsmith (https://orcid.org/0000-0002-0620-7868): Formal analysis, Methodology, Writing – reviewing and editing.

Nahel Yaziji (https://orcid.org/0000-0002-2273-7957): Formal analysis, Writing – reviewing and editing.

Huajie Jin (https://orcid.org/0000-0002-3872-3998): Formal analysis, Writing – reviewing and editing, Methodology.

Alessandro Colasanti (https://orcid.org/0000-0001-6017-801X): Resources, Supervision, Investigation, Writing – reviewing and editing.

John Geddes (https://orcid.org/0000-0002-5281-5960): Resources, Funding acquisition, Supervision, Investigation, Writing – reviewing and editing.

David Kessler (https://orcid.org/0000-0001-5333-132X): Resources, Supervision, Investigation, Writing – reviewing and editing.

R Hamish McAllister-Williams (https://orcid.org/0000-0001-9966-1834): Resources, Funding acquisition, Supervision, Investigation, Writing – reviewing and editing.

Allan H Young (https://orcid.org/0000-0003-2291-6952): Funding acquisition, Supervision, Writing – reviewing and editing.

Alvaro Barrera (https://orcid.org/0000-0003-4716-8487): Investigation, Writing – reviewing and editing.

Lindsey Marwood (https://orcid.org/0000-0002-5818-2199): Investigation, Project administration, Writing – reviewing and editing.

Rachael W Taylor (https://orcid.org/0000-0001-6471-537X): Investigation, Project administration, Writing – reviewing and editing.

Helena Tee (https://orcid.org/0000-0001-6444-5787): Investigation, Project administration, Writing – reviewing and editing.

Anthony J Cleare (https://orcid.org/0000-0002-6990-939X): Conceptualisation, Funding acquisition, Methodology, Supervision, Writing – reviewing and editing.

Acknowledgements

This study represents independent research supported by the National Institute for Health and Care Research (NIHR)-Wellcome King's Clinical Research Facility and the NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. KG at KCL is also supported by the National Institute for Health and Care Research (NIHR) Applied Research Collaboration South London (NIHR ARC South London) at King's College Hospital NHS Foundation Trust.

We would like to thank the independent Trial Steering Committee (TSC; members: David Baldwin (chair), Jasmine Stuart Smith, Joanna Clarke, Muzaffer Kaser, and Peter Haddad; observer: Mohamed Abdelghani) and independent Data Monitoring and Ethics Committee (DMEC; members: Mike Bradburn (chair), Christopher Dickens, and Louise Marston). For details of the TSC and DMEC charters and roles, please contact the corresponding author. We thank the contribution of Patient and Public Involvement (PPI) members who provided feedback and valuable suggestions in the development of this trial and the study documents.

We acknowledge the support of the Maudsley Biomedical Research Centre's Feasibility and Acceptability Support Team for Researchers, the invaluable administrative support provided by Caroline Loveland, and Baptiste Leurent from the University College London Department of Statistical Science for helpful conversations about the missing not at random sensitivity analyses. Finally, we would like to thank the NIHR Clinical Research Network for their support in delivering this research.

Lithium versus Quetiapine for Depression Study Group members

South London and Maudsley NHS Foundation Trust/King's College London:

Study Clinicians: James Rucker, Ryan Williams, Camilla Day, Ilia Bountouni, Mario Juruena, Luke Jelen, Allison Edwards, Carol Kan, Chaitra Jairaj, Lauren Waterman, Robert McCutcheon, Romayne Gadelrab, Rashmi Patel, Patrick McLoughlin.

Researchers: Elana Day, Rupal Shah, Joseph Cattell, Valeria de Angel, Emma Incecik, Anda Ancane, Emmanuella Oprea, Andrea Ulrichsen, Viktoriya Nikolova, Tim Mantingh, Dimosthenis Tsapekos.

Advisors: Andrew Pickles, Daniel Dietch, Rebecca Strawbridge.

Former statisticians/health economists: Renee Romeo, Rachel Holland.

Cumbria, Northumberland, Tyne and Wear NHS Foundation Trust/Newcastle University:

Study clinicians: Stuart Watson, Hossam Marey, Niraj Ahuja, Eman Arebi, Daniel Armstrong, Dheeraj Buruju, Afamefuna Emechebe, Oghenefejiro Esi, Jayne Healicon, Selma Osman, William Stageman, Neeti Sud, Vineet Raj, Jennifer Burgess

Researchers: Jake Hutchinson, Kimberley Nortey, Susan Wilson, Wendy Hall, Joe Swift, Rosie Carr, Jahnese Hamilton

Tees, Esk and Wear Valleys NHS Foundation Trust:

Site PI: Joe Reilly.

Study clinicians: Rajesh Nair.

Oxford Health NHS Foundation Trust/University of Oxford:

Study clinicians: Phil Cowen, Mike Browning.

Researchers: Lucas McKeown, Sarah Mather, Jen Potts, Chris Millar, Sarah Hollingsworth, Rebecca Dean, Ceri Morgan, Jade Harvey, Fay Davies.

Advisors: Andrea Cipriani, Mary-Jane Attenburrow.

Avon and Wiltshire Mental Health Partnership NHS Trust/University of Bristol:

Clinicians: Jonathan Evans, Jonathan Davies

Researchers: Rwth Leach, Catherine Roiz de S'a.

Sussex Partnership NHS Foundation Trust/University of Sussex:

Researchers: Joanna Pooley, Guy Emery, Serena Gregory .

Patient data statement

This work uses data provided by patients and collected by the NHS as part of their care and support. Using patient data is vital to improve health and care for everyone. There is huge potential to make better use of information from people's patient records, to understand more about disease, develop new treatments, monitor safety and plan NHS services. Patient data should be kept safe and secure, to protect everyone's privacy, and it's important that there are safeguards to make sure that they are stored and used responsibly. Everyone should be able to find out about how patient data are used. #datasaveslives You can find out more about the background to this citation here: https://understandingpatientdata.org.uk/data-citation.

Data-sharing statement

Patient-level data will not be made publicly available. All data requests should be submitted to the corresponding author for consideration. Access to anonymised data may be granted following review.

Ethics statement

A favourable opinion was given by the East of England – Cambridge South REC (reference number: 16/EE/0318) on 20 September 2016. The trial received approval from the Health Research Authority (HRA) and the Medicines and Healthcare Products Regulatory Agency (MHRA; EudraCT reference number: 2016-001637-27).

Information governance statement

King's College London and South London and Maudsley NHS Foundation Trust are committed to handling all personal information in line with the UK Data Protection Act (2018) and the UK General Data Protection Regulation. Under the Data Protection legislation, King's College London and South London and Maudsley NHS Foundation Trust are the Data Controllers, and you can find out more about how we handle personal data, including how to exercise your individual rights and the contact details for our Data Protection Officer here: www.kcl.ac.uk/research/support/rgei/research-ethics/use-of-personal-data-in-research.

Disclosure of interests

Full disclosure of interests: Completed ICMJE forms for all authors, including all related interests, are available in the toolkit on the NIHR Journals Library report publication page at https://doi.org/10.3310/YQVF5347.

Primary conflicts of interest: Jess Kerr-Gaffney owns shares in AstraZeneca and GSK plc. Kimberley Goldsmith has received grant funding from NIHR, National Institutes of Health, Juvenile Diabetes Research Foundation International, Stroke Association, Medical Research Council (MRC), and UK research councils, received support from NIHR for conference and meeting attendance, and sits on trial steering committees for the following NIHR-funded studies: DPACT, EPICC-ID, ASSURED, ORBIT, and SATURN. David Kessler has received grant funding from NIHR (RP-PG-0514-20012, INTERACT), received payment from the Royal College of General Practitioners for a presentation, and sits on advisory boards for NIHR-funded studies. R Hamish McAllister-Williams has received payments/consultation fees from LivaNova, Janssen-Cilag, Sage Therapeutics, P1Vital, Sothema, Takeda, and Lundbeck, support for attending meetings from Janssen-Cilag, is director of education for the British Association for Psychopharmacology, and chair of the trial steering committee for the

NIHR-funded SNAPPER trial. Allan H Young has received grant funding from the European Commission – Horizon 2020, EU Brain Council, NIHR (SNAPPER, Rhapsody, PsiDeR, PAX-BD, BRC translational therapies), MRC, Alzheimer's Research UK, Probiotics International Ltd, LivaNova, Janssen, COMPASS Pathways plc., Novartis, CIHR, NARSAD, Stanley Medical Research Institute, Wellcome Trust, Royal College of Physicians, British Medical Association, UBC-VGH Foundation, WEDC, CCS Depression Research Fund, and MSFHR, has received payments or honoraria for presentations or advisory roles from Flow Neuroscience, Novartis, Roche, Janssen, Takeda, Noema pharma, COMPASS Pathways Plc., Astrazenaca, Boehringer Ingelheim, Eli Lilly, LivaNova, Lundbeck, Sunovion, Servier, Janssen, Allergan, Bionomics, Sumitomo Dainippon Pharma, Sage, Novartis, and Neurocentrx, is a trustee of Bipolar UK and the Drug Safety Research Unit, is editor of the Journal of Psychopharmacology and deputy editor of BJPsych Open. Lindsey Marwood and Rachael W Taylor are current employees of COMPASS Pathways plc and hold shares or options in COMPASS Pathways plc. This work is unrelated to COMPASS Pathways plc. Anthony J Cleare has received grant funding from the MRC, ADM Protexin Ltd, NIHR (PAX-D, SLaM BRC), European Union Horizon Europe/Innovate UK, Beckley Psytech Ltd, and Wellcome Trust, has received payment or honoraria for presentations and/or consulting from Janssen, Otsuka, Viatris, COMPASS Pathways Plc., and Medscape, and is president of the International Society for Affective Disorders.

Publications

Publications and conference presentations resulting from the study

Marwood L, Taylor R, Goldsmith K, Romeo R, Holland R, Pickles A, *et al.* Study protocol for a randomised pragmatic trial comparing the clinical and cost effectiveness of lithium and quetiapine augmentation in treatment resistant depression (the LQD study). *BMC Psychiatry* 2017;**17**:1–12. https://doi.org/10.1186/s12888-017-1393-0

Incecik E, Taylor RW, Valentini B, Hatch SL, Geddes JR, Cleare AJ, *et al.* Online mood monitoring in treatment-resistant depression: qualitative study of patients' perspectives in the NHS. *BJPsych Bull* 2020;**44**:47–52. https://doi.org/10.1192/bjb.2019.92

Day E, Shah R, Taylor RW, Marwood L, Nortey K, Harvey J, *et al.* A retrospective examination of care pathways in individuals with treatment-resistant depression. *BJPsych Open* 2021;**7**:e101,1–11. https://doi.org/10.1192/bjo.2021.59

McKeown L, Taylor RW, Day E, Shah R, Marwood L, Tee H, *et al.* Patient perspectives of lithium and quetiapine augmentation treatment in treatment-resistant depression: a qualitative assessment. *J Psychopharmacol* 2022;**36**:557–65. https://doi.org/10.1177/02698811221089042

Cleare AJ. *A Review of Antidepressant Augmentation for TRD, Including Results from the Lithium versus Quetiapine for Depression (LQD) study* [Oral presentation]. British Association for Psychopharmacology Summer Meeting, Manchester, UK, 2023.

Cleare AJ. *Latest Evidence on Pharmacological Augmentation for Treatment-Resistant Depression: Does the Lithium versus Quetiapine (LQD) Study Change Best Practice?* [Oral presentation]. International Society for Affective Disorders, Milan, Italy, 2023.

Cleare AJ. *Pharmacological Management of Treatment Resistant Depression: Choosing and Using Augmentation Therapies* [Oral presentation]. RCPsych International Congress, Liverpool, UK, 2023.

Kerr-Gaffney J, Zenasni Z, Goldsmith K, Yaziji N, Jin H, Colasanti A, *et al. A Randomised Pragmatic Trial Comparing the Clinical and Cost-Effectiveness of Lithium Versus Quetiapine Augmentation for Treatment Resistant Depression (LQD)* [Poster presentation]. International Society for Affective Disorders, Milan, Italy, 2023.

Cleare AJ, Kerr-Gaffney J, Goldsmith K, Zenasni Z, Yaziji N, Jin H, *et al.* Clinical and cost-effectiveness of lithium versus quetiapine augmentation for treatment-resistant depression: a pragmatic, open-label, parallel-group, randomised controlled superiority trial in the UK. *Lancet Psychiatry* 2025;**12**(4):276–88. https://doi.org/10.1016/S2215-0366(25)00028-8

References

1. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, *et al.*; National Comorbidity Survey Replication. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA* 2003;**289**:3095–105. https://doi.org/10.1001/jama.289.23.3095
2. Üstün TB, Ayuso-Mateos JL, Chatterji S, Mathers C, Murray CJL. Global burden of depressive disorders in the year 2000. *Br J Psychiatry* 2004;**184**:386–92. https://doi.org/10.1192/bjp.184.5.386
3. Cleare A, Pariante C, Young A, Anderson I, Christmas D, Cowen P, *et al.*; the Members of the Consensus Meeting. Evidence-based guidelines for treating depressive disorders with antidepressants: a revision of the 2008 British Association for Psychopharmacology guidelines. *J Psychopharmacol* 2015;**29**:459–525. https://doi.org/10.1177/0269881115581093
4. 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. https://doi.org/10.1176/ajp.2006.163.11.1905
5. Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, *et al.*; STAR*D Study Team. 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. https://doi.org/10.1176/appi.ajp.163.1.28
6. Fekadu A, Rane LJ, Wooderson SC, Markopoulou K, Poon L, Cleare AJ. Prediction of longer-term outcome of treatment-resistant depression in tertiary care. *Br J Psychiatry* 2012;**201**:369–75. https://doi.org/10.1192/bjp.bp.111.102665
7. Adekkanattu P, Olfson M, Susser LC, Patra B, Vekaria V, Coombes BJ, *et al*. Comorbidity and healthcare utilization in patients with treatment resistant depression: a large-scale retrospective cohort analysis using electronic health records. *J Affect Disord* 2023;**324**:102–13. https://doi.org/10.1016/j.jad.2022.12.044
8. Li G, Fife D, Wang G, Sheehan JJ, Bodén R, Brandt L, *et al*. All-cause mortality in patients with treatment-resistant depression: a cohort study in the US population. *Ann Gen Psychiatry* 2019;**18**:1–8. https://doi.org/10.1186/s12991-019-0248-0
9. Zhdanava M, Pilon D, Ghelerter I, Chow W, Joshi K, Lefebvre P, Sheehan JJ. The prevalence and national burden of treatment-resistant depression and major depressive disorder in the United States. *J Clin Psychiatry* 2021;**82**:20m13699. https://doi.org/10.4088/JCP.20m13699
10. Johnston KM, Powell LC, Anderson IM, Szabo S, Cline S. The burden of treatment-resistant depression: a systematic review of the economic and quality of life literature. *J Affect Disord* 2019;**242**:195–210. https://doi.org/10.1016/j.jad.2018.06.045
11. Wooderson SC, Fekadu A, Markopoulou K, Rane LJ, Poon L, Juruena MF, *et al*. Long-term symptomatic and functional outcome following an intensive inpatient multidisciplinary intervention for treatment-resistant affective disorders. *J Affect Disord* 2014;**166**:334–42. https://doi.org/10.1016/j.jad.2014.05.013
12. Strawbridge R, Carter B, Marwood L, Bandelow B, Tsapekos D, Nikolova VL, *et al*. Augmentation therapies for treatment-resistant depression: systematic review and meta-analysis. *Br J Psychiatry* 2019;**214**:42–51. https://doi.org/10.1192/bjp.2018.233
13. Thase ME, Shelton RC, Khan A. Treatment with venlafaxine extended release after SSRI nonresponse or intolerance. *J Clin Psychopharmacol* 2006;**26**:250–8. https://doi.org/10.1097/01.jcp.0000219922.19305.08
14. Wade AG, Crawford GM, Yellowlees A. Efficacy, safety and tolerability of escitalopram in doses up to 50 mg in Major Depressive Disorder (MDD): an open-label, pilot study. *BMC Psychiatry* 2011;**11**:42. https://doi.org/10.1186/1471-244X-11-42

15. Adli M, Baethge C, Heinz A, Langlitz N, Bauer M. Is dose escalation of antidepressants a rational strategy after a medium–dose treatment has failed? *Eur Arch Psychiatry Clin Neurosci* 2005;**255**:387–400. https://doi.org/10.1007/s00406-005-0579-5

16. Dold M, Bartova L, Rupprecht R, Kasper S. Dose escalation of antidepressants in unipolar depression: a meta-analysis of double-blind, randomized controlled trials. *Psychother Psychosom* 2017;**86**:283–91. https://doi.org/10.1159/000477770

17. Rush AJ, Trivedi MH, Stewart JW, Nierenberg AA, Fava M, Kurian BT, *et al*. Combining Medications to Enhance Depression Outcomes (CO-MED): acute and long-term outcomes of a single-blind randomized study. *Am J Psychiatry* 2011;**168**:689–701. https://doi.org/10.1176/appi.ajp.2011.10111645

18. Bauer M, Adli M, Ricken R, Severus E, Pilhatsch M. Role of lithium augmentation in the management of major depressive disorder. *CNS Drugs* 2014;**28**:331–42. https://doi.org/10.1007/s40263-014-0152-8

19. Nelson JC, Papakostas GI. Atypical antipsychotic augmentation in major depressive disorder: a meta-analysis of placebo-controlled randomized trials. *Am J Psychiatry* 2009;**166**:980–91. https://doi.org/10.1176/appi.ajp.2009.09030312

20. Zhou X, Ravindran AV, Qin B, Del Giovane C, Li Q, Bauer M, *et al*. Comparative efficacy, acceptability, and tolerability of augmentation agents in treatment-resistant depression: Systematic review and network meta-analysis. *J Clin Psychiatry* 2015;**76**:e487–98. https://doi.org/10.4088/JCP.14r09204

21. Bauer M, Dell'Osso L, Kasper S, Pitchot W, Vansvik ED, Köhler J, *et al*. Extended-release quetiapine fumarate (quetiapine XR) monotherapy and quetiapine XR or lithium as add-on to antidepressants in patients with treatment-resistant major depressive disorder. *J Affect Disord* 2013;**151**:209–19. https://doi.org/10.1016/j.jad.2013.05.079

22. Fekadu A, Wooderson SC, Markopoulou K, Cleare AJ. The Maudsley staging method for treatment-resistant depression: prediction of longer-term outcome and persistence of symptoms. *J Clin Psychiatry* 2009;**70**:952–7. https://doi.org/10.4088/JCP.08m04728

23. Bauer M, Pfennig A, Severus E, Whybrow PC, Angst J, Möller H-J; World Federation of Societies of Biological Psychiatry; Task Force on Unipolar Depressive Disorders. World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for biological treatment of unipolar depressive disorders, part 1: update 2013 on the acute and continuation treatment of unipolar depressive disorders. *World J Biol Psychiatry* 2013;**14**:334–85. https://doi.org/10.3109/15622975.2013.804195

24. National Institute for Health and Care Excellence. *Depression in Adults: Treatment and Management*. 2022. URL: www.nice.org.uk/guidance/ng222 (accessed 1 February 2023).

25. Bauer M, Pretorius HW, Constant EL, Earley WR, Szamosi J, Brecher M. Extended-release quetiapine as adjunct to an antidepressant in patients with major depressive disorder. *J Clin Psychiatry* 2009;**70**:540–9. https://doi.org/10.4088/JCP.08m04629

26. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, *et al*. The 16-item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry* 2003;**54**:573–83. https://doi.org/10.1016/S0006-3223(02)01866-8

27. Vergunst FK, Fekadu A, Wooderson SC, Tunnard CS, Rane LJ, Markopoulou K, Cleare AJ. Longitudinal course of symptom severity and fluctuation in patients with treatment-resistant unipolar and bipolar depression. *Psychiatry Res* 2013;**207**:143–9. https://doi.org/10.1016/j.psychres.2013.03.022

28. Rush AJ, Carmody TJ, Ibrahim HM, Trivedi MH, Biggs MM, Shores-Wilson K, *et al*. Comparison of self-report and clinician ratings on two inventories of depressive symptomatology. *Psychiatr Serv* 2006;**57**:829–37. https://doi.org/10.1176/ps.2006.57.6.829

29. Geddes JR, Gardiner A, Rendell J, Voysey M, Tunbridge E, Hinds C, *et al*.; CEQUEL Investigators and Collaborators. Comparative evaluation of quetiapine plus lamotrigine combination versus quetiapine

monotherapy (and folic acid versus placebo) in bipolar depression (CEQUEL): a 2 × 2 factorial randomised trial. *Lancet Psychiatry* 2016;**3**:31–9. https://doi.org/10.1016/S2215-0366(15)00450-2

30. Marwood L, Taylor R, Goldsmith K, Romeo R, Holland R, Pickles A, *et al*. Study protocol for a randomised pragmatic trial comparing the clinical and cost effectiveness of lithium and quetiapine augmentation in treatment resistant depression (the LQD study). *BMC Psychiatry* 2017;**17**:1–12. https://doi.org/10.1186/s12888-017-1393-0
31. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, *et al*. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 1998;**59**:22–33;quiz 34.
32. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960;**23**:56–62. https://doi.org/10.1136/jnnp.23.1.56
33. Taylor D, Patron C, Kapur S. *The Maudsley Prescribing Guidelines in Psychiatry*. 12th edn. New York: Wiley-Blackwell; 2015.
34. Joint Formularly Committee. *British National Formulary*. London: Joint Formularly Committe; 2023.
35. Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979;**134**:382–9. https://doi.org/10.1192/bjp.134.4.382
36. Busner J, Targum SD. Global impressions scale: applying a research tool in clinical practice. *Psychiatry* 2007;**4**:28–37.
37. Brooks R, Rabin R, Charro F. *The Measurement and Valuation of Health Status Using EQ-5D: A European Perspective*. Dordrecht: Springer; 2003.
38. Mundt JC, Marks IM, Shear MK, Greist JM. The Work and Social Adjustment Scale: a simple measure of impairment in functioning. *Br J Psychiatry* 2002;**180**:461–4. https://doi.org/10.1192/bjp.180.5.461
39. Clatworthy J, Bowskill R, Parham R, Rank T, Scott J, Horne R. Understanding medication non-adherence in bipolar disorders using a Necessity-Concerns Framework. *J Affect Disord* 2009;**116**:51–5. https://doi.org/10.1016/j.jad.2008.11.004
40. Rush AJ, Fava M, Wisniewski SR, Lavori PW, Trivedi MH, Sackeim HA, *et al*.; STAR*D Investigators Group. Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. *Control Clin Trials* 2004;**25**:119–42. https://doi.org/10.1016/S0197-2456(03)00112-0
41. Chisholm D, Knapp MRJ, Knudsen HC, Amaddeo F, Gaite L, van Wijngaarden B; EPSILON Study Group. Client Socio-Demographic and Service Receipt Inventory – European Version: development of an instrument for international research. *Br J Psychiatry* 2000;**177**:s28–33. https://doi.org/10.1192/bjp.177.39.s28
42. Wechsler D. *Wechsler Adult Intelligence Scale*. 4th edn. San Antonio, TX: Psychological Corporation; 2014.
43. Fekadu A, Donocik JG, Cleare AJ. Standardisation framework for the Maudsley staging method for treatment resistance in depression. *BMC Psychiatry* 2018;**18**:100. https://doi.org/10.1186/s12888-018-1679-x
44. Incecik E, Taylor RW, Valentini B, Hatch SL, Geddes JR, Cleare AJ, Marwood L. Online mood monitoring in treatment-resistant depression: qualitative study of patients' perspectives in the NHS. *BJPsych Bull* 2020;**44**:47–52. https://doi.org/10.1192/bjb.2019.92
45. McKeown L, Taylor RW, Day E, Shah R, Marwood L, Tee H, *et al*. Patient perspectives of lithium and quetiapine augmentation treatment in treatment-resistant depression: a qualitative assessment. *J Psychopharmacol* 2022;**36**:557–65. https://doi.org/10.1177/02698811221089042
46. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 1995;**57**:289–300.
47. White IR, Thompson SG. Adjusting for partially missing baseline measurements in randomized trials. *Stat Med* 2005;**24**:993–1007. https://doi.org/10.1002/sim.1981

48. Stone JK, Shafer LA, Graff LA, Lix L, Witges K, Targownik LE, *et al*. Utility of the MARS-5 in assessing medication adherence in IBD. *Inflamm Bowel Dis* 2021;**27**:317–24. https://doi.org/10.1093/ibd/izaa056

49. Irshaidat S, Gustafsson M, Norberg H. Self-reported medication adherence among older people admitted to hospital: a descriptive study. *Drugs Real World Outcomes* 2023;**10**:23–9. https://doi.org/10.1007/s40801-023-00352-8

50. Yen YF, Lai HH, Kuo YC, Chan SY, Chen LY, Chen CC, *et al*. Association of depression and antidepressant therapy with antiretroviral therapy adherence and health-related quality of life in men who have sex with men. *PLOS ONE* 2022;**17**:e0264503. https://doi.org/10.1371/journal.pone.0264503

51. Cro S, Morris TP, Kahan BC, Cornelius VR, Carpenter JR. A four-step strategy for handling missing outcome data in randomised trials affected by a pandemic. *BMC Med Res Methodol* 2020;**20**:208. https://doi.org/10.1186/s12874-020-01089-6

52. National Institute for Health and Care Excellence. *Guide to the Methods of Technology Appraisal (PMG9)*. London: National Institute for Health and Care Excellence; 2013. URL: www.nice.org.uk/process/pmg9/ (accessed 1 February 2023).

53. Joint Formulary Committee. *British National Formulary*. London: BMJ Group and Pharmaceutical Press; 2020.

54. Curtis L, Burns A. *Unit Costs of Health and Social Care 2020*. Canterbury: Personal Social Services Research Unit; 2020.

55. NHS Digital. *Reference Costs (Ref Costs)*. 2021. URL: https://www.england.nhs.uk/publication/2020-21-national-cost-collection-data-publication/ (accessed 1 February 2023).

56. Office for National Statistics. *Earnings and Working Hours Census Data*. 2021. URL: www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours (accessed 1 February 2023).

57. Little RJA. Modeling the drop-out mechanism in repeated-measures studies. *J Am Stat Assoc* 1995;**90**:1112–21. https://doi.org/10.1080/01621459.1995.10476615

58. Van Hout BA, Al MJ, Gordon GS, Rutten FFH. Costs, effects and C/E-ratios alongside a clinical trial. *Health Econ* 1994;**3**:309–19. https://doi.org/10.1002/hec.4730030505

59. Montgomery SA, Möller HJ. Is the significant superiority of escitalopram compared with other antidepressants clinically relevant? *Int Clin Psychopharmacol* 2009;**24**:111–8. https://doi.org/10.1097/YIC.0b013e32832a8eb2

60. Gaynes BN, Warden D, Trivedi MH, Wisniewski SR, Fava M, Rush AJ. What did STAR*D teach us? Results from a large-scale, practical, clinical trial for patients with depression. *Psychiatr Serv* 2009;**60**:1439–45. https://doi.org/10.1176/ps.2009.60.11.1439

61. Öhlund L, Ott M, Oja S, Bergqvist M, Lundqvist R, Sandlund M, *et al*. Reasons for lithium discontinuation in men and women with bipolar disorder: a retrospective cohort study. *BMC Psychiatry* 2018;**18**:37. https://doi.org/10.1186/s12888-018-1895-4

62. Anderson IM, Sarsfield A, Haddad PM. Efficacy, safety and tolerability of quetiapine augmentation in treatment resistant depression: an open-label, pilot study. *J Affect Disord* 2009;**117**:116–9. https://doi.org/10.1016/j.jad.2008.12.016

63. McIntyre A, Gendron A, McIntyre A. Quetiapine adjunct to selective serotonin reuptake inhibitors or venlafaxine in patients with major depression, comorbid anxiety, and residual depressive symptoms: a randomized, placebo-controlled pilot study. *Depress Anxiety* 2007;**24**:487–94. https://doi.org/10.1002/da.20275

64. Day E, Shah R, Taylor RW, Marwood L, Nortey K, Harvey J, *et al*. A retrospective examination of care pathways in individuals with treatment-resistant depression. *BJPsych Open* 2021;**7**:e101. https://doi.org/10.1192/bjo.2021.59

65. Wiles N, Taylor A, Turner N, Barnes M, Campbell J, Lewis G, *et al*. Management of treatment-resistant depression in primary care: a mixed-methods study. *Br J Gen Pract* 2018;**68**:e673–81.

66. Sansone RA, Forbis JS, Sosa T. Primary care perspectives on treating bipolar disorder: a cross-sectional survey. *Prim Care Companion CNS Disord* 2011;**13**:10m01072. https://doi.org/10.4088/PCC.10m01072

67. Buckman JEJ, Saunders R, Stott J, Cohen ZD, Arundell LL, Eley TC, *et al*. Socioeconomic indicators of treatment prognosis for adults with depression. *JAMA Psychiatry* 2022;**79**:406–16. https://doi.org/10.1001/jamapsychiatry.2022.0100

68. McAllister-Williams RH, Arango C, Blier P, Demyttenaere K, Falkai P, Gorwood P, *et al*. The identification, assessment and management of difficult-to-treat depression: an international consensus statement. *J Affect Disord* 2020;**267**:264–82. https://doi.org/10.1016/j.jad.2020.02.023

69. Office for National Statistics. *England and Wales 2021 Census*. London: Office for National Statistics; 2021.

70. Nuñez NA, Joseph B, Pahwa M, Kumar R, Resendez MG, Prokop LJ, *et al*. Augmentation strategies for treatment resistant major depression: a systematic review and network meta-analysis. *J Affect Disord* 2022;**302**:385–400. https://doi.org/10.1016/j.jad.2021.12.134

71. Zhu M, Hong RH, Yang T, Yang X, Wang X, Liu J, *et al*. The efficacy of measurement-based care for depressive disorders. *J Clin Psychiatry* 2021;**82**:21r14034. https://doi.org/10.4088/JCP.21r14034

72. Leurent B, Gomes M, Faria R, Morris S, Grieve R, Carpenter JR. Sensitivity analysis for not-at-random missing data in trial-based cost-effectiveness analysis: a tutorial. *PharmacoEconomics* 2018;**36**:889–901. https://doi.org/10.1007/s40273-018-0650-5

73. White IR, Horton NJ, Carpenter J, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. *BMJ* 2011;**342**:d40. https://doi.org/10.1136/bmj.d40

Appendix 1 National Health Service/Personal Social Services health and social care unit costs

TABLE 25 Unit costs, all costs are inflated to year 2022–3 using the NHS cost inflation index (NHSCII)

Service	£ sterling	Source
Community-based services		
GP visit (9.22 minutes)	41.70	PSSRU 2020–21
Practice nurse visit (hour)	49.54	PSSRU 2015–16
District nurse visit (hour)	57.6	PSSRU 2015–16
Community nurse visit (hour)	57.6	PSSRU 2015–16
Occupational therapist	52.08	PSSRU 2020–21
Community psychiatrist (hour)	155.53	PSSRU 2020–21
Community psychologist (hour)	66.95	PSSRU 2020–21
Dentist (hour)	108.15	PSSRU 2020–21
Mental health professional	57.6	PSSRU 2015–16
Counsellor (hour)	66.95	PSSRU 2020–21
Physiotherapist (hour)	55.62	PSSRU 2020–21
Other doctors (same as GP)	41.70	PSSRU 2020–21
Alternative therapist	57.6	PSSRU 2015–16
Optician (hour)	22.14	Gov.uk
NHS direct/NHS psychiatric help line	8.06	PSSRU 2015–16
Other professional	57.6	PSSRU 2015–16
Social worker (hour)	53.56	PSSRU 2020–21
Daycentre (per client attendance)	40.20	PSSRU 2020–21
Hospital contacts		
Mental health admission (bed-day)	445.98	PSSRU 2010
Day-case	799.38	Reference Cost Year 2018–9
Non-elective short stay (bed-days)	852.53	Reference Cost Year 2019–20
Non-elective long stay (bed-days)	3740.70	Reference Cost Year 2019–20
Excess day	314.51	National Tariff Year 2022–3
Hospital outpatient clinic – physical (attendance)	157.32	Reference Cost Year 2019–20
A&E – not admitted	143.50	Reference Cost Year 2019–20
Accident and emergency – admitted	191.34	Reference Cost Year 2019–20

Appendix 2 Amendments to protocol

TABLE 26 Amendments to protocol throughout the trial

Amendment	REC approval date	Summary
Substantial amendment 1	10 November 2016	Primary, secondary, tertiary and ancillary outcomes clarified; addition of FIBSER, PRISE, Maudsley VAS, DSCT, concomitant therapy questionnaire, and optional Bioresource hair and saliva collection; removal of physical activity questionnaire, pain VAS, and study ECGs; reporting of AEs reduced to SAEs only; quetiapine instant release added as an acceptable formulation; addition of week 4 study phone call; addition of option to split screening and baseline visit; removal of clinician medication log; removal of primary care collaborator.
Substantial amendment 2	10 January 2017	Prescribing by a trial clinician reduced from 8 weeks to the point of first prescription; recent test results allowed as pre-prescription safety tests; change of PIs at OHT, SLaM, and TEWV; addition of optional True Colours qualitative interview; addition of optional THINC-it assessment; clarified wording of outcome measures.
Substantial amendment 3	30 October 2017	Initial prescribing could take place in primary care with oversight from a trial clinician; change to definition of treatment resistance (minimum therapeutic trial of mirtazapine changed from 15 to 30 mg/day); change to known contraindication exclusion criteria (only *congenital* long QT syndrome is contraindicated); lithium/quetiapine use in current episode exclusion criteria changed to adequate treatment trials only; clarification regarding current antidepressant remaining unchanged for ≥ 6 weeks (applies to type, not dose); change of PI at TEWV; addition of CTQ; clarified wording of outcome measures; clarification regarding clinician judgement for current psychosis exclusion criteria.
Substantial amendment 5	27 September 2018	Addition of optional qualitative interview assessing patient views of trial medications; clarification of which measures are screening or baseline measures; improved wording of analysis section and SAE reporting.
Substantial amendment 6	4 February 2019	Improved wording regarding ancillary analysis publications, trial medication discontinuation monitoring, laboratory tests, and telephone or home research visit options; addition of recruitment method (re-contacting patients from other studies, with their consent).
Substantial amendment 8	14 October 2019	Adequate treatment trial could be reached any time before the 52-week study visit (rather than before week 8); therapeutic serum lithium level changed from 0.4–1.2 mmol/l to 0.6–1.2 mmol/l; clarification regarding randomisation stratification by site (including two new sites, AWP and SPFT).
Substantial amendment 9	30 September 2020	Reduction in power (from 90% to 80%) and sample size (from 276 to 214); CSRI updated to capture information relating to COVID-19; addition of the Maudsley Staging Model of Treatment Resistance, Modified Multi-Therapy resistance version at week 52; improved wording regarding initial prescription; change of trial manager and co-sponsor contact details.
Non-substantial amendment 28	22 November 2022	Change of trial manager and SLaM PI, sponsor contact details updated.

AE, adverse event; CTQ, childhood trauma questionnaire; FIBSER, Frequency, Intensity, and Burden of Side Effects scale; MSM-MTR, Maudsley Staging Model of Treatment Resistance, Modified Multi-Therapy Resistance Version; PI, principal investigator; REC, Research Ethics Committee; SAE, serious adverse event; THINC-it tool for Cognitive Dysfunction in Major Depressive Disorder; VAS, visual analogue scale.

Appendix 3 Tertiary and ancillary outcome measures

Tertiary outcomes:

- secondary outcomes analysed at 8 and 52 weeks will also be analysed at 26 weeks
- global severity (change in CGI-severity) from baseline to 8, 26 and 52 weeks
- global efficacy (CGI-efficacy score) at 8, 26 and 52 weeks
- side effects (frequency of individual items on the PRISE) at 8, 26 and 52 weeks
- side effects (FIBSER subscales: Frequency, Intensity, and Burden) at 8 and 52 weeks
- physical health changes (continuous blood parameters and waist circumference) from baseline to 8, 26 and 52 weeks
- satisfaction with lithium/quetiapine treatment at 8, 26 and 52 weeks as assessed using the four subscale scores on the Treatment Satisfaction Questionnaire for Medication (TSQM)
- manic symptoms (total Altman Mania Self Rating scale (AMSR) score) at baseline, 8, 26 and 52 weeks
- anxiety symptoms (total Generalised Anxiety Disorder Questionnaire (GAD-7) score) at baseline, 8, 26 and 52 weeks
- time to prescription, defined as the first date the participant is given a prescription for the treatment they were randomised to (proportion of participants given a prescription in each arm will also be summarised)
- baseline adherence to antidepressant treatment pre-trial (baseline MARS-5 using total MARS-5 scores and exploratory cut-offs [categorising participants as either adherent or non-adherent on the MARS-5)]
- cognition (total DSCT scores) at baseline, 8, 26 and 52 weeks
- adherence of clinicians to prescribing and monitoring guidelines for clinical practice as published and recommended (e.g. series of yes/no variables indicating whether recommended tests have been done)
- proportion of participants having an adequate treatment trial (as defined in *Primary outcomes*) by the 52-week visit
- number of hospital admissions for a depressive episode over 52 weeks
- personality measure [total Standard Assessment of Personality – Abbreviated Scale (SAPAS) score] from baseline to 8, 26 and 52 weeks
- work and social functioning (total WSAS score) AUC, measured weekly over 52 weeks.

Ancillary analyses:

- Economic cost over 52 weeks (CSRI, modified for TRD, measured at baseline, 8-, 26-, and 52-week visits). This included a comparison of NHS and PSS costs between treatment arms, as well as a comparison of societal costs (including costs to other statutory and non-statutory services, impacts on caregivers and families, days off work due to health problems for those in employment, and time spent by paid and unpaid caregivers in addition to NHS and PSS costs).
- Predictors of treatment response: baseline severity of treatment resistance (MSM), baseline depression characteristics [e.g. severity (HAMD-17), chronicity (MSM)] childhood trauma (CTQ), family history, recurrence (MINI 7.0), psychiatric comorbidity (MINI 7.0), subtype (e.g. typical vs. atypical, IDS-C), personality (SAPAS), type of antidepressant (SSRI vs. non SSRI, MTI), smoking status, alcohol use and illicit substance use at baseline, hypomanic screening (HCL-16 at baseline) and sociodemographic factors (e.g. sex, age, ethnicity, BMI).
- Exploration of longitudinal depression severity (assessed weekly via the QIDS-SR) until time to all cause and side effect treatment discontinuation, for lithium versus quetiapine, using varying definitions of treatment discontinuation, taking into account that patients may restart the trial medications after discontinuation (defined as not taking trial medications for at least 2 weeks).
- Analysis of biological samples for those who consented to provide blood/hair/saliva samples to the BioResource.
- Reliability and validity of the Maudsley Visual Analogue Scale (VAS) current and change measures compared to the QIDS-SR and MADRS at baseline, 8, 26 and 52 weeks.
- Discrepancy between the self-rated and clinician-rated 16-item Inventory of Depressive Symptomatology (IDS-C) at baseline and 8 weeks.
- Relationship between quetiapine and lithium serum levels, prescribed dose and depressive symptom severity and change from baseline to 8 and 52 weeks (if applicable) on the MADRS.

- Exploration of new interventions and amount of concomitant psychological treatment for depression (psychotropic medication and/or non-pharmacological) over 12 months (determined using questionnaires assessing new/concomitant treatments completed at baseline, 8-, 26- and 52-week study visits).
- Change in cognitive function (THINC-it composite and individual test scores).
- Patient rated experience of the True Colours weekly monitoring system, measured using qualitative interview at a follow-up appointment (either 8-, 26- or 52-week visit).
- Patient views and experiences of lithium and quetiapine, measured using qualitative interview (at, or after, 52-week visit).

Appendix 4 Schoenfeld residual plots testing proportional hazards assumptions

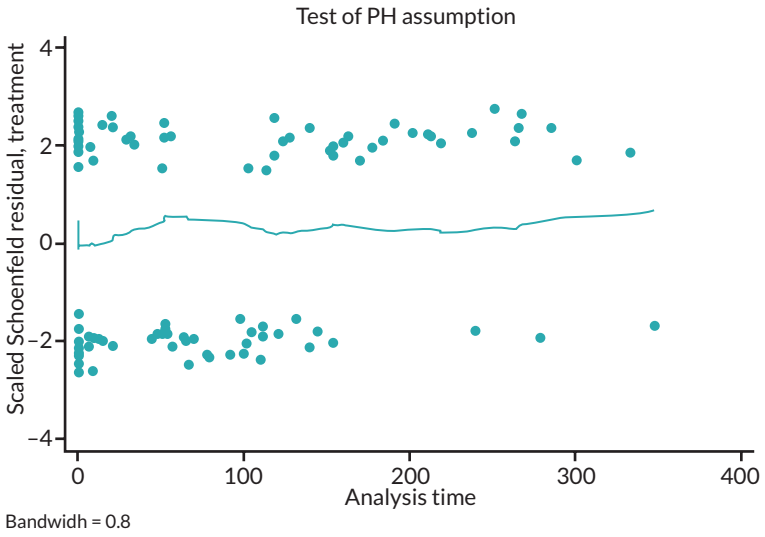

FIGURE 27 Schoenfeld residuals plot testing the assumption of proportional hazards for time to trial medication discontinuation outcome (ITT).

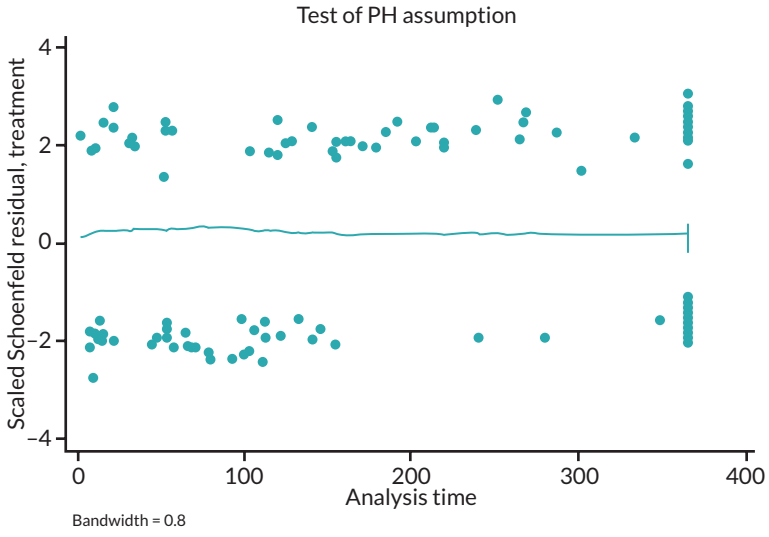

FIGURE 28 Schoenfeld residuals plot testing the assumption of proportional hazards for the time to initiation of trial medication outcome (ITT).

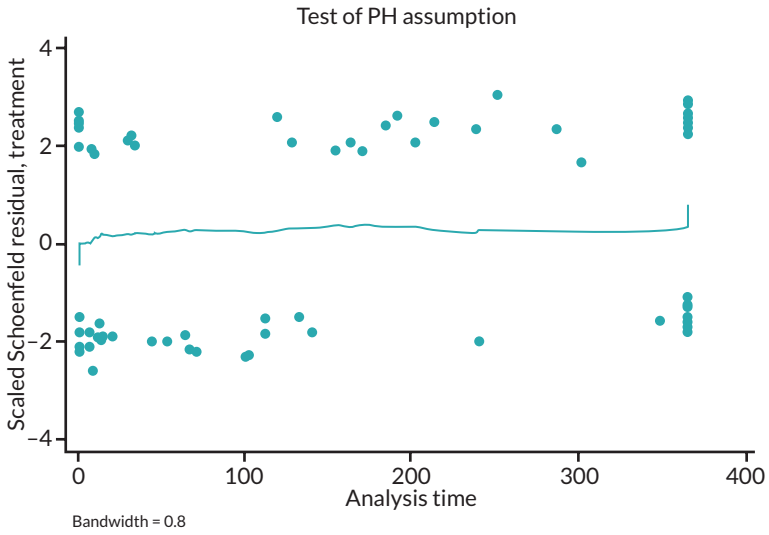

FIGURE 29 Schoenfeld residuals plot testing the assumption of proportional hazards for time to initiation of any new intervention for depression (ITT).

Appendix 5 Per-protocol clinical effectiveness analyses

For participants who were prescribed medication (see *Table 27*), in the quetiapine arm the mean (SD) QIDS-SR score was 17.4 (4.3) at baseline, which decreased to 12.4 (6.1) at week 52. In comparison the mean (SD) QIDS-SR at baseline in the lithium arm was slightly higher at 18.1 (4.3), decreasing to 13.8 (5.9) at week 52. *Figure 30* shows the unadjusted mean QIDS-SR scores, and 95% CIs, at study visits (baseline, week 8, week 26 and week 52), by trial arm for participants who were prescribed the trial medication. *Figure 31* shows the unadjusted mean QIDS-SR scores, and 95% CIs, collected at study visits and each week via True Colours by trial arm for participants who were prescribed the trial medication.

TABLE 27 Self-rated Quick Inventory of Depressive Symptomatology scores unadjusted descriptive statistics in the PP population

		Quetiapine	Lithium	Overall
		N = 95	*N* = 86	*N* = 181
QIDS-SR score at baseline	Mean (SD)	17.4 (4.3)	18.1 (4.3)	17.7 (4.3)
	Median (IQR)	17.0 (15.0–21.0)	18.0 (15.0–21.0)	18.0 (15.0–21.0)
QIDS-SR score at week 8	Mean (SD)	14.3 (5.5)	15.2 (5.5)	14.8 (5.5)
	Median (IQR)	14.0 (10.0–18.0)	16.0 (12.0–19.0)	15.0 (11.0–19.0)
QIDS-SR score at week 26	Mean (SD)	13.2 (5.6)	13.8 (6.0)	13.5 (5.8)
	Median (IQR)	13.0 (9.0–17.0)	13.0 (10.0–19.0)	13.0 (9.0–18.5)
QIDS-SR score at week 52	Mean (SD)	12.4 (6.1)	13.8 (5.9)	13.0 (6.0)
	Median (IQR)	12.0 (8.0–16.0)	14.0 (10.0–19.0)	13.0 (9.0–18.0)

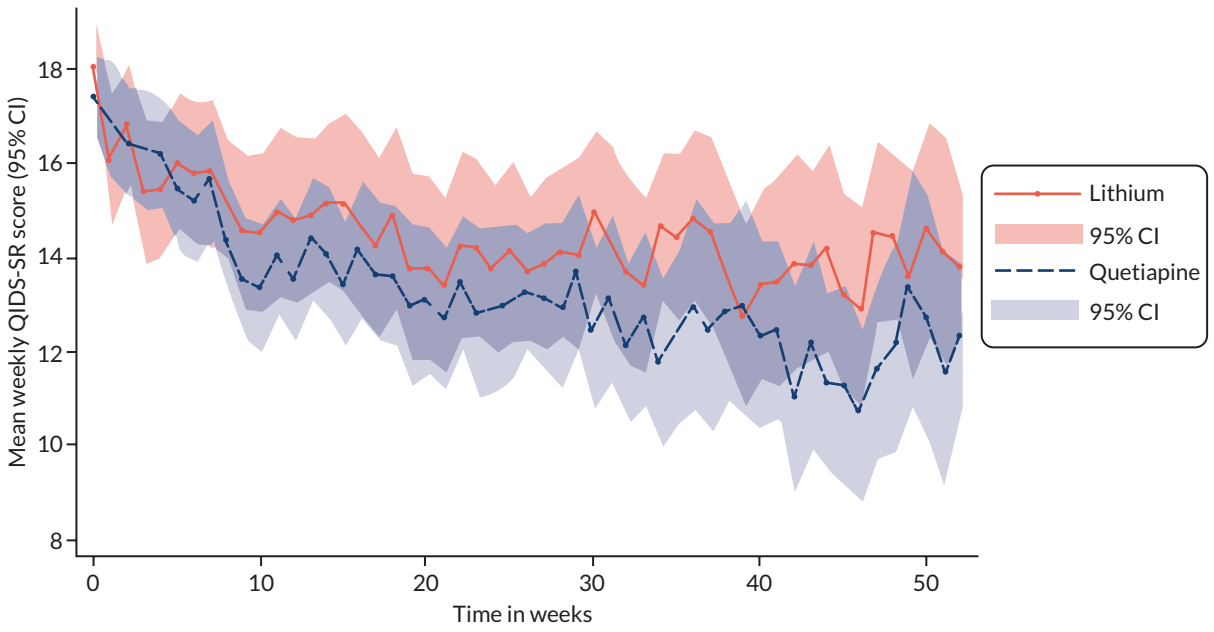

FIGURE 30 Unadjusted mean QIDS-SR scores at study visits, by arm (PP).

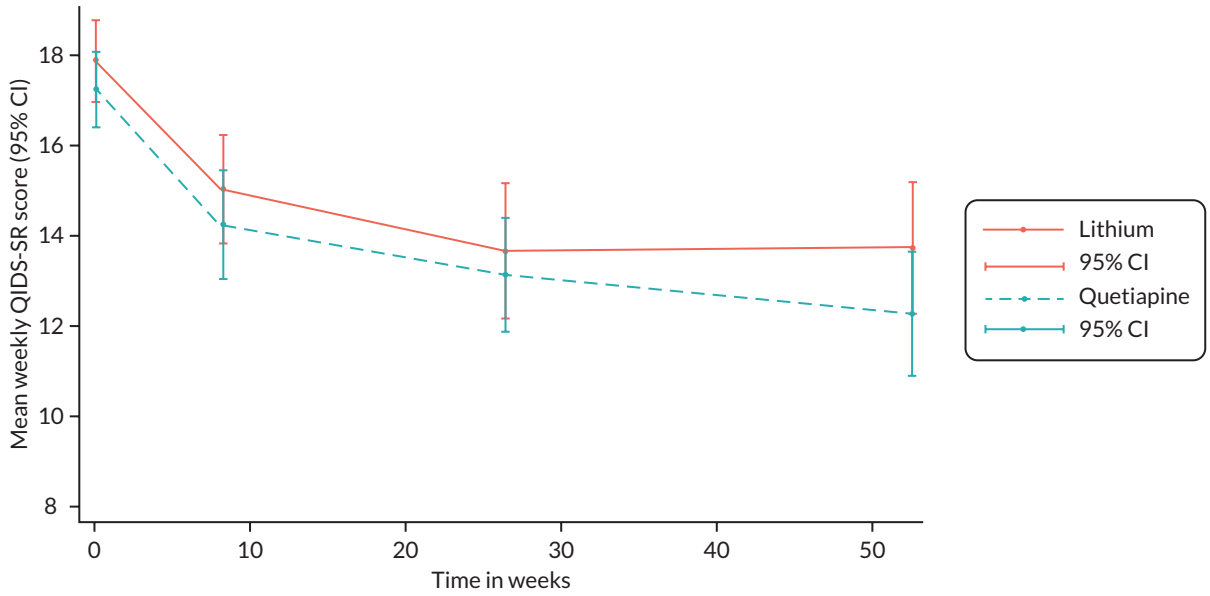

FIGURE 31 Unadjusted mean weekly QIDS-SR scores, by arm (PP).

TABLE 28 Between arm PP area under the QIDS-SR difference curve estimates

	Quetiapine vs. lithium difference in AUC	95% CI	*p*-value
Weekly QIDS-SR adjusted by stratification factors[a]	−73.72	(−138.47 to −8.97)	0.0256
Weekly QIDS-SR adjusted by stratification factors[a] and missingness predictors[b]	−72.66	(−137.03 to −8.29)	0.0269

a Stratification factors (TRD severity, depression severity and site).
b Missing predictors number of comorbidities and ethnicity.

These analyses used the same models as the ITT analyses but in the PP population, that is, individuals to whom randomised treatment was prescribed. As shown in *Table 28*, the model adjusted by baseline QIDS-SR score and the stratification factors yielded a significant negative quetiapine versus lithium area of −73.72 (95% CI: −138.47 to −8.97) points over 52 weeks ($p = 0.0256$), indicating significantly worse depression over the period in the lithium arm in the PP population. The full model that had also been adjusted by baseline variables that predicted missing 52-week QIDS-SR score yielded a significant negative quetiapine versus lithium area of −72.66 (95% CI: −137.03 to −8.29), again indicating significantly worse depression in the lithium arm ($p = 0.0269$). *Figure 32* shows the weekly observed mean QIDS-SR scores by arm, against the predicted QIDS-SR scores as predicted by the fully adjusted (PP) model (adjusted on baseline QIDS-SR score, TRD severity, depression severity, site, ethnicity and number of comorbidities). These results in the PP population were consistent with those in the ITT population.

The median (IQR) days to discontinuation for participants in the quetiapine arm was 365.0 (100.0–365.0), which is somewhat longer than those in the lithium arm, 301.0 (140.0–365.0), see *Table 29*. This difference is smaller than for the ITT analysis, as the PP analysis by its nature removed individuals defined as having short times to discontinuation (i.e. those not prescribed), and there were more such individuals in the lithium arm.

As shown in *Table 30*, participants in the quetiapine arm had 0.80 times the hazard of discontinuing (95% CI, 0.48 to 1.32) as compared to those in the lithium arm, which was not statistically significant ($p = 0.3866$). Similar to the ITT analysis, the Kaplan–Meier plot in *Figure 33* suggests that the proportional hazard assumption was violated, so we also present the RMST difference. Over a period of a year follow-up, participants in the quetiapine arm remained on the trial medication for 2.16 more days on average (−34.00 to 38.32) as compared to lithium, again this result was not statistically significant ($p = 0.9068$).

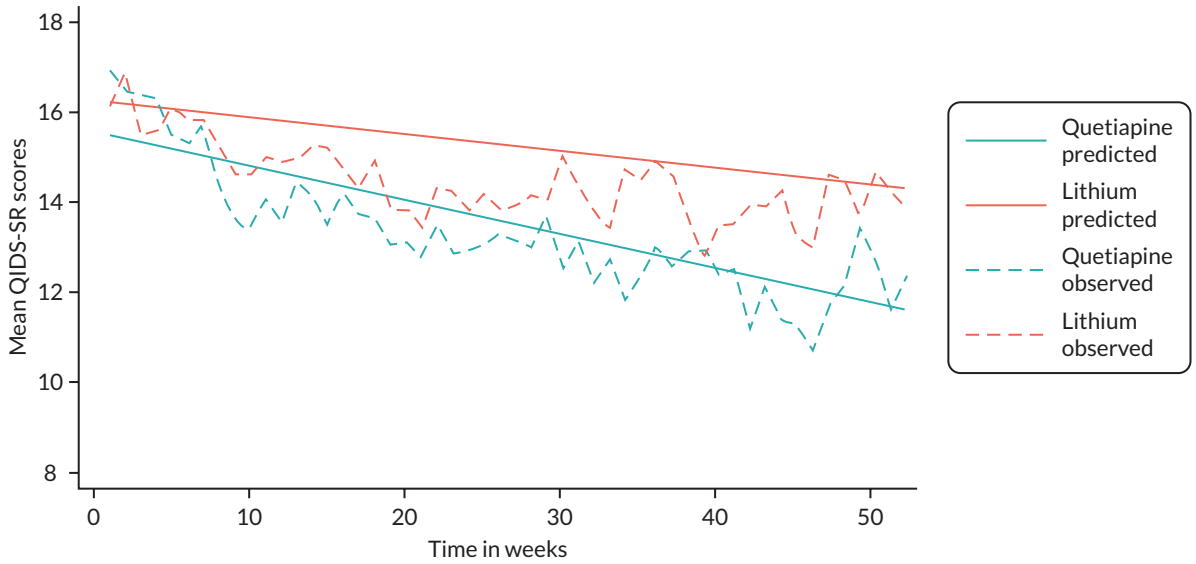

FIGURE 32 Weekly QIDS-SR scores, by arm, observed and predicted (PP).

TABLE 29 Time to discontinuation descriptive statistics (PP)

		Quetiapine, *N* = 95	Lithium, *N* = 85
Time to discontinuation	Mean days (95% CI)	256.81 (227.48 to 286.14)	246.74 (217.90–275.58)
	Median days (IQR)	365.0 (100.0 to 365.0)	301.0 (140.0–365.0)

TABLE 30 Between arm time to discontinuation comparisons (PP)

Time to discontinuation[a]	Hazard ratio (95% CI)	*p*-value	RMST difference (95% CI)	*p*-value
Quetiapine vs. lithium	0.80 (0.48 to 1.32)	0.3866	2.16 (−34.00 to 38.32)	0.9068
Lithium vs. quetiapine	1.25 (0.75 to 2.07)	0.3866		

a Adjusted by stratification factors (TRD severity, depression severity and site).

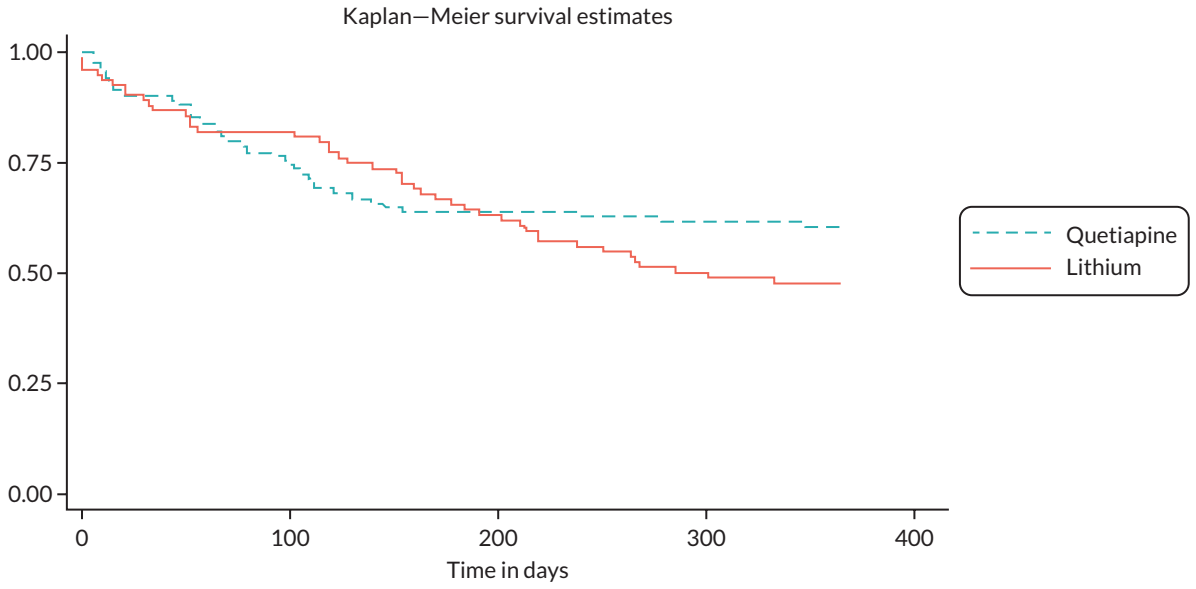

FIGURE 33 Time to discontinuation Kaplan–Meier curve (PP) by arm.

Appendix 6 Clinical effectiveness sensitivity analyses

Sensitivity analysis 1: therapeutic treatment trial

A minimum therapeutic treatment trial of the medication the participant was randomised to, for purposes of analyses, at or at any point up to 52-week follow-up visit, was defined as:

- A documented lithium serum level between 0.6 and 1.2 mmol/l[23,33] and participant reporting to have taken lithium for at least 4 weeks by the time of the 52-week study visit (determined via True Colours, treatment initiation and discontinuation forms and medical records).
- Quetiapine prescribed at ≥ 150 mg/day (determined via medical records/True Colours) for at least 4 weeks by the time of the 52-week study visit (determined via True Colours, treatment initiation and discontinuation forms and medical records).

This is as per *Section 5.2* of the Protocol v8.2. The estimate in the ITT population addresses the policy question of the effect of prescribing the medications when pragmatically some individuals will not be able to take them. In comparison, this sensitivity analysis addresses the effect of the medications in reducing depressive symptoms when taken as recommended, for a minimum therapeutic trial. For participants who fulfilled a therapeutic trial, in the quetiapine arm the mean (sd) QIDS-SR score was 17.4 (4.2) at baseline, which decreased to a mean (SD) of 11.9 (6.1) at the 52-week follow-up. In comparison, the mean (SD) QIDS-SR score at baseline in the lithium arm was 17.6 (4.1), which decreased to 13.5 (6.0) at week 52. In general, the unadjusted scores did not differ substantially between the arms, see *Table 31* and *Figure 34*.

As reported in *Table 32*, there was a quetiapine versus lithium area under the difference curve of −83.11 points over 52 weeks (95% CI: −161.86 to −4.36), which was statistically significant (p = 0.0386), and similar to the ITT estimate.

The median (IQR) days to discontinuation in the quetiapine arm in the therapeutic trial population was 365.0 (240.0–365.0) and in the lithium arm 365.0 (202.0–365.0), see *Table 33*. The inferential Cox regression and RMST results showed no significant differences between the arms in time to discontinuation in this population, which was consistent with the ITT result, see *Table 34*. *Figure 35* shows the Kaplan–Meier curve for time to discontinuation between the treatment arms over 365 days, for participants who were clinically assessed as having had a therapeutic trial.

TABLE 31 Mean weekly QIDS-SR scores (unadjusted) descriptive statistics for participants who received a therapeutic treatment trial

		Quetiapine	Lithium	Overall
		N = 71	*N* = 56	*N* = 127
QIDS-SR score at baseline	Mean (SD)	17.4 (4.2)	17.6 (4.1)	17.5 (4.1)
	Median (IQR)	17.0 (15.0–20.0)	18.0 (14.0–21.0)	17.0 (15.0–21.0)
QIDS-SR score at week 8	Mean (SD)	14.2 (5.3)	15.3 (5.2)	14.7 (5.2)
	Median (IQR)	14.0 (10.0–18.0)	15.0 (12.0–19.0)	15.0 (11.0–18.0)
QIDS-SR score at week 26	Mean (SD)	13.7 (5.4)	13.8 (6.6)	13.7 (5.9)
	Median (IQR)	13.0 (10.0–17.0)	14.0 (9.0–19.0)	13.5 (10.0–18.0)
QIDS-SR score at week 52	Mean (SD)	11.9 (6.1)	13.5 (6.0)	12.6 (6.0)
	Median (IQR)	11.5 (8.0–16.0)	12.7 (9.0–18.0)	12.0 (8.0–17.0)

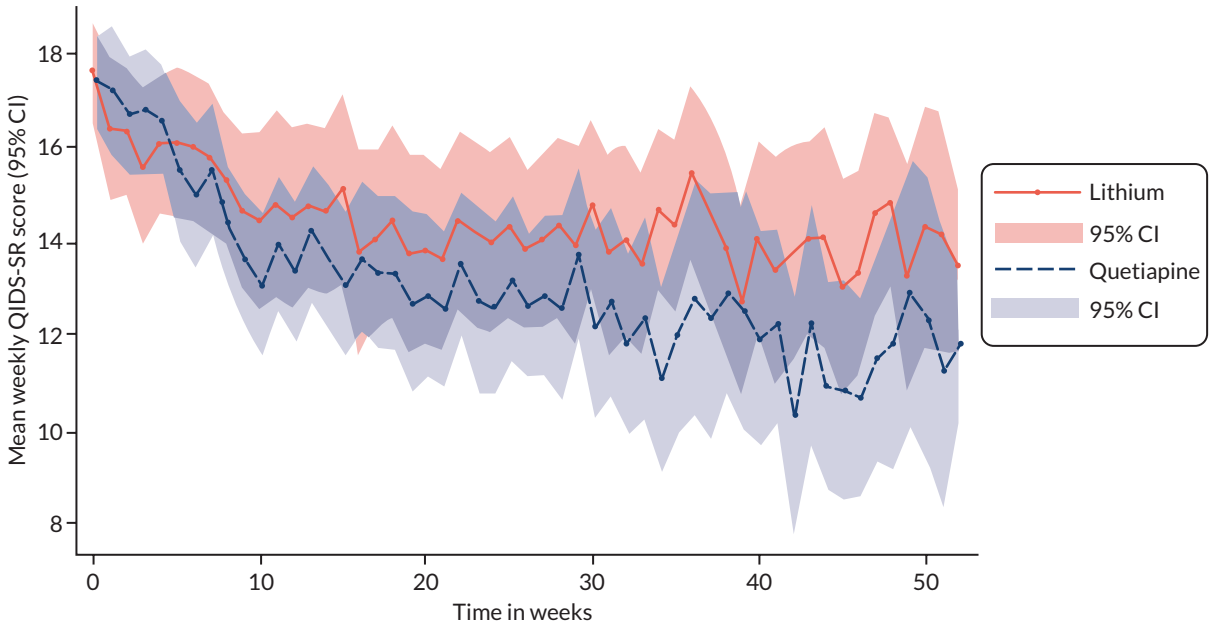

FIGURE 34 Mean weekly QIDS-SR scores (unadjusted) for participants who received a therapeutic treatment trial.

TABLE 32 Between arm area under the QIDS-SR difference curve estimate for participants who received a therapeutic treatment trial

	Quetiapine vs. lithium difference in AUC	95% CI	*p*-value
Weekly QIDs-SR adjusted by stratification factors[a] and missingness predictors[b]	−83.11	(−161.86 to −4.36)	0.0386

a Stratification factors (TRD severity, depression severity and site).
b Missing predictors by number of comorbidities and ethnicity.

TABLE 33 Time to discontinuation descriptive statistics for participants who received a therapeutic treatment trial

		Quetiapine, *N* = 71	Lithium, *N* = 55
Time to discontinuation	Mean days (95% CI)	296.89 (269.06 to 324.71)	294.16 (268.04 to 320.28)
	Median days (IQR)	365.0 (240.0–365.0)	365.0 (202.0–365.0)

TABLE 34 Between arm time to discontinuation comparisons for participants who received a therapeutic treatment trial

Time to discontinuation[a]	Hazard ratio (95% CI)	*p*-value	RMST difference (95% CI)	*p*-value
Quetiapine vs. lithium	0.73 (0.35 to 1.48)	0.3789	−4.71 (−31.81 to 22.38)	0.7332
Lithium vs. quetiapine	1.38 (0.67 to 2.82)	0.3789		

a Adjusted by stratification factors (TRD severity, depression severity and site).

Sensitivity analysis 2: adherence to trial medication

Given that the intervention being tested in the trial was pragmatic, that is, to prescribe the medication rather than directly for the medication to be taken, one sensitivity analysis of interest was to estimate effects in the subset who adhered to the medication. This analysis used the self-report treatment adherence MARS-5 measure. We re-estimated differences between arms for both primary outcomes for those having a score of 23 or greater (80% adherence) on the

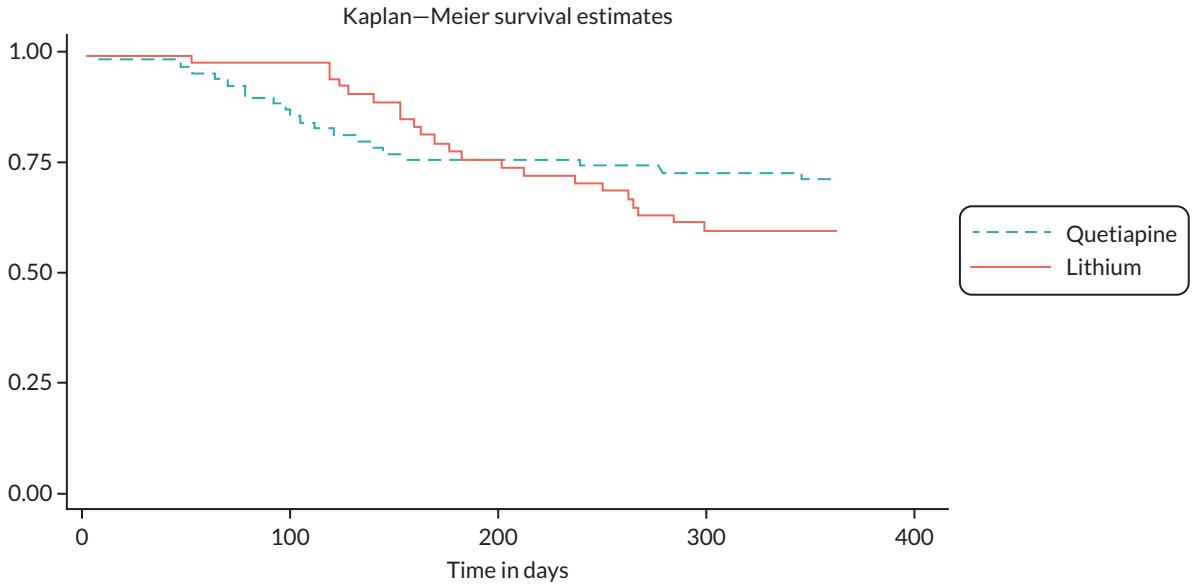

FIGURE 35 Time to discontinuation Kaplan–Meier curve by arm, for participants who had a therapeutic treatment trial.

first MARS-5 measure collected during the time the participant was on the trial medication. *Table 35* shows unadjusted QIDS-SR scores at study visits by treatment arm for participants who met the cut-off of having a score of 23 or greater (80% adherence), based on the MARS-5. In general, the unadjusted scores did not differ substantially between the arms. In the quetiapine arm the mean (SD) QIDS-SR score was 17.4 (4.0), which decreased to a mean (SD) of 12.1 (6.1) at the 52-week follow-up. In comparison the mean (SD) QIDS-SR score at baseline in the lithium arm was 17.9 (4.2), which decreased to 13.1 (6.0) at week 52. As shown in *Figure 36*, the QIDS-SR scores over 52 weeks between the arms show a similar pattern in this MARS adherence sensitivity analysis population as in the ITT population.

The inferential analysis is shown in *Table 36*. This yielded an area under the difference curve between quetiapine and lithium of −71.13 (95% CI: −145.18 to 2.91), which while not strictly statistically significant ($p = 0.0597$), was broadly consistent with the ITT estimate.

As shown in *Table 37*, the median (IQR) days to discontinuation in the quetiapine arm was 365.0 (154.0–365.0) and in the lithium arm 365.0 (166.5–365.0). The inferential Cox regression and RMST results (*Table 38*) show that participants in the quetiapine arm had 0.62 times the hazard of discontinuing (95% CI: 0.32 to 1.17) as compared to lithium participants, which was not statistically significant and was consistent with the ITT result. As for the analysis in the ITT population, the Kaplan–Meier plot in *Figure 37* suggests that the proportional hazard assumption was violated. Because of this we also present the RMST difference. Over a period of a year, participants in the quetiapine arm remained on the trial medication for 7.45 days longer on average (−22.24 to 37.15) as compared to lithium, again this result was not statistically significant ($p = 0.6227$).

Sensitivity analysis 3: initiation of trial medication

Another pre-specified sensitivity analysis was to estimate the between arm effects on the primary outcomes for those who initiated medication. Both the primary outcome analyses have been repeated dropping those participants who were not prescribed treatment, and/or did not record any date of first initiation of trial treatment. We note that the results of this analysis are very similar to those in the PP population, as this analysis in the population that initiated treatment only drops two additional people in the lithium arm who were prescribed medication but did not initiate it.

The descriptive summaries are shown in *Table 39* and *Figure 38*. For the participants who initiated treatment, in the quetiapine arm the mean (SD) QIDS-SR was 17.4 (4.3), which decreased to 12.4 (6.1) at the week 52 follow-up. The mean (SD) QIDS-SR at baseline was 18.0 (4.3) in the lithium arm and decreased to 13.8 (5.9) at week 52.

TABLE 35 Mean QIDS-SR scores descriptive statistics (unadjusted) for the MARS-5 medication adherent population

		Quetiapine	Lithium	Overall
		N = 73	*N* = 69	*N* = 142
QIDS-SR score at baseline	Mean (SD)	17.4 (4.0)	17.9 (4.2)	17.6 (4.1)
	Median (IQR)	17.0 (15.0–21.0)	18.0 (15.0–21.0)	18.0 (15.0–21.0)
QIDS-SR score at week 8	Mean (SD)	14.4 (5.4)	15.0 (5.5)	14.7 (5.4)
	Median (IQR)	14.5 (10.0–18.0)	16.0 (12.0–19.0)	15.0 (11.0–19.0)
QIDS-SR score at week 26	Mean (SD)	14.0 (5.5)	14.0 (6.6)	14.0 (6.0)
	Median (IQR)	13.5 (10.0–18.0)	15.0 (9.0–19.0)	14.0 (10.0–18.5)
QIDS-SR score at week 52	Mean (SD)	12.1 (6.1)	13.1 (6.0)	12.6 (6.0)
	Median (IQR)	12.0 (8.0–16.0)	11.7 (9.0–18.0)	12.0 (8.0–17.0)

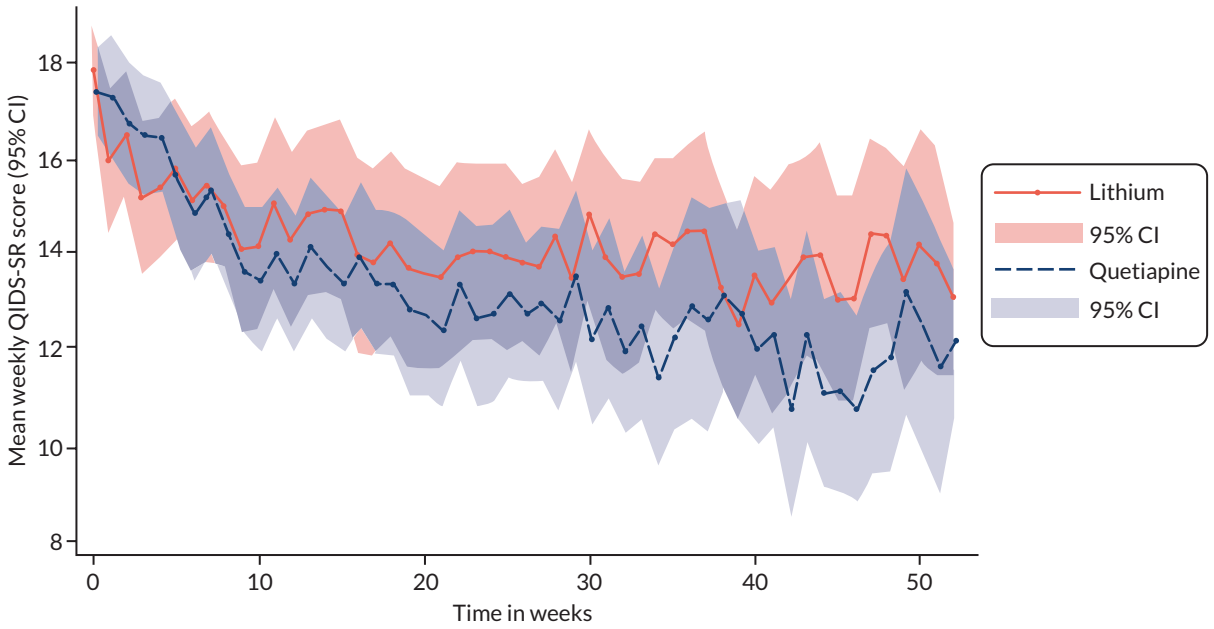

FIGURE 36 Mean weekly QIDS-SR scores (unadjusted) and 95% CIs for the MARS-5 medication adherent population.

TABLE 36 Between arm area under the QIDS-SR difference curve estimate in the MARS-5 medication adherent population

	Quetiapine vs. lithium difference in AUC	95% CI	*p*-value
Weekly QIDS-SR adjusted by stratification factors[a] and missingness predictors[b]	−71.13	(−145.18 to 2.91)	0.0597

a Stratification factors (TRD severity, depression severity and site).
b Missing predictors by number of comorbidities and ethnicity.

TABLE 37 Time to discontinuation descriptive statistics in the MARS-5 medication adherent population

		Quetiapine, *N* = 73	Lithium, *N* = 68
Time to discontinuation	Mean days (95% CI)	290.14 (260.58 to 319.70)	271.99 (244.03 to 299.94)
	Median days (IQR)	365.0 (154.0–365.0)	365.0 (166.5–365.0)

TABLE 38 Between arm time to discontinuation comparisons in the MARS-5 medication adherent population

Time to discontinuation[a]	Hazard ratio (95% CI)	*p*-value	RMST difference (95% CI)	*p*-value
Quetiapine vs. lithium	0.62 (0.32 to 1.17)	0.1384	7.45 (−22.24 to 37.15)	0.6227
Lithium vs. quetiapine	1.62 (0.86 to 3.08)	0.1384		

a Adjusted by stratification factors (TRD severity, depression severity and site).

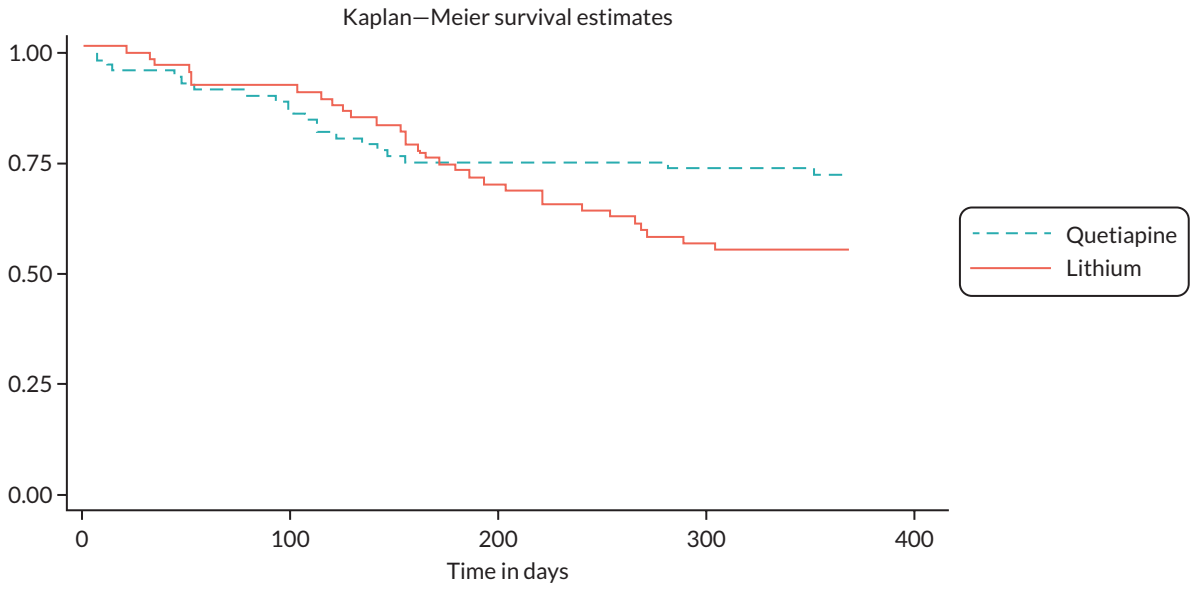

FIGURE 37 Time to discontinuation Kaplan–Meier curve by arm, in the MARS-5 medication adherent population.

TABLE 39 Mean QIDS-SR scores (unadjusted) descriptive statistics in those who initiated the trial medication

		Quetiapine	Lithium	Overall
		N = 95	*N* = 84	*N* = 179
QIDS-SR score at baseline	Mean (SD)	17.4 (4.3)	18.0 (4.3)	17.7 (4.3)
	Median (IQR)	17.0 (15.0–21.0)	18.0 (15.0–21.0)	18.0 (15.0–21.0)
QIDS-SR score at week 8	Mean (SD)	14.3 (5.5)	15.2 (5.5)	14.8 (5.5)
	Median (IQR)	14.0 (10.0–18.0)	16.0 (12.0–19.0)	15.0 (11.0–19.0)
QIDS-SR score at week 26	Mean (SD)	14.2 (5.5)	14.7 (6.8)	14.4 (6.1)
	Median (IQR)	14.0 (10.0–19.0)	16.0 (10.0–20.0)	15.0 (10.0–19.0)
QIDS-SR score at week 52	Mean (SD)	12.4 (6.1)	13.8 (5.9)	13.0 (6.0)
	Median (IQR)	12.0 (8.0–16.0)	14.0 (10.0–19.0)	13.0 (9.0–18.0)

The area under the between arm difference curve results are presented in *Table 40*. There was a significant negative quetiapine versus lithium area of −72.51 (95% CI: −137.21 to −7.80), indicating significantly worse depression in the lithium arm (p = 0.0281) in those who initiated treatment, which was consistent with the ITT result.

As shown in *Table 41*, the median (IQR) days to discontinuation for participants in the quetiapine arm was 365.0 (100.0–365.0), compared to 333.0 (152.0–365.0) for those in the lithium arm in the population who initiated the trial medication. Similar to the analysis in the PP population, these estimates are more similar as compared to the ITT

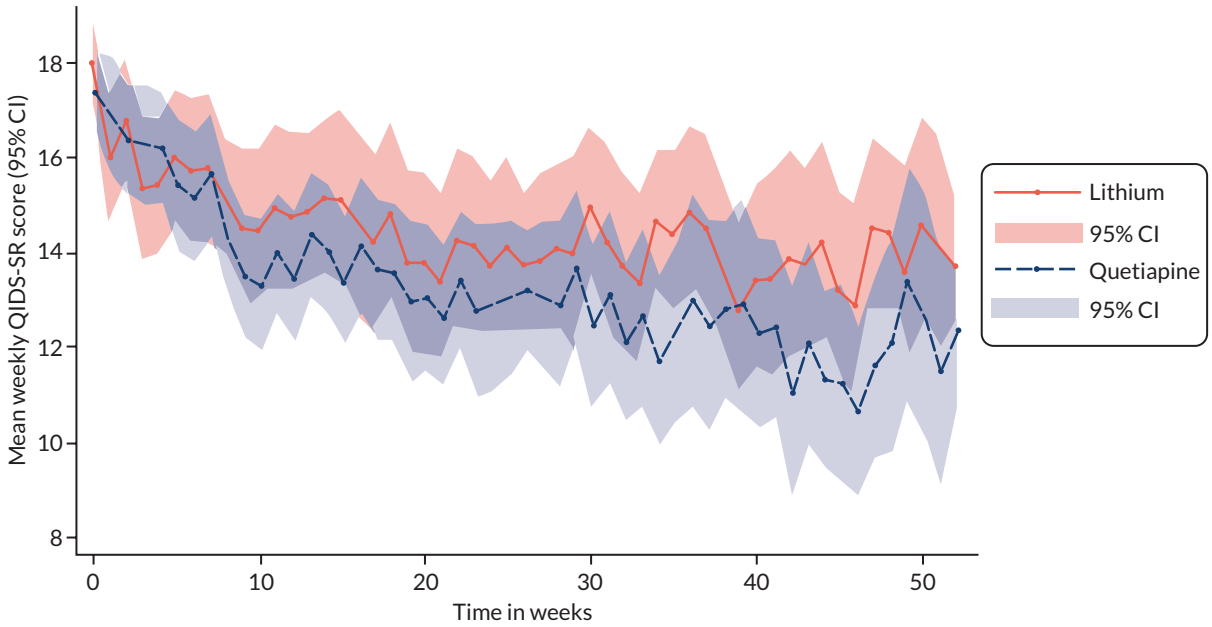

FIGURE 38 Mean weekly QIDS-SR scores (unadjusted) and 95% CIs in in those who initiated the trial medication.

TABLE 40 Between arm area under the QIDS-SR difference curve estimate in the population in those who initiated the trial medication

	Quetiapine vs. lithium difference in AUC	95% CI	*p*-value
Weekly QIDs-SR adjusted by stratification factors[a] and missingness predictors[b]	−72.51	(−137.21 to −7.80)	0.0281

a Stratification factors (TRD severity, depression severity and site).
b Missing predictors by number of comorbidities and ethnicity.

TABLE 41 Time to discontinuation descriptive statistics in those who initiated the trial medication

		Quetiapine, *N* = 95	Lithium, *N* = 83
Time to discontinuation	Mean days (95% CI)	256.81 (227.48 to 286.14)	252.67 (224.37 to 280.97)
	Median days (IQR)	365.0 (100.0–365.0)	333.0 (152.0–365.0)

population, as the population of those initiating treatment removes those who had short times to discontinuation (i.e. in this case both those not prescribed or initiated), and there are more such individuals in the lithium arm.

As shown in *Table 42*, participants in the quetiapine arm had 0.83 times the hazard of discontinuing (95% CI: 0.51 to 1.36) as compared to those in the lithium arm, which was not statistically significant, consistent with the ITT result. Similar to the analysis in the ITT population, the Kaplan–Meier plot in *Figure 39* suggests that the proportional hazard assumption was violated. Because of this the RMST difference is also presented. Over the 52-week period, participants in the quetiapine arm remained on the trial medication for 2.78 days longer on average (−33.04 to 38.59) as compared to lithium, this result was not statistically significant (p = 0.8793).

TABLE 42 Between arm time to discontinuation comparisons in those who initiated the trial medication

Time to discontinuation[a]	Hazard ratio (95% CI)	*p*-value	RMST difference (95% CI)	*p*-value
Quetiapine vs. lithium	0.83 (0.51 to 1.36)	0.4655	2.78 (−33.04 to 38.59)	0.8793
Lithium vs. quetiapine	1.20 (0.73 to 1.97)	0.4655		

a Adjusted by stratification factors (TRD severity, depression severity and site).

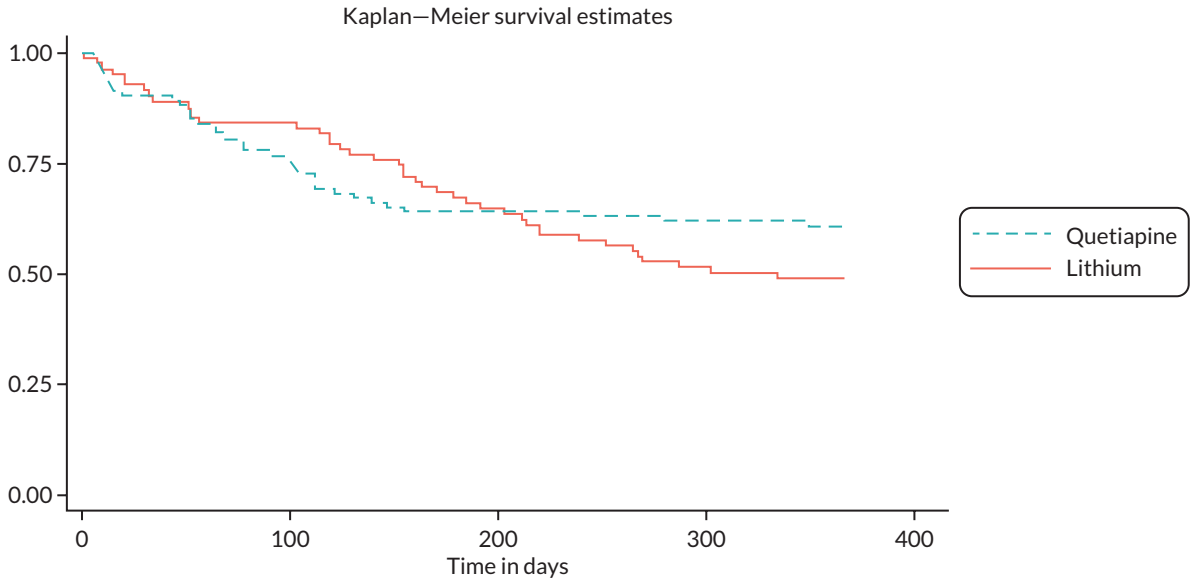

FIGURE 39 Time to discontinuation Kaplan–Meier curve by arm, for participants who initiated the trial medication.

Sensitivity analysis 4: missing at random assumption

We aimed to fit the models outlined in Section 3.1.5.4 of the SAP, but were not able to for technical reasons. We then tried to take the approach outlined in the trial protocol and Section 1.10 of the SAP (see *Report Supplementary Material 1*), using pattern mixture models to evaluate the effect of assuming systematically worse (increased) QIDS-SR for people who did not provide data. This would involve multiply imputing missing values, followed by applying systematically worse values of QIDS-SR (delta values[72,73]) across all 52 weekly measures, only for time points after participants had discontinued trial medication. However, this approach also posed technical problems.

For those with data at 52 weeks, participants in the quetiapine arm (n = 77, 72%) scored 1.79 points lower than those in the lithium arm (n = 68, 65%) on the QIDS-SR (95% CI: −3.57 to −0.02, p = 0.0473). *Figure 40* shows that this quetiapine versus lithium QIDS-SR effect at 52 weeks is quite robust to deviations from the MAR assumption applied in the main analysis. The effect becomes non-significant when worsening of 0.2 points or greater on the QIDS-SR is assumed, but only if this is assumed to occur solely in the quetiapine arm. The statistically significant quetiapine versus lithium effect at 52 weeks is robust to delta values across the range if these are assumed to apply to the lithium arm only, or to both arms.

Sensitivity analysis 5: impact of the COVID-19 pandemic

Descriptive summaries

In total 193 participants had baseline data collected and were randomised before 1 February 2020, see *Table 43*. 96 were randomised to quetiapine and 97 to lithium. There was no missing sociodemographic data at baseline. The mean (SD) age was 42.6 (13.8) years overall, this was balanced between arms with a mean (SD) of 41.7 (14.4) years in the

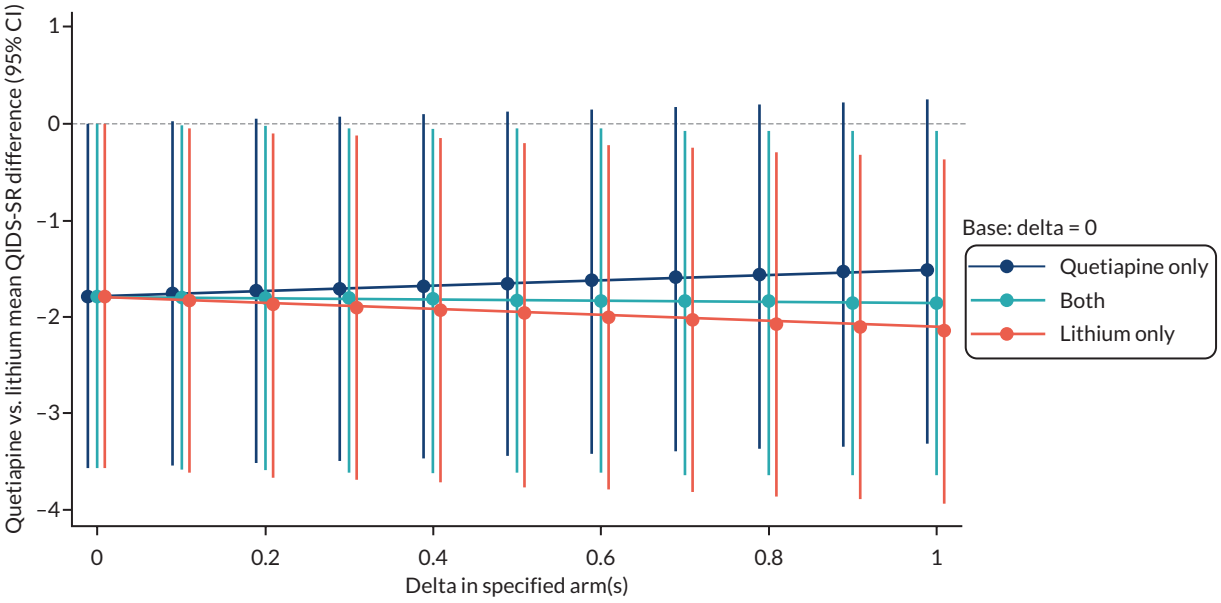

FIGURE 40 Data missing not at random at 52 weeks sensitivity analysis of quetiapine vs. lithium adjusted mean QIDS-SR difference.

TABLE 43 Sociodemographic characteristics at baseline pre COVID-19

		Quetiapine	Lithium	Overall
Total	*N*	96	97	193
Characteristics				
Age	Mean (SD)	41.7 (14.4)	43.4 (13.3)	42.6 (13.8)
	Median (IQR)	40.0 (29.0–52.0)	43.0 (33.0–53.0)	42.0 (31.0–53.0)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
		n (%)	*n* (%)	*n* (%)
Sex	Male	39 (40.6)	50 (51.5)	89 (46.1)
	Female	57 (59.4)	46 (47.4)	103 (53.4)
	Intersex	0 (0.0)	0 (0.0)	0 (0.0)
	Female to male	0 (0.0)	1 (1.0)	1 (0.5)
	Male to female	0 (0.0)	0 (0.0)	0 (0.0)
	Neither	0 (0.0)	0 (0.0)	0 (0.0)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
Ethnicity	White background	87 (90.6)	86 (88.7)	173 (89.6)
	Mixed/Multiple ethnic background	3 (3.1)	3 (3.1)	6 (3.1)
	Asian background	1 (1.0)	5 (5.2)	6 (3.1)
	Black/African/Caribbean background	2 (2.1)	2 (2.1)	4 (2.1)
	Any other background	2 (2.1)	1 (1.0)	3 (1.6)
	Unrecorded	1 (1.0)	0 (0.0)	1 (0.5)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)

TABLE 43 Sociodemographic characteristics at baseline pre COVID-19 (*continued*)

		Quetiapine	Lithium	Overall
Highest completed level of education	Primary education or less (no formal qualifications)	1 (1.0)	6 (6.2)	7 (3.6)
	Secondary education (GCSE, O Levels)	15 (15.6)	15 (15.5)	30 (15.5)
	College-level education or equivalent (A Level, NVQ, International Baccalaureate, BTEC nationals)	40 (41.7)	35 (36.1)	75 (38.9)
	Degree level education/diploma (e.g. BSc, BA)	24 (25.0)	30 (30.9)	54 (28.0)
	Post-graduate degree (e.g. MSc, MA, PhD)	16 (16.7)	11 (11.3)	27 (14.0)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
Current main employment status	Paid employment	47 (49.0)	45 (46.4)	92 (47.7)
	Unemployed	26 (27.1)	40 (41.2)	66 (34.2)
	Student	10 (10.4)	3 (3.1)	13 (6.7)
	Housewife/husband	1 (1.0)	2 (2.1)	3 (1.6)
	Retired	12 (12.5)	7 (7.2)	19 (9.8)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)

quetiapine arm and a mean (SD) of 43.4 (13.3) years in the lithium arm. Overall, there was a majority of females (53.4%), this was also the case for those randomised to quetiapine (59.4% female), however there was a majority of males in the lithium arm (51.5% male), that is, we see the same imbalance in this pre-COVID-19 subset as in the whole trial population. The majority of the sample were from a white background (89.6% overall), this was also reflected in both arms (90.6% in the quetiapine arm and 88.7% in the lithium arm).

Most participants had a college-level education or equivalent as the highest completed level of education (38.9% overall) and this was similar within the arms too (41.7% in the quetiapine arm; 36.1% in the lithium arm). The majority overall were in paid employment (47.7%), and this was similar between arms (49.0% in the quetiapine arm; 46.4% in the lithium arm). As with the complete data set, there was an imbalance between arms in the proportion of participants who were unemployed. Overall, 34.2% of participants were unemployed. However, there was a greater proportion of participants unemployed in the lithium arm (41.2%) compared to the quetiapine arm (27.1%).

Clinical characteristics at baseline for participants randomised before 1 February 2020 are shown in *Table 44*. There was no data missing for number of comorbidities (as recorded using the MINI v7.0), HAMD-17 score (an inclusion criteria measure) and severity of mental illness. All other measures had various degrees of missing data. Waist circumference was missing the most measures (23.3% overall with 24.0% in the quetiapine arm, and 22.7% in the lithium arm). Pulse rate, diastolic blood pressure and systolic blood pressure were missing similar proportions of data (14.5%, 13.0% and 12.4% overall, respectively). Weight was missing in 8.3% of the sample overall, however there was a larger proportion of participants in the quetiapine arm missing weight than in the lithium arm, (11.5%, compared to 5.2% in the lithium arm). All other measures had < 6% missing data overall with similar proportions missing between arms.

Mean (SD) QIDS-SR score was 17.4 (4.3) overall, which appeared to be relatively balanced between arms at baseline [17.0 (4.4) in the quetiapine arm, and 17.8 (4.1) in the lithium arm]. All other measures were balanced between the arms.

In total, 19 participants had baseline data collected and were randomised on or after 1 February 2020. Eleven were randomised to quetiapine and eight to lithium, see *Table 45*. There was no missing sociodemographic data collected at baseline. The mean (SD) age was 40.9 (15.5) years overall, this was balanced between the arms with a mean (SD) of 40.9

TABLE 44 Clinical characteristics at baseline pre COVID-19

Characteristics		Quetiapine, *N* = 96	Lithium, *N* = 97	Overall, *N* = 193
Number of comorbidities	Mean (SD)	1.8 (1.7)	1.9 (1.6)	1.8 (1.6)
	Median (IQR)	2.0 (0.0–3.0)	2.0 (1.0–3.0)	2.0 (0.0–3.0)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
HAM-17 score	Mean (SD)	21.3 (5.1)	22.2 (5.4)	21.8 (5.3)
	Median (IQR)	21.0 (18.0–24.0)	21.0 (18.0–26.0)	21.0 (18.0–25.0)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
Weight (kg)	Mean (SD)	88.2 (22.8)	88.4 (22.1)	88.3 (22.4)
	Median (IQR)	86.8 (73.0–98.0)	88.8 (71.0–102.3)	87.7 (71.9–101.5)
	Missing, *n* (%)	11 (11.5)	5 (5.2)	16 (8.3)
Waist circumference (cm)	Mean (SD)	99.4 (17.3)	101.5 (22.2)	100.5 (19.9)
	Median (IQR)	99.0 (87.5–110.0)	99.0 (88.0–113.3)	99.0 (87.8–111.5)
	Missing, *n* (%)	23 (24.0)	22 (22.7)	45 (23.3)
Systolic blood pressure (mmHg)	Mean (SD)	130.5 (16.1)	131.1 (17.2)	130.8 (16.6)
	Median (IQR)	130.0 (119.0–141.0)	128.0 (120.0–144.0)	129.0 (120.0–141.0)
	Missing, *n* (%)	11 (11.5)	13 (13.4)	24 (12.4)
Diastolic blood pressure (mmHg)	Mean (SD)	77.8 (11.0)	80.2 (11.4)	79.0 (11.2)
	Median (IQR)	78.5 (69.5–83.5)	80.0 (73.0–86.5)	79.5 (71.0–86.0)
	Missing, *n* (%)	12 (12.5)	13 (13.4)	25 (13.0)
Pulse rate (bpm)	Mean (SD)	76.8 (15.3)	73.9 (10.7)	75.4 (13.3)
	Median (IQR)	74.0 (65.0–85.0)	74.0 (67.0–80.0)	74.0 (66.0–82.0)
	Missing, *n* (%)	13 (13.5)	15 (15.5)	28 (14.5)
MARS-5 score	Mean (SD)	23.5 (1.9)	23.2 (2.1)	23.3 (2.0)
	Median (IQR)	24.0 (23.0–25.0)	24.0 (22.0–25.0)	24.0 (23.0–25.0)
	Missing, *n* (%)	3 (3.1)	5 (5.2)	8 (4.1)
QIDS-SR score	Mean (SD)	17.0 (4.4)	17.8 (4.1)	17.4 (4.3)
	Median (IQR)	17.0 (14.5–21.0)	18.0 (15.0–21.0)	17.0 (15.0–21.0)
	Missing, *n* (%)	0 (0.0)	1 (1.0)	1 (0.5)
WSAS score	Mean (SD)	27.0 (7.2)	28.1 (7.9)	27.5 (7.6)
	Median (IQR)	28.0 (23.0–31.0)	28.9 (22.5–34.0)	28.0 (23.0–33.0)
	Missing, *n* (%)	1 (1.0)	3 (3.1)	4 (2.1)
HCL-16 score	Mean (SD)	6.5 (3.6)	6.8 (3.9)	6.7 (3.8)
	Median (IQR)	7.0 (4.0–9.0)	7.0 (4.0–10.0)	7.0 (4.0–10.0)
	Missing, *n* (%)	4 (4.2)	7 (7.2)	11 (5.7)
MADRS score	Mean (SD)	30.7 (6.3)	31.8 (7.3)	31.2 (6.8)
	Median (IQR)	31.0 (27.0–35.0)	32.0 (26.0–36.0)	31.0 (26.0–36.0)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)

TABLE 44 Clinical characteristics at baseline pre COVID-19 (*continued*)

Characteristics		Quetiapine, *N* = 96	Lithium, *N* = 97	Overall, *N* = 193
GAD-7	Mean (SD)	12.5 (5.5)	12.5 (5.7)	12.5 (5.6)
	Median (IQR)	13.0 (8.0–17.0)	13.0 (8.0–18.0)	13.0 (8.0–17.0)
	Missing, *n* (%)	0 (0.0)	2 (2.1)	2 (1.0)
		n (%)	*n* (%)	*n* (%)
Severity of mental illness	Normal, not at all ill	1 (1.0)	0 (0.0)	1 (0.5)
	Borderline mentally ill	0 (0.0)	0 (0.0)	0 (0.0)
	Mildly ill	10 (10.4)	8 (8.2)	18 (9.3)
	Moderately ill	48 (50.0)	49 (50.5)	97 (50.3)
	Markedly ill	29 (30.2)	30 (30.9)	59 (30.6)
	Severely ill	8 (8.3)	10 (10.3)	18 (9.3)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)

TABLE 45 Sociodemographic characteristics at baseline post COVID-19

		Quetiapine	Lithium	Overall
Total	*N*	11	8	19
Characteristics				
Age	Mean (SD)	40.9 (17.0)	41.0 (14.4)	40.9 (15.5)
	Median (IQR)	37.0 (24.0–56.0)	36.0 (31.5–56.0)	37.0 (26.0–56.0)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
		n (%)	*n* (%)	*n* (%)
Sex	Male	3 (27.3)	4 (50.0)	7 (36.8)
	Female	8 (72.7)	4 (50.0)	12 (63.2)
	Intersex	0 (0.0)	0 (0.0)	0 (0.0)
	Female to male	0 (0.0)	0 (0.0)	0 (0.0)
	Male to female	0 (0.0)	0 (0.0)	0 (0.0)
	Neither	0 (0.0)	0 (0.0)	0 (0.0)
	Missing	0 (0.0)	0 (0.0)	0 (0.0)
Ethnicity	White background	9 (81.8)	6 (75.0)	15 (78.9)
	Mixed/Multiple ethnic background	1 (9.1)	0 (0.0)	1 (5.3)
	Asian background	1 (9.1)	2 (25.0)	3 (15.8)
	Black/African/Caribbean background	0 (0.0)	0 (0.0)	0 (0.0)
	Any other background	0 (0.0)	0 (0.0)	0 (0.0)
	Unrecorded	0 (0.0)	0 (0.0)	0 (0.0)
	Missing	0 (0.0)	0 (0.0)	0 (0.0)

continued

TABLE 45 Sociodemographic characteristics at baseline post COVID-19 (*continued*)

		Quetiapine	Lithium	Overall
Highest completed level of education	Primary education or less (no formal qualifications)	0 (0.0)	0 (0.0)	0 (0.0)
	Secondary education (GCSE, O Levels)	1 (9.1)	0 (0.0)	1 (5.3)
	College-level education or equivalent (A Level, NVQ, International Baccalaureate, BTEC nationals)	4 (36.4)	2 (25.0)	6 (31.6)
	Degree level education/diploma (e.g. BSc, BA)	4 (36.4)	1 (12.5)	5 (26.3)
	Post-graduate degree (e.g. MSc, MA, PhD)	2 (18.2)	5 (62.5)	7 (36.8)
	Missing	0 (0.0)	0 (0.0)	0 (0.0)
Current main employment status?	Paid employment	7 (63.6)	4 (50.0)	11 (57.9)
	Unemployed	3 (27.3)	3 (37.5)	6 (31.6)
	Student	0 (0.0)	1 (12.5)	1 (5.3)
	Housewife/husband	0 (0.0)	0 (0.0)	0 (0.0)
	Retired	1 (9.1)	0 (0.0)	1 (5.3)
	Missing	0 (0.0)	0 (0.0)	0 (0.0)

(17.0) years in the quetiapine arm and a mean (SD) of 41.0 (14.4) years in the lithium arm. Overall, there was a majority of females (63.2%). This was also the case for those randomised to quetiapine (72.7% female), while 50.0% of those to lithium were female. The majority of the sample were from a white background (78.9% overall), this was also reflected in the two arms (81.8% in the quetiapine arm and 75.0% in the lithium arm). There were several apparent imbalances between the arms, however, we note this is a very small sample and so imbalances might be expected.

Clinical characteristics are shown in *Table 46*. There was no clinical characteristic data missing for number of comorbidities (as recorded using the MINI v7.0), HAM-17 score (an inclusion criteria measure), MARS-5 score, QIDS-SR score, WSAS score, HCL-16 score, GAD-7 score, and the severity of mental illness. Waist circumference was missing the most measures (47.4% overall with 45.5% in the quetiapine arm, and 50.0% in the lithium arm). Pulse rate, diastolic blood pressure and systolic blood pressure are missing the same proportions of data, (26.3% overall, respectively). Weight was missing in 21.1% of the sample overall, however there is larger proportion of participants in the quetiapine arm missing weight than in the lithium arm, (27.3%, compared to 12.5% in the lithium arm).

While keeping the small sample size in mind, there are a few points to note regarding clinical characteristics of the sample. QIDS-SR and HCL scores were somewhat higher overall/in both arms than in the pre-COVID-19 subset. It also appears that participants in the lithium arm may have scored worse on some depression and anxiety-related measures (QIDS-SR, HCL-16, GAD-7), however for other measures this is reversed (MADRS and WSAS). There was a mean (SD) QIDS-SR score of 18.8 (3.6) overall. The mean (SD) was 18.5 (3.8) in the quetiapine arm, however it was slightly higher at baseline 19.3 (3.5) in the lithium arm. The HCL-16 score was 7.4 (3.8) in the quetiapine arm and 9.1 (2.6) in the lithium arm, and GAD-7 was 11.5 (6.6) in the quetiapine arm and 14.1 (4.9) in the lithium arm. In terms of WSAS and MADRS, which were lower/better in the lithium arm, the mean (SD) WSAS score was 28.7 (6.6) in the quetiapine arm and 25.4 (6.7) in the lithium arm, with the MADRS score being 31.4 (7.6) in the quetiapine arm and 28.0 (4.0) in the lithium arm. It is unclear whether the magnitudes of these differences are likely to be clinically significant, especially given the small sample.

Self-rated Quick Inventory of Depressive Symptomatology descriptive summaries for participants for whom their measures were collected before 1 March 2020 (pre-COVID-19) are shown in *Table 47*. In the quetiapine arm the mean (SD) QIDS-SR score was 14.3 (5.7) at 8 weeks, which decreased to a mean (SD) of 12.7 (6.6) at the 52-week follow-up.

TABLE 46 Clinical characteristics at baseline post COVID-19

Characteristics		Quetiapine, *N* = 11	Lithium, *N* = 8	Overall, *N* = 19
Number of comorbidities	Mean (SD)	1.6 (1.6)	2.0 (1.2)	1.8 (1.4)
	Median (IQR)	1.0 (0.0–3.0)	1.5 (1.0–3.0)	1.0 (1.0–3.0)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
HAM-17 score	Mean (SD)	19.5 (4.4)	19.1 (3.3)	19.3 (3.9)
	Median (IQR)	17.0 (17.0–23.0)	18.5 (17.0–20.5)	18.0 (17.0–21.0)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
Weight (kg)	Mean (SD)	90.3 (18.9)	87.6 (23.5)	89.0 (20.4)
	Median (IQR)	89.4 (79.5–105.0)	85.7 (70.0–112.7)	85.7 (77.0–110.3)
	Missing, *n* (%)	3 (27.3)	1 (12.5)	4 (21.1)
Waist circumference (cm)	Mean (SD)	97.7 (16.5)	98.8 (20.1)	98.1 (16.9)
	Median (IQR)	97.0 (81.0–114.0)	100.5 (82.0–115.5)	97.0 (81.0–114.0)
	Missing, *n* (%)	5 (45.5)	4 (50.0)	9 (47.4)
Systolic blood pressure (mmHg)	Mean (SD)	137.7 (23.9)	125.9 (17.5)	131.8 (21.0)
	Median (IQR)	135.0 (118.0–148.0)	130.0 (117.0–141.0)	130.5 (118.0–142.0)
	Missing, *n* (%)	4 (36.4)	1 (12.5)	5 (26.3)
Diastolic blood pressure (mmHg)	Mean (SD)	77.6 (9.9)	73.7 (10.2)	75.6 (9.9)
	Median (IQR)	81.0 (71.0–85.0)	70.0 (66.0–84.0)	75.5 (67.0–84.0)
	Missing, *n* (%)	4 (36.4)	1 (12.5)	5 (26.3)
Pulse rate (bpm)	Mean (SD)	72.0 (14.7)	79.6 (24.0)	75.8 (19.5)
	Median (IQR)	72.0 (58.0–85.0)	88.0 (57.0–94.0)	74.0 (58.0–92.0)
	Missing, *n* (%)	4 (36.4)	1 (12.5)	5 (26.3)
MARS-5 score	Mean (SD)	24.4 (0.7)	23.4 (1.7)	23.9 (1.3)
	Median (IQR)	24.0 (24.0–25.0)	24.0 (22.5–24.5)	24.0 (24.0–25.0)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
QIDS-SR score	Mean (SD)	18.5 (3.8)	19.3 (3.5)	18.8 (3.6)
	Median (IQR)	19.0 (17.0–21.0)	20.5 (17.0–21.0)	20.0 (17.0–21.0)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
WSAS score	Mean (SD)	28.7 (6.6)	25.4 (6.7)	27.3 (6.7)
	Median (IQR)	30.0 (25.0–35.0)	25.0 (23.0–31.0)	26.0 (24.0–32.0)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
HCL-16 score	Mean (SD)	7.4 (3.8)	9.1 (2.6)	8.1 (3.4)
	Median (IQR)	8.0 (6.0–10.0)	9.0 (8.0–11.3)	9.0 (6.4–10.7)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
MADRS score	Mean (SD)	31.4 (7.6)	28.0 (4.0)	30.1 (6.5)

continued

TABLE 46 Clinical characteristics at baseline post COVID-19 (*continued*)

Characteristics		Quetiapine, *N* = 11	Lithium, *N* = 8	Overall, *N* = 19
	Median (IQR)	32.0 (24.0–37.0)	28.0 (26.0–30.0)	28.0 (26.0–35.0)
	Missing, *n* (%)	0 (0.0)	1 (12.5)	1 (5.3)
GAD-7	Mean (SD)	11.5 (6.6)	14.1 (4.9)	12.6 (5.9)
	Median (IQR)	12.0 (5.0–18.0)	14.0 (12.0–17.5)	14.0 (7.0–18.0)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
		n (%)	*n* (%)	*n* (%)
Severity of mental illness	Normal, not at all ill	0 (0.0)	0 (0.0)	0 (0.0)
	Borderline mentally ill	0 (0.0)	0 (0.0)	0 (0.0)
	Mildly ill	0 (0.0)	1 (12.5)	1 (5.3)
	Moderately ill	7 (63.6)	6 (75.0)	13 (68.4)
	Markedly ill	2 (18.2)	1 (12.5)	3 (15.8)
	Severely ill	2 (18.2)	0 (0.0)	2 (10.5)
	Missing, n(%)	0 (0.0)	0 (0.0)	0 (0.0)

TABLE 47 Self-rated Quick Inventory of Depressive Symptomatology scores where the visit took place pre COVID-19 measures

		Quetiapine	Lithium	Overall
QIDS-SR score at week 8	*N*	95	96	191
	Mean (SD)	14.3 (5.7)	15.2 (5.8)	14.7 (5.7)
	Median (IQR)	14.0 (9.0–18.0)	16.0 (11.0–20.0)	15.0 (11.0–19.0)
	Missing, *n* (%)	17 (17.9)	14 (14.6)	31 (16.2)
QIDS-SR score at week 26	*N*	82	86	168
	Mean (SD)	13.4 (5.8)	14.8 (6.0)	14.0 (5.9)
	Median (IQR)	13.0 (9.5–17.5)	16.5 (11.0–19.5)	14.0 (11.0–19.0)
	Missing, *n* (%)	18 (22.0)	34 (39.5)	52 (31.0)
QIDS-SR score at week 52	*N*	64	63	127
	Mean (SD)	12.7 (6.6)	14.6 (5.9)	13.5 (6.3)
	Median (IQR)	13.0 (8.0–18.0)	15.0 (10.0–18.5)	13.0 (9.0–18.0)
	Missing, *n* (%)	15 (23.4)	23 (36.5)	38 (29.9)

In comparison, the mean (SD) QIDS-SR score at 8 weeks was approximately one point higher in the lithium arm at 15.2 (5.8) and decreased to a mean (SD) of 14.6 (5.9) in the lithium arm. These values are broadly similar to those in the whole trial population.

Self-rated Quick Inventory of Depressive Symptomatology descriptive summaries for participants for whom measures were collected on or after 1 March 2020 (post-COVID) are shown in *Table 48*. In the quetiapine arm the mean (SD) QIDS-SR was 15.4 (3.5) at 8 weeks, which decreased to mean (SD) 11.2 (5.3) at the 52-week follow-up. In comparison, the mean (SD) QIDS-SR at 8 weeks was lower in the lithium arm at 14.2 (5.6), with the mean (SD) decreasing to a of

TABLE 48 Self-rated Quick Inventory of Depressive Symptomatology scores where the visit took place post COVID-19 measures

		Quetiapine	Lithium	Overall
QIDS-SR score at week 8	*N*	12	9	21
	Mean (SD)	15.4 (3.5)	14.2 (5.6)	14.9 (4.3)
	Median (IQR)	16.0 (14.5–17.0)	13.5 (9.0–19.0)	15.5 (12.0–17.0)
	Missing, *n* (%)	4 (33.3)	3 (33.3)	7 (33.3)
QIDS-SR score at week 26	*N*	25	19	44
	Mean (SD)	11.3 (5.7)	11.2 (5.9)	11.3 (5.7)
	Median (IQR)	9.0 (7.5–16.5)	9.5 (7.0–15.5)	9.0 (7.5–16.5)
	Missing, *n* (%)	9 (36.0)	7 (36.8)	16 (36.4)
QIDS-SR score at week 52	*N*	43	42	85
	Mean (SD)	11.2 (5.3)	13.4 (6.5)	12.3 (6.0)
	Median (IQR)	11.0 (7.5–16.0)	12.5 (8.0–19.5)	11.0 (8.0–17.0)
	Missing, *n* (%)	15 (34.9)	14 (33.3)	29 (34.1)

13.4 (6.5) at 52 weeks. These values were similar to or even somewhat lower than in the whole trial population. There was not appreciably more missing QIDS-SR data post-1 March as compared to pre, so we did not perform multiple imputation as per the SAP section 3.4.5 (see *Report Supplementary Material 1*).

Discontinuation and initiation data for participants with a discontinuation date before and on or after 1 March 2020 are shown in *Table 49*. There were 81 participants in the quetiapine arm and 80 participants in the lithium arm who had a discontinuation date before 1 March 2020. For these participants, the median (IQR) time to discontinuation was 240.0 (53.0–365.0) days for those in the quetiapine arm, compared to a median (IQR) of 170.0 (0.5–365.0) days in the lithium arm. Participants in the quetiapine arm have a longer average time to discontinuation than the lithium arm, as was the case for the full sample. The reason the 25th percentile of the time to discontinuation before 1 March 2020 in the lithium arm is 0.5 days and is smaller than the 25th percentile in the overall trial population is because the majority of participants who were assigned a discontinuation time of 0.5 (were not prescribed or didn't initiate) had a date of discontinuation before 1 March 2020. Only one participant with a time to discontinuation time of 0.5 days had their discontinuation date on or after 1 March 2020. There is one participant who was randomised more than a year before 1 March 2020 for whom we are missing discontinuation data, so their date of discontinuation could not have been after 1 March/even if they had not discontinued their year-long follow-up would precede this. For this reason, they were included in the group with discontinuation dates pre 1 March where possible/applicable.

There were 26 participants in the quetiapine arm and 25 participants in the lithium arm who had a discontinuation date on or after March 2020. The median (IQR) days to discontinuation was 365.0 (365.0–365.0) in the quetiapine arm and 365.0 (163.0–365.0) in the lithium arm, which was longer than participants with discontinuation dates pre-COVID. The proportions of participants who were prescribed and initiated treatment and had discontinuation dates post-1 March were larger than those with dates pre-1 March: 96.2% in the quetiapine arm and 96.0% in the lithium arm (initiation post-1 March), compared to 86.4% in the quetiapine arm and 75.0% in the lithium arm (initiation post-1 March). This suggests no major issues with prescription and initiation of trial medication during the pandemic.

Montgomery–Åsberg Depression Rating Scale descriptive summaries for visits that took place pre-1 March are shown in *Table 50*. Mean MADRS score decreased over the follow-up period for both arms however, participants in the quetiapine arm tended to score lower on average than those in the lithium arm, with a mean (SD) of 23.4 (9.6) compared to 25.3 (9.3) at week 8, and a mean (SD) of 21.2 (10.5) compared to 24.5 (9.7) at week 26; and a mean (SD) of 20.5 (10.4) compared to 22.6 (9.8) at week 52. These values appear similar to those reported in the whole trial

TABLE 49 Time to discontinuation before and on or after 1 March 2020

		Quetiapine	Lithium	Overall
Date of discontinuation before 1 March 2020	Total *N* at risk	107	105	212
	n (%)	81 (75.7)	80 (76.2)	161 (75.9)
Prescribed treatment	*N*	81	80	161
	n (%)	70 (86.4)	62 (77.5)	132 (82.0)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
Initiated treatment	*N*	81	80	161
	n (%)	70 (86.4)	60 (75.0)	130 (80.7)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
Discontinued before 12 months	*N*	81	80	161
	n (%)	32 (39.5)	34 (42.5)	66 (41.0)
	Missing, *n* (%)	0 (0.0)	1 (1.3)	1 (0.6)
Time to discontinuation	Mean (SD)	206.9 (159.8)	179.1 (151.3)	193.2 (155.8)
	Median (IQR)	240.0 (53.0–365.0)	170.0 (0.5–365.0)	174.0 (17.5–365.0)
Date of discontinuation on or after 1 March 2020	Total *N* at risk	107	105	212
	n (%)	26 (24.3)	25 (23.8)	51 (24.1)
Prescribed treatment	*N*	26	25	51
	n (%)	25 (96.2)	24 (96.0)	49 (96.1)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
Initiated treatment	*N*	26	25	51
	n (%)	25 (96.2)	24 (96.0)	49 (96.1)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
Discontinued before 12 months	*N*	26	25	51
	n (%)	5 (19.2)	8 (32.0)	13 (25.5)
	Missing, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)
Time to discontinuation	Mean (SD)	293.9 (135.1)	273.4 (142.6)	283.8 (137.8)
	Median (IQR)	365.0 (365.0–365.0)	365.0 (163.0–365.0)	365.0 (163.0–365.0)

population. Similarly, the proportion of responders and remitters at each follow-up point in each arm was similar to those reported in the whole trial population.

Montgomery–Åsberg Depression Rating Scale descriptive summaries for visits that took place on or after 1 March are shown in *Table 51*. Mean MADRS scores decreased over the follow-up period in the quetiapine arm. In the lithium arm, mean MADRS scores were lower at the week 26 follow-up than at week 8, but then increased at the week 52 follow-up. As in both the whole trial population and for outcomes measured prior to 1 March 2020, scores were higher in the lithium arm. The responder and remitter summaries also show a similar pattern to those reported in the whole trial population, with a greater proportion of responders and remitters at each follow-up point in the quetiapine arm than in the lithium arm. There did not appear to be any clear pattern indicating higher MADRS scores after the beginning of the COVID-19 pandemic. This was generally also the case for responders and remitters, with possible exceptions being for lithium at week 8 and quetiapine response at week 52.

TABLE 50 Montgomery–Åsberg Depression Rating Scale secondary outcome measures for visits that took place pre COVID-19 measures

Outcomes		Week 8			Week 26			Week 52		
		Quetiapine, *N* = 94	Lithium, *N* = 95	Overall, *N* = 189	Quetiapine, *N* = 82	Lithium, *N* = 84	Overall, *N* = 166	Quetiapine, *N* = 64	Lithium, *N* = 63	Overall, *N* = 127
MADRS score	Mean (SD)	23.4 (9.6)	25.3 (9.3)	24.3 (9.5)	21.2 (10.5)	24.5 (9.7)	22.7 (10.2)	20.5 (10.4)	22.6 (9.8)	21.4 (10.1)
	Median (IQR)	24.0 (16.0–31.0)	26.0 (20.0–32.0)	24.5 (17.0–31.0)	21.5 (14.0–29.0)	24.5 (18.0–32.0)	23.0 (15.0–32.0)	20.0 (12.0–30.0)	23.0 (17.0–30.0)	22.5 (13.0–30.0)
	Missing	15 (16.0)	20 (21.1)	35 (18.5)	20 (24.4)	32 (38.1)	52 (31.3)	21 (32.8)	32 (50.8)	53 (41.7)
		n (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)
Responder		14 (14.9)	10 (10.5)	24 (12.7)	18 (22.0)	10 (11.9)	28 (16.9)	18 (28.1)	7 (11.1)	25 (19.7)
Remission		9 (9.6)	6 (6.3)	15 (7.9)	9 (11.0)	6 (7.1)	15 (9.0)	8 (12.5)	6 (9.5)	14 (11.0)

TABLE 51 Montgomery–Åsberg Depression Rating Scale secondary outcome measures where the visit took place post COVID-19 measures

Outcomes		Week 8			Week 26			Week 52		
		Quetiapine, *N* = 13	Lithium, *N* = 10	Overall, *N* = 23	Quetiapine, *N* = 25	Lithium, *N* = 21	Overall, *N* = 46	Quetiapine, *N* = 43	Lithium, *N* = 42	Overall, *N* = 85
MADRS score	Mean (SD)	23.1 (11.9)	24.5 (9.1)	23.5 (10.8)	21.6 (10.9)	20.9 (10.4)	21.3 (10.5)	19.7 (8.7)	22.0 (8.5)	20.8 (8.6)
	Median (IQR)	24.0 (21.0–31.0)	22.5 (18.0–31.0)	24.0 (20.0–31.0)	22.0 (11.0–30.0)	18.0 (14.0–26.0)	18.0 (13.0–30.0)	20.5 (13.0–26.0)	21.0 (16.0–28.0)	21.0 (16.0–26.0)
	Missing	4 (30.8)	6 (60.0)	10 (43.5)	7 (28.0)	8 (38.1)	15 (32.6)	17 (39.5)	19 (45.2)	36 (42.4)
		n (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)
Responder		2 (15.4)	0 (0.0)	2 (8.7)	6 (24.0)	5 (23.8)	11 (23.9)	7 (16.3)	5 (11.9)	12 (14.1)
Remission		2 (15.4)	0 (0.0)	2 (8.7)	4 (16.0)	1 (4.8)	5 (10.9)	4 (9.3)	3 (7.1)	7 (8.2)

Treatment and visit windows

Table 52 presents time to discontinuation, prescription and initiation for those who were randomised before 1 February 2020 (which differs from what is presented in *Table 49*). For those randomised pre-COVID-19, the time to discontinuation was median (IQR) 365.0 (66.0–365.0) days in the quetiapine arm and 219.0 (25.5–365.0) days in the lithium arm. Time to prescription was similar between arms, as well as the times to initiation. In the quetiapine arm, 91.7% were prescribed and initiated treatment, whereas in the lithium arm, 81.4% were prescribed treatment and 79.4% initiated treatment. These values are broadly similar to those in the whole population.

As reported in *Table 53*, the median (IQR) time to discontinuation for those randomised post COVID-19, was 145.0 (0.5–365.0) days in the quetiapine arm and 42.0 (18.0–213.5) days in the lithium arm, which were both appreciably shorter as compared to the times for those randomised pre COVID-19. Times to prescription and initiation were similar in the quetiapine arm pre- and post-1 February; however these were shorter in the lithium arm for people randomised during the pandemic period. In the quetiapine arm, 63.6% were prescribed and initiated treatment, whereas in the lithium arm, 87.5% were prescribed and initiated treatment. The proportion being prescribed and initiating lithium in the pandemic period was similar to those pre 1 February, whereas these proportions were lower in the quetiapine arm during the pandemic (although both based on small numbers post-1 February).

TABLE 52 Time to discontinuation, prescription, and initiation for those randomised pre COVID-19

Outcomes		Quetiapine *N* = 96	Lithium *N* = 97	Overall *N* = 193
Time to discontinuation	Mean (SD)	233.5 (155.3)	209.1 (153.8)	221.3 (154.6)
	Median (IQR)	365.0 (66.0–365.0)	219.0 (25.5–365.0)	293.5 (53.0–365.0)
Time to prescription	Mean (SD)	23.7 (36.6)	22.9 (24.5)	23.3 (31.4)
	Median (IQR)	14.0 (4.0–30.0)	15.0 (6.0–32.0)	14.0 (5.0–30.0)
Time to initiation	Mean (SD)	26.9 (37.3)	26.4 (25.6)	26.7 (32.2)
	Median (IQR)	15.0 (6.0–33.0)	20.0 (8.0–38.0)	19.5 (7.0–34.0)
		n (%)	*n* (%)	*n* (%)
Prescribed treatment		88 (91.7)	79 (81.4)	167 (86.5)
Initiated treatment		88 (91.7)	77 (79.4)	165 (85.5)

TABLE 53 Time to discontinuation, prescription and initiation for those randomised post COVID-19

Outcomes		Quetiapine *N* = 11	Lithium *N* = 8	Overall *N* = 19
Time to discontinuation	Mean (SD)	180.6 (181.2)	114.1 (136.2)	152.6 (163.1)
	Median (IQR)	145.0 (0.5–365.0)	42.0 (18.0–213.5)	52.0 (0.5–365.0)
Time to prescription	Mean (SD)	23.7 (15.3)	16.4 (11.4)	20.1 (13.5)
	Median (IQR)	22.0 (15.0–38.0)	15.0 (8.0–26.0)	16.5 (15.0–28.0)
Time to initiation	Mean (SD)	25.6 (16.6)	17.9 (12.2)	21.7 (14.6)
	Median (IQR)	22.0 (15.0–45.0)	16.0 (8.0–31.0)	18.5 (15.0–33.0)
		n (%)	*n* (%)	*n* (%)
Prescribed treatment		7 (63.6)	7 (87.5)	14 (73.7)
Initiated treatment		7 (63.6)	7 (87.5)	14 (73.7)

Table 54 presents time to each visit in the full sample. *Table 55* presents these data for those who had their outcome data collected before 1 March 2020 and for those who had their data collected on or after 1 March 2020. Time to visit was similar in lithium and quetiapine arms in both the pre- and post-1 March 2020 periods.

Serious adverse events

Serious adverse events starting before and after COVID-19 measures were implemented (i.e. before and on or before 1 March 2020) are shown in *Table 56*. There was a total of 30 SAEs from 16 participants before 1 March 2020. Fifteen SAEs were from seven participants randomised to the quetiapine arm, where 46.7% of SAEs were unlikely to be related to quetiapine, and 53.3% were not related to quetiapine. Fifteen SAEs were from nine participants randomised to the lithium arm, where 6.7% were possibly related to lithium treatment, 46.7% were unlikely to be related, and 46.7% were not related to lithium.

After 1 March 2020, there was a total of two SAEs from two participants, solely in the lithium arm, both unrelated to lithium. There did not appear to be an excess of SAEs during the pandemic period, nor was there a difference between arms, but it is difficult to tell with the small numbers of SAEs occurring during the pandemic period. No SAEs were COVID-19 related.

Primary outcome descriptives and inferential analyses

Sensitivity analyses were performed on the primary outcomes (QIDS-SR and discontinuation data), observing the following scenarios:

1. completely excluding participants that were randomised after 1 February 2020; and
2. excluding any of the outcome measures taken after 1 March 2020. For the discontinuation outcome this equated to setting 1 March 2020 as the censoring date and excluding anyone receiving a prescription after 1 March 2020.

Self-rated Quick Inventory of Depressive Symptomatology descriptive summaries are shown in *Table 57*. For participants who were randomised before 1 February 2020, in the quetiapine arm the mean (SD) QIDS-SR score was 17.0 (4.4) at baseline, which decreased at the 52-week follow-up to mean (SD) 12.2 (6.3). In the lithium arm, the mean (SD) QIDS-SR score at baseline was 17.8 (4.1) and decreased to 14.1 (6.2) at week 52.

The unadjusted mean QIDS-SR scores, and 95% CIs, collected each week via the trial weekly visits and True Colours, for participants who were randomised before 1 February 2020, are presented by trial arm in *Figure 41*. Excluding individuals randomised on or after 1 February 2020 yielded an area under the between arm difference curve of quetiapine versus lithium of −73.07 (95% CI: −137.43 to −8.71), which was statistically significant (p = 0.0261), see *Table 58*. This result was similar to that reported in the whole trial population.

TABLE 54 Visit windows in the full sample

Visit window		Quetiapine, *N* = 107	Lithium, *N* = 105	Overall, *N* = 212
Time to week 8 visit	Mean (SD)	59.6 (13.4)	58.4 (11.9)	59.0 (12.7)
	Median (IQR)	56.0 (56.0–58.0)	56.0 (56.0–58.0)	56.0 (56.0–58.0)
	Missing, *n* (%)	9 (8.4)	12 (11.4)	21 (9.9)
Time to week 26 visit	Mean (SD)	185.0 (13.3)	186.0 (13.6)	185.5 (13.4)
	Median (IQR)	182.0 (182.0–185.0)	182.5 (182.0–187.0)	182.0 (182.0–186.0)
	Missing, *n* (%)	18 (16.8)	33 (31.4)	51 (24.1)
Time to week 52 visit	Mean (SD)	368.8 (11.5)	371.7 (15.4)	370.2 (13.5)
	Median (IQR)	365.0 (364.0–372.0)	367.0 (364.0–372.0)	366.0 (364.0–372.0)
	Missing, *n* (%)	27 (25.2)	34 (32.4)	61 (28.8)

TABLE 55 Visit windows pre and post COVID-19 measures

Visit windows pre COVID-19 measures		Quetiapine	Lithium	Overall
Time to week 8 visit	*N*	88	87	175
	Mean (SD)	59.2 (12.2)	59.1 (11.7)	59.1 (11.9)
	Median (IQR)	56.0 (56.0–58.0)	56.0 (56.0–59.0)	56.0 (56.0–58.0)
Time to week 26 visit	*N*	70	57	127
	Mean (SD)	185.4 (14.6)	185.7 (12.6)	185.6 (13.7)
	Median (IQR)	182.0 (182.0–185.0)	183.0 (182.0–187.0)	183.0 (182.0–187.0)
Time to week 52 visit	*N*	50	40	90
	Mean (SD)	369.6 (12.3)	370.1 (15.3)	369.8 (13.6)
	Median (IQR)	364.5 (364.0–372.0)	366.0 (364.0–371.5)	365.5 (364.0–372.0)
Visit window post COVID-19 measures		**Quetiapine**	**Lithium**	**Overall**
Time to week 8 visit	*N*	10	6	16
	Mean (SD)	63.6 (21.3)	49.3 (12.3)	58.3 (19.4)
	Median (IQR)	55.5 (54.0–58.0)	54.0 (49.0–57.0)	55.5 (53.0–57.0)
Time to week 26 visit	*N*	19	15	34
	Mean (SD)	183.5 (6.1)	187.1 (17.4)	185.1 (12.3)
	Median (IQR)	182.0 (181.0–186.0)	182.0 (175.0–187.0)	182.0 (180.0–186.0)
Time to week 52 visit	*N*	30	31	61
	Mean (SD)	367.5 (10.0)	373.7 (15.6)	370.7 (13.4)
	Median (IQR)	365.0 (364.0–371.0)	368.0 (364.0–378.0)	367.0 (364.0–374.0)

TABLE 56 Serious adverse events starting before and after COVID-19 measures

Serious adverse events starting before implementation of COVID-19 measures (before 1 March 2020)				
		Quetiapine	Lithium	Overall
		n	*n*	*n*
Number of SAEs		15	15	30
Related to lithium or quetiapine add on therapy, *n* (%)	Definitely related	0 (0.0)	0 (0.0)	0 (0.0)
	Likely related	0 (0.0)	0 (0.0)	0 (0.0)
	Possibly related	0 (0.0)	1 (6.7)	1 (3.3)
	Unlikely related	7 (46.7)	7 (46.7)	14 (46.7)
	Not related	8 (53.3)	7 (46.7)	15 (50.0)
Number of participants with SAEs		7	9	16

continued

TABLE 56 Serious adverse events starting before and after COVID-19 measures (*continued*)

SAEs starting after implementation of COVID-19 measures (on or after 1 March 2020)				
		Quetiapine	Lithium	Overall
		n	*n*	*n*
Number of SAEs		0	2	2
Related to lithium or quetiapine add on therapy, *n* (%)	Definitely related	0 (0.0)	0 (0.0)	0 (0.0)
	Likely related	0 (0.0)	0 (0.0)	0 (0.0)
	Possibly related	0 (0.0)	0 (0.0)	0 (0.0)
	Unlikely related	0 (0.0)	0 (0.0)	0 (0.0)
	Not related	0 (0.0)	2 (100.0)	2 (100.0)
Number of participants with SAEs		0	2	2

TABLE 57 Mean QIDS-SR (unadjusted) descriptive statistics at study visits excluding post-COVID-19 participants

		Quetiapine	Lithium	Overall
		N = 96	*N* = 97	*N* = 193
QIDS-SR score at baseline	Mean (SD)	17.0 (4.4)	17.8 (4.1)	17.4 (4.3)
	Median (IQR)	17.0 (14.5–21.0)	18.0 (15.0–21.0)	17.0 (15.0–21.0)
QIDS-SR score at week 8	Mean (SD)	14.3 (5.7)	15.2 (5.8)	14.7 (5.7)
	Median (IQR)	14.0 (9.0–18.0)	16.0 (11.0–20.0)	15.0 (11.0–19.0)
QIDS-SR score at week 26	Mean (SD)	14.2 (5.5)	14.6 (6.9)	14.4 (6.1)
	Median (IQR)	14.0 (10.0–18.0)	15.5 (9.0–20.5)	15.0 (10.0–19.0)
QIDS-SR score at week 52	Mean (SD)	12.2 (6.3)	14.1 (6.2)	13.1 (6.3)
	Median (IQR)	12.0 (8.0–16.0)	15.0 (10.0–19.0)	13.0 (9.0–18.0)

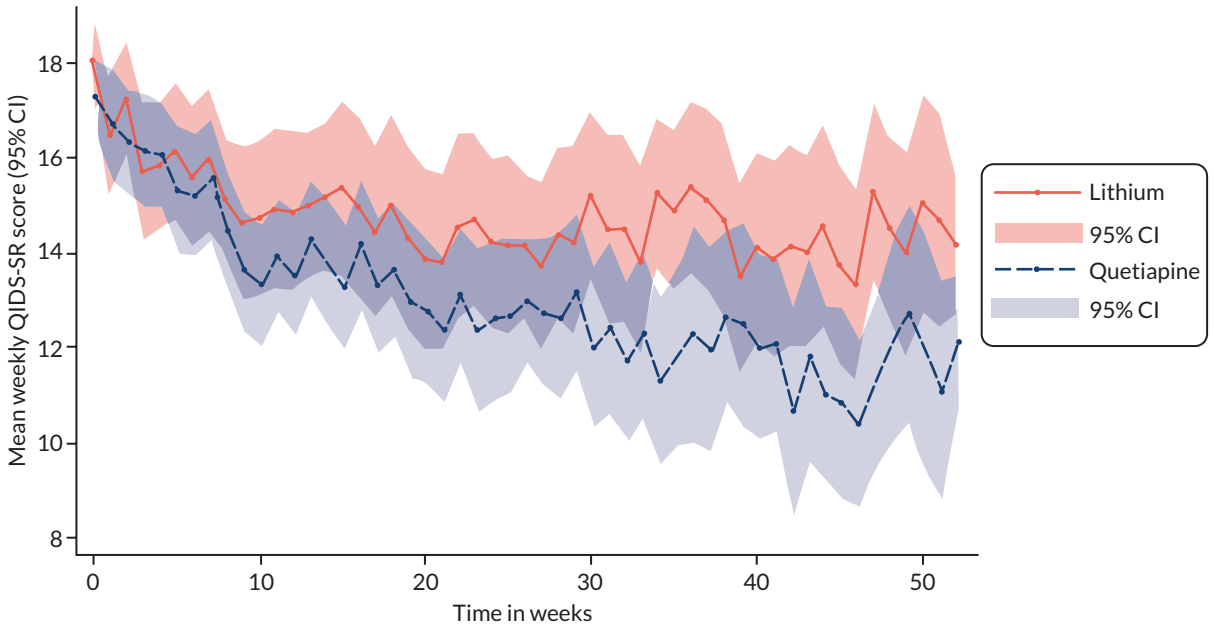

FIGURE 41 Mean weekly QIDS-SR scores (unadjusted) excluding post-COVID-19 participants.

TABLE 58 Between arm area under the QIDS-SR difference curve, excluding post-COVID participants

	Quetiapine vs. lithium difference in AUC	95% CI	*p*-value
Weekly QIDS-SR adjusted by stratification factors[a] and missingness predictors[b]	−73.07	(−137.43 to −8.71)	0.0261

a Stratification factors (TRD severity, depression severity and site).
b Missing predictors by number of comorbidities and ethnicity.

The median time to discontinuation and inferential results were similar in participants randomised before 1 February 2020 as compared to the results in the overall trial population, see *Tables 59* and *60*. The Kaplan–Meier curve in *Figure 42* shows time to discontinuation between the treatment arms over 365 days, excluding participants who were randomised on or after 1 February 2020.

Self-rated Quick Inventory of Depressive Symptomatology descriptive summaries for participants whose measures were collected before 1 March 2020 are shown in *Table 61*. In the quetiapine arm the mean (SD) QIDS-SR score was

TABLE 59 Time to discontinuation descriptives excluding post-COVID-19 participants

		Quetiapine, *N* = 96	Lithium, *N* = 96
Time to discontinuation	Mean days (95% CI)	233.50 (202.02 to 264.98)	209.06 (177.90 to 240.22)
	Median days (IQR)	365.0 (66.0–365.0)	219.0 (25.5–365.0)

TABLE 60 Between-arm time to discontinuation analyses excluding post-COVID-19 participants

Time to discontinuation[a]	Hazard ratio (95% CI)	*p*-value	RMST difference (95% CI)	*p*-value
Quetiapine vs. lithium	0.73 (0.47 to 1.11)	0.1413	20.12 (−19.16 to 59.41)	0.3153
Lithium vs. quetiapine	1.38 (0.90 to 2.11)	0.1413		

a Adjusted by stratification factors (TRD severity, depression severity and site).

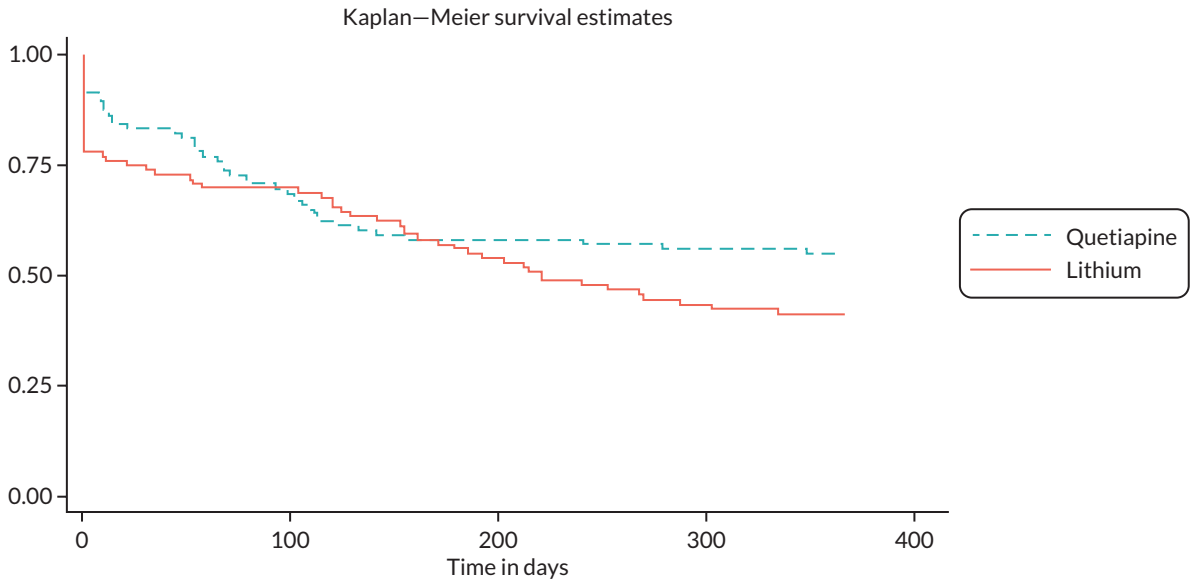

FIGURE 42 Time to discontinuation Kaplan–Meier curve by arm excluding participants who were randomised post-COVID-19.

TABLE 61 Self-rated Quick Inventory of Depressive Symptomatology (unadjusted) descriptive statistics excluding post-COVID-19 measures

		Quetiapine	Lithium	Overall
QIDS-SR score at baseline	*N*	101	99	200
	Mean (SD)	17.2 (4.4)	17.9 (4.1)	17.6 (4.3)
	Median (IQR)	17.0 (15.0–21.0)	18.0 (15.0–21.0)	18.0 (15.0–21.0)
QIDS-SR score at week 8	*N*	95	96	191
	Mean (SD)	14.3 (5.7)	15.2 (5.8)	14.7 (5.7)
	Median (IQR)	14.0 (9.0–18.0)	16.0 (11.0–20.0)	15.0 (11.0–19.0)
QIDS-SR score at week 26	*N*	82	86	168
	Mean (SD)	13.4 (5.8)	14.8 (6.0)	14.0 (5.9)
	Median (IQR)	13.0 (9.5–17.5)	16.5 (11.0–19.5)	14.0 (11.0–19.0)
QIDS-SR score at week 52	*N*	64	63	127
	Mean (SD)	12.7 (6.6)	14.6 (5.9)	13.5 (6.3)
	Median (IQR)	13.0 (8.0–18.0)	15.0 (10.0–18.5)	13.0 (9.0–18.0)

17.2 (4.4) at baseline, which decreased at the 52-week follow-up to mean (SD) 12.7 (6.6). In the lithium arm, the mean (SD) QIDS-SR score at baseline was 17.9 (4.1) and decreased to 14.6 (5.9) at week 52. The unadjusted mean QIDS-SR scores, and 95% CIs, collected each week via the trial visits and True Colours, for participants whose measure was collected before 1 March 2020, are presented by trial arm in *Figure 43*.

Excluding QIDS-SR outcomes measured on or after 1 March 2020 yielded an area under the between arm difference curve for quetiapine versus lithium of −76.02 (95% CI: −144.27 to −7.78) which was statistically significant ($p = 0.0290$), see *Table 62*. This result was similar to that reported in the whole trial population.

The time to discontinuation inferential results were similar in the group both prescribed before 1 March 2020 and censored at this date as compared to the results in the overall trial population, see *Table 63* and *Table 64*. However, we

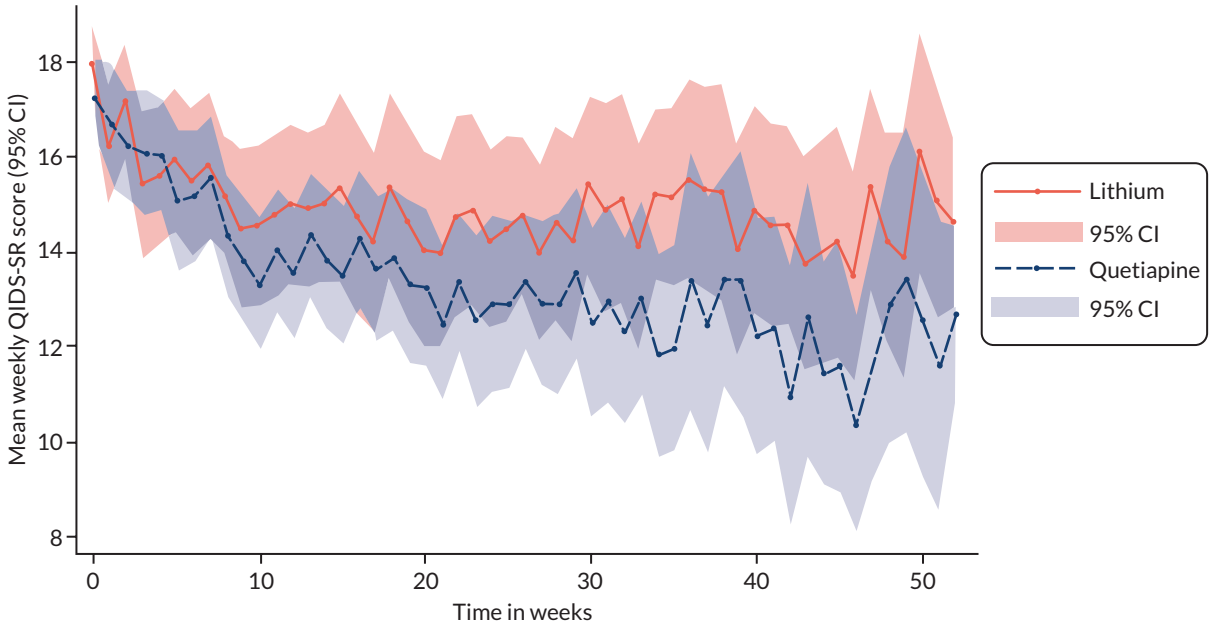

FIGURE 43 Mean weekly QIDS-SR scores (unadjusted) excluding post-COVID-19 measures.

TABLE 62 Between arm area under the QIDS-SR difference curve estimates, excluding post-COVID-19 measures

	Quetiapine vs. lithium difference in AUC	95% CI	*p*-value
Weekly QIDS-SR adjusted by stratification factors[a] and missingness predictors[b]	−76.02	(−144.27 to −7.78)	0.0290

a Stratification factors (TRD severity, depression severity and site).
b Missing predictors by number of comorbidities and ethnicity.

TABLE 63 Time to discontinuation descriptives excluding post-COVID-19 measures

		Quetiapine, *N* = 100	Lithium, *N* = 97
Time to discontinuation	Mean days (95% CI)	197.75 (167.80 to 227.69)	179.57 (150.94 to 208.19)
	Median days (IQR)	173.0 (53.0–365.0)	178.0 (21.0–365.0)

TABLE 64 Between arm time to discontinuation analyses excluding post-COVID-19 measures

Time to discontinuation[a]	Hazard ratio (95% CI)	*p*-value	RMST difference (95% CI)	*p*-value
Quetiapine vs. lithium	0.76 (0.49 to 1.18)	0.2240	11.81 (−22.19 to 45.82)	0.4959
Lithium vs. quetiapine	1.32 (0.85 to 2.05)	0.2240		

a Adjusted by stratification factors (TRD severity, depression severity and site).

note the median times to discontinuation are shorter in both arms, although the IQR are similar. This will in part be due to the earlier censoring at this follow-up time point. The fact that there are shorter times to discontinuation in both arms leads to the inferential results being similar to those in the whole trial. The Kaplan–Meier curve in *Figure 44* shows time to discontinuation between the treatment arms over 365 days, excluding post-COVID-19 measures and censoring time to discontinuation at 1 March 2020.

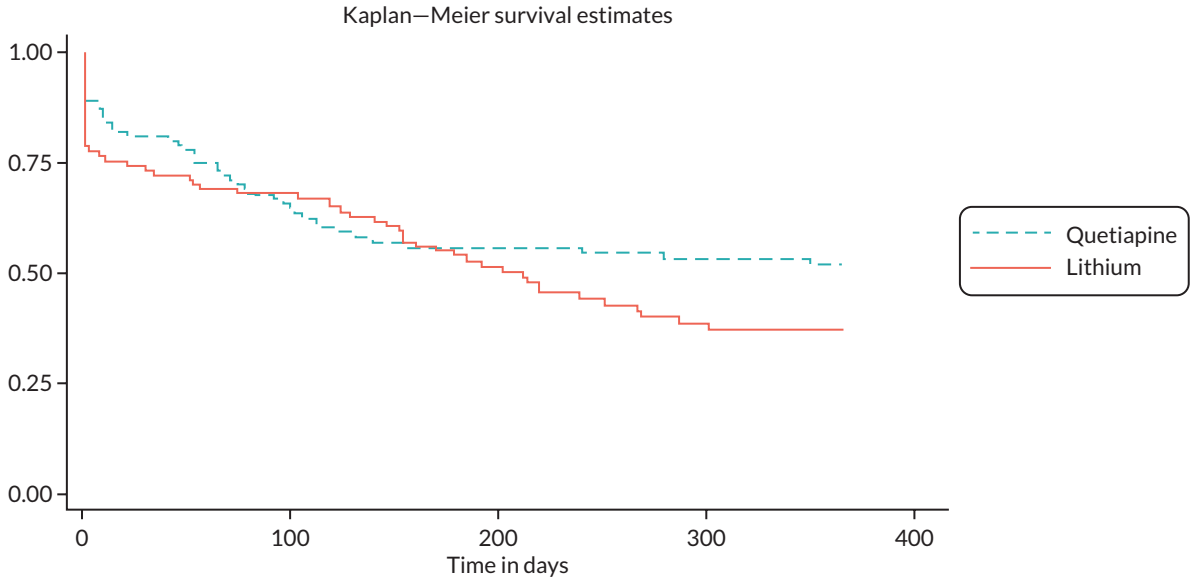

FIGURE 44 Time to discontinuation Kaplan–Meier curve by arm excluding post-COVID-19 measures and censoring time to discontinuation at 1 March 2020.

Appendix 7 Additional secondary outcome inferential analysis figures

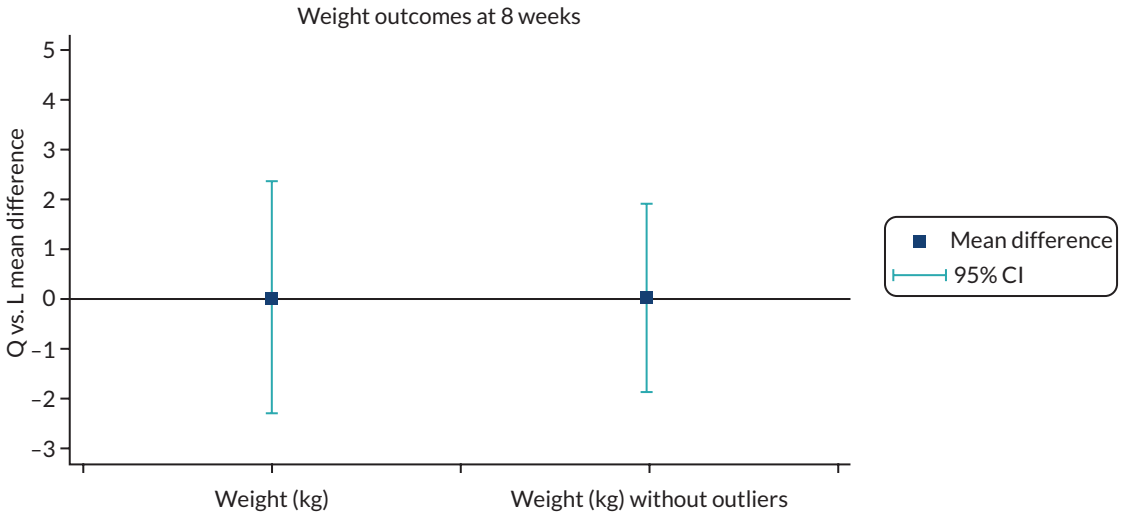

FIGURE 45 Adjusted between arm differences for weight (including and excluding outliers) at week 8. L, lithium; Q, quetiapine.

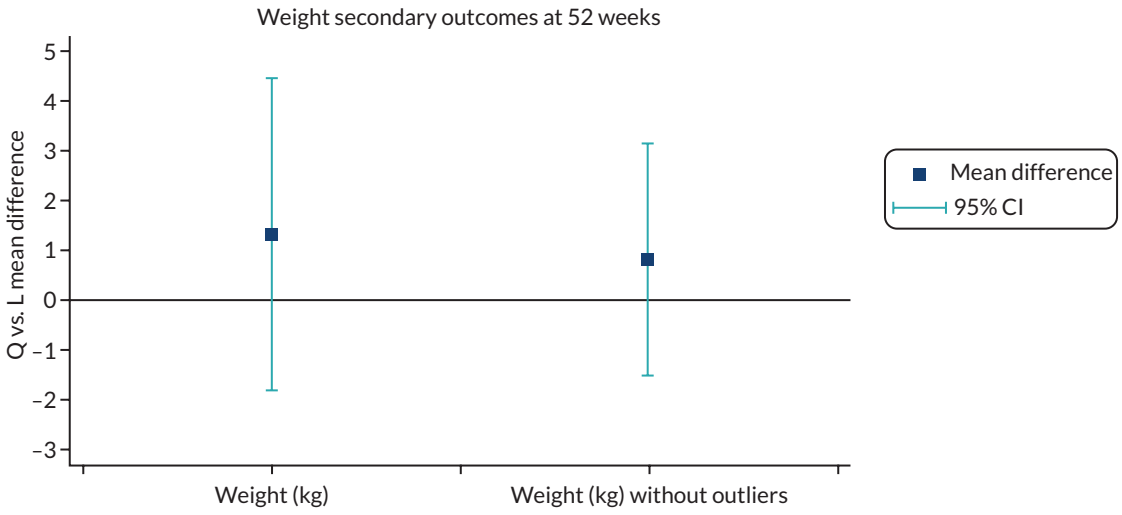

FIGURE 46 Adjusted between arm differences for weight (including and excluding outliers) at week 52. L, lithium; Q, quetiapine.

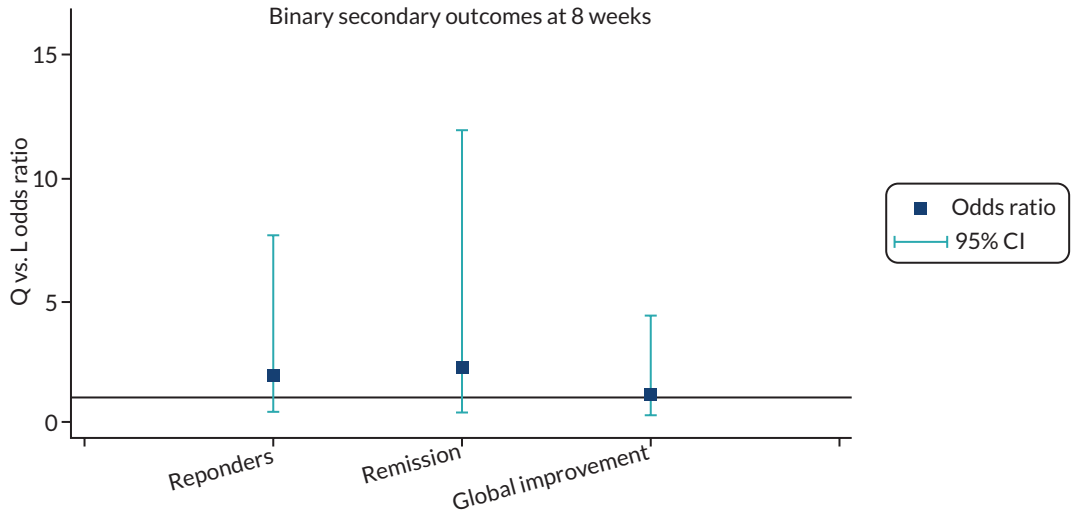

FIGURE 47 Adjusted between arm differences for binary secondary outcomes at week 8. L, lithium; Q, quetiapine.

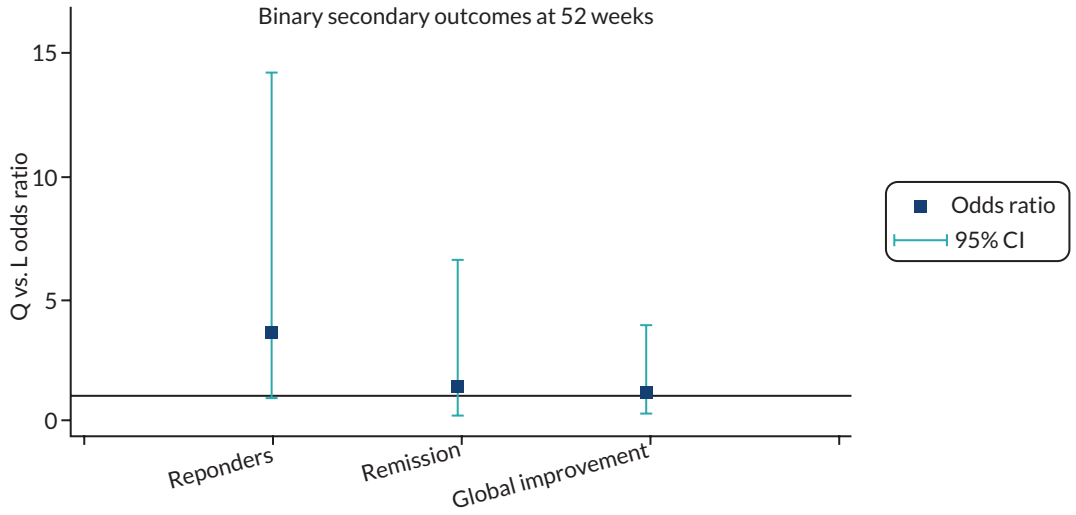

FIGURE 48 Adjusted between arm differences for binary secondary outcomes at week 52. L, lithium; Q, quetiapine.

Appendix 8 Health economic sensitivity analyses

Sensitivity analysis 1: changing the cost of the intervention drug

Primary analysis

Table 65 presents the cost-effectiveness results from the NHS and PSS perspective using the cost of the generic drug. The probability that quetiapine was the more cost-effective option dropped. Results of bootstrapping (*Figure 49*) showed that the most likely scenario with 90.4% of the re-samples was that quetiapine resulted in higher NHS and PSS costs and a higher QALY gain compared with lithium. The probability that quetiapine was the cost-effective option was 0.92 at NICE's WTP threshold of £20,000 per additional unit of QALY (*Figure 50*).

Secondary analysis

Adopting a societal perspective

Table 66 presents the cost-effectiveness analysis when the cost of the generic drug was used and a societal costing perspective was adopted. Results of bootstrapping (*Figure 51*) showed that the most likely scenario with 96.2% of

TABLE 65 Cost-effectiveness results from the NHS and PSS perspective using the cost of generic drug

Intervention	NHS and PSS cost (£)	QALY	Incremental cost[a]	Incremental QALY[b]	ICER	INB
Lithium	3151.46	0.468	–	–	–	–
Quetiapine	3582.31	0.540	421.60	0.074	5697.29	0.052

a Missing values are multiple imputed and the estimated difference is adjusted for arm, EQ-5D baseline index value, baseline cost of the same item, QIDS baseline score, site and failure of antidepressant treatment.
b Missing values are multiple imputed and the estimated difference is adjusted for EQ-5D baseline index value, QIDS baseline score, site and failure of antidepressant treatment.

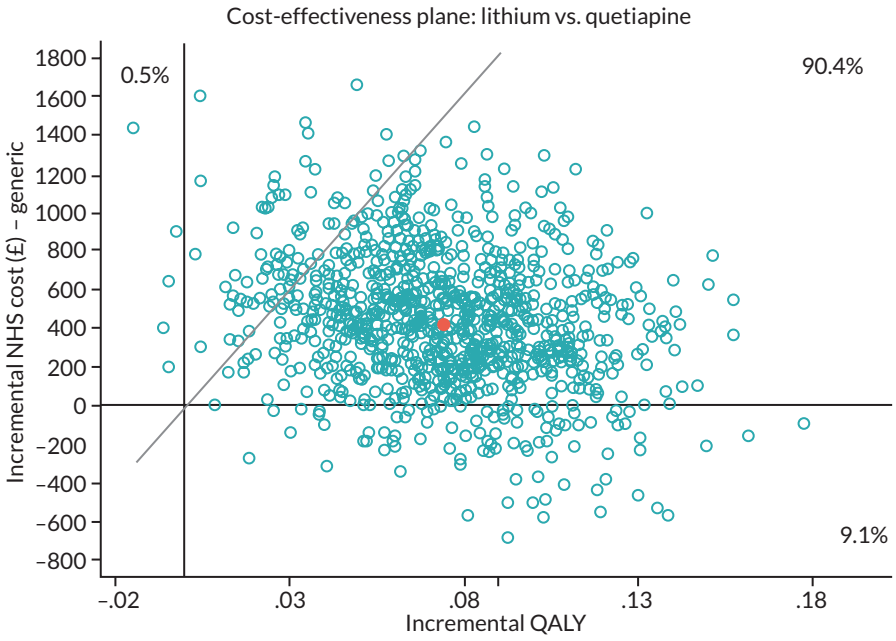

FIGURE 49 Cost-effectiveness plane of NHS and PSS cost and QALY differences – using generic drug costs.

TABLE 66 Cost-effectiveness results from the societal perspective using the cost of generic drug

Intervention	Societal cost (£)	QALY	Incremental cost[a]	Incremental QALY[b]	ICER	INB
Lithium	4366.97	0.468	–	–	–	–
Quetiapine	5577.87	0.540	1173.16	0.074	16818.05	0.015

a Missing values are multiple imputed and the estimated difference is adjusted for arm, EQ-5D baseline index value, baseline cost of the same item, QIDS baseline score, site and failure of antidepressant treatment.
b Missing values are multiple imputed and the estimated difference is adjusted for EQ-5D baseline index value, QIDS baseline score, site and failure of antidepressant treatment.

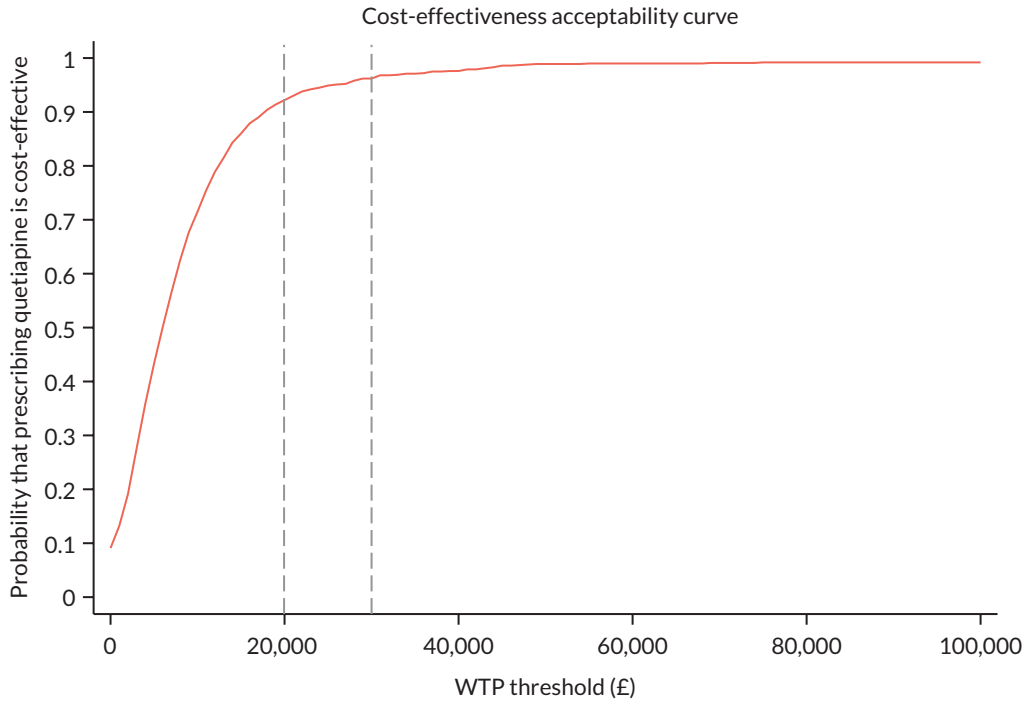

FIGURE 50 Cost-effectiveness acceptability curves showing probability that quetiapine was the most cost-effective option at different WTP thresholds for improvement in QALYs – NHS and PSS perspective and using generic drug costs.

the re-samples was that quetiapine resulted in higher societal costs and higher QALY gain compared to lithium. The probability that quetiapine was the cost-effective option dropped to slightly over 0.5 at NICE's WTP threshold of £20,000 per additional unit of QALY (*Figure 52*).

Using the self-rated Quick Inventory of Depressive Symptomatology as the effectiveness outcome

When the cost of generic drug was used and the QIDS-SR was used as the effectiveness outcome, results of bootstrapping (*Figure 53*) showed that the most likely scenario with 71.5% of the re-samples was that quetiapine resulted in higher NHS and PSS costs and an improvement in QIDS-SR scores compared with lithium. The probability that quetiapine was the cost-effective option is 0.1 at a WTP threshold of £0 per unit improvement in QIDS-SR score (*Figure 54*). As a unit improvement is valued at higher levels the probability increases to reach over 0.7 for a WTP threshold of £20,000.

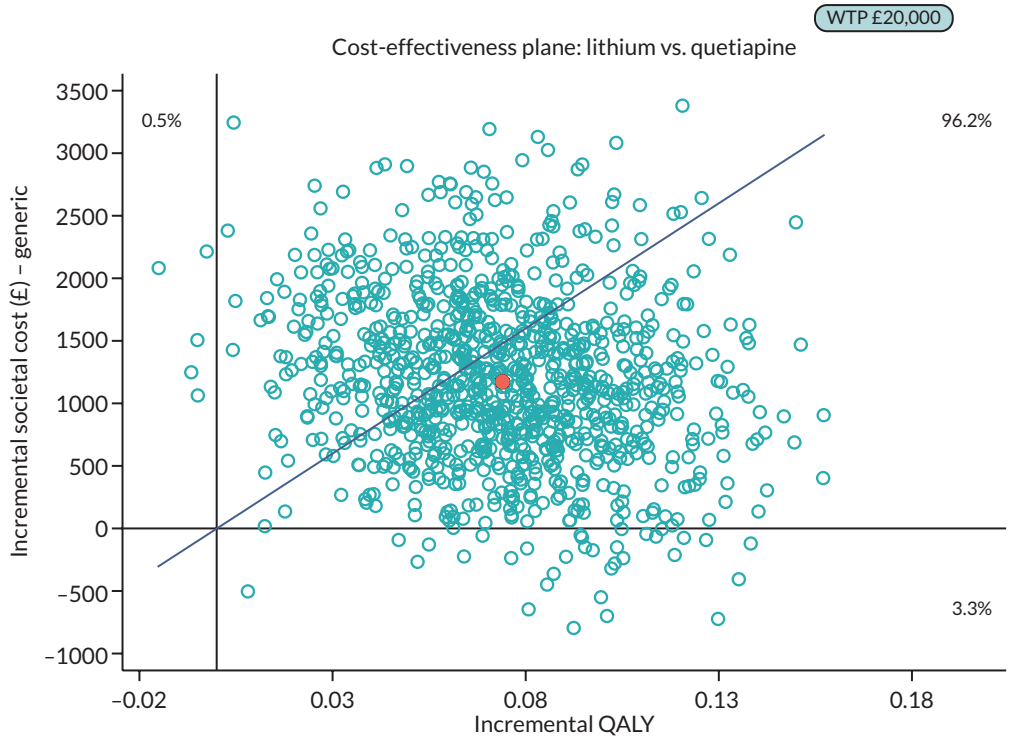

FIGURE 51 Cost-effectiveness plane of societal cost and QALY differences – using generic drug costs.

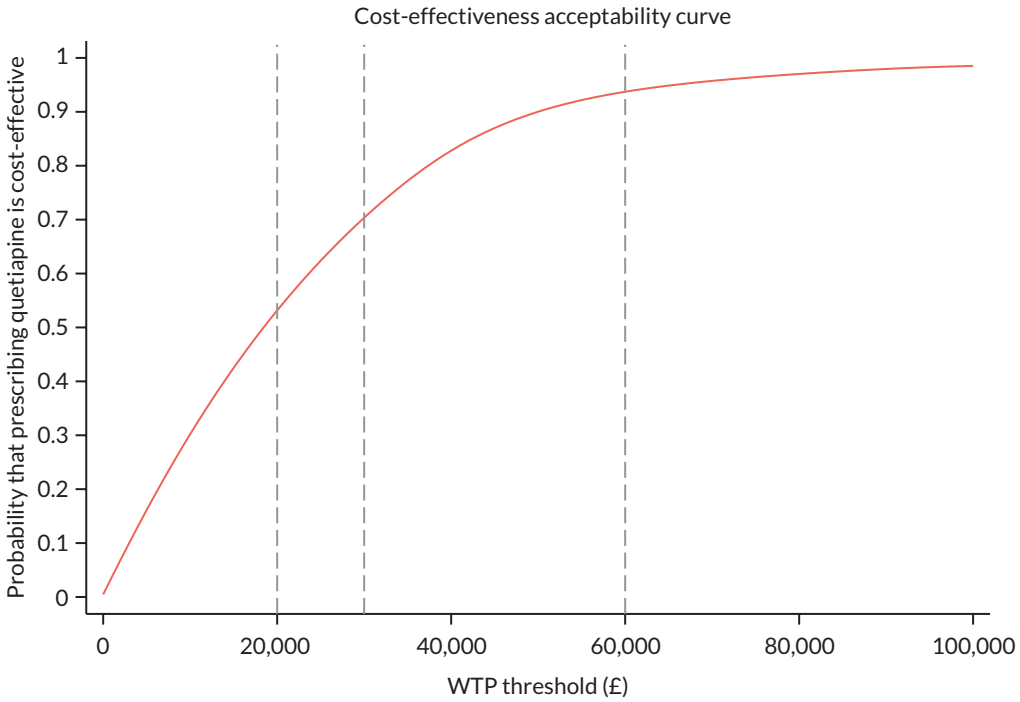

FIGURE 52 Cost-effectiveness acceptability curves showing probability that quetiapine is most cost-effective option at different WTP threshold for improvement in QALYs – societal perspective and using generic drug costs.

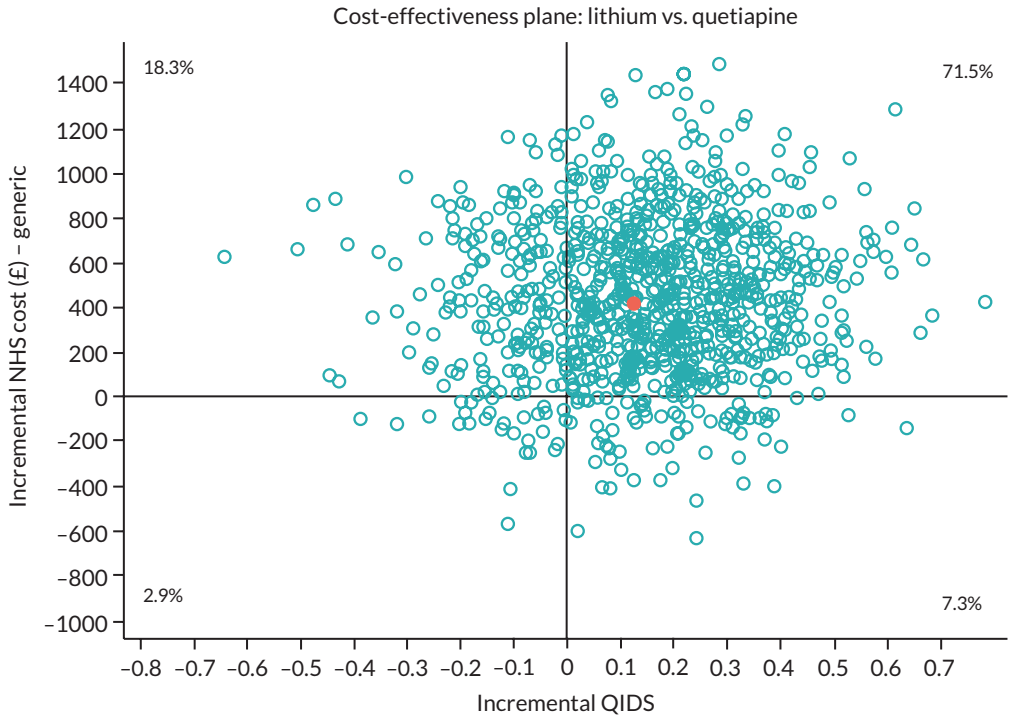

FIGURE 53 Cost-effectiveness plane of NHS and PSS cost and QIDS differences – using generic drug costs.

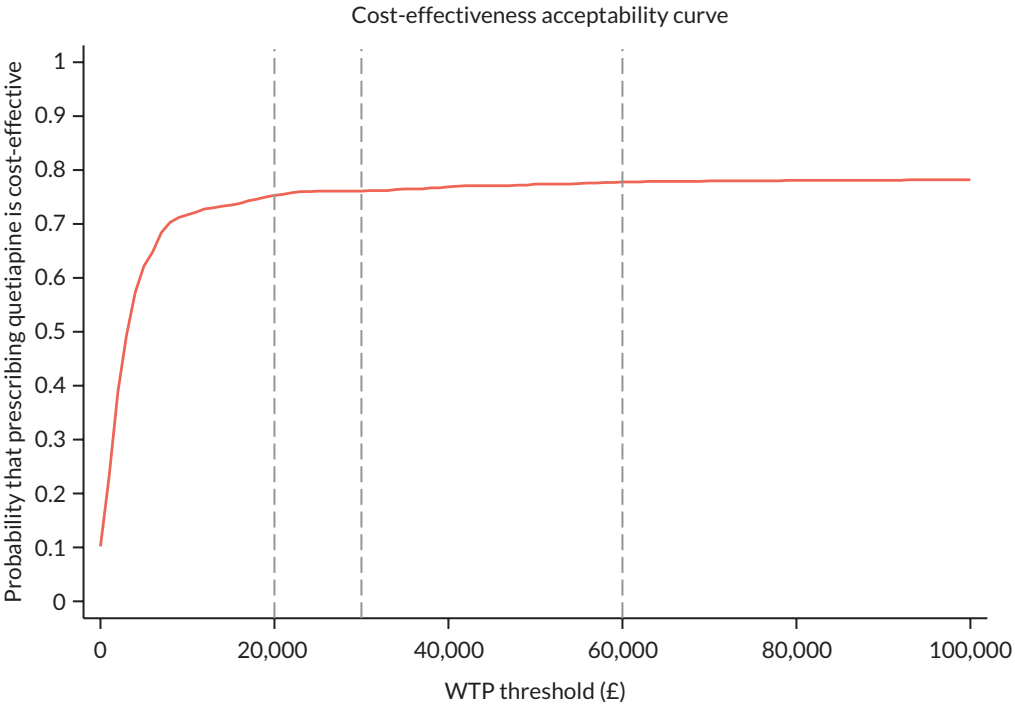

FIGURE 54 Cost-effectiveness acceptability curves showing probability that quetiapine is the most cost-effective option at different WTP threshold for improvement in QIDS-SR score – NHS and PSS perspective, generic drug cost.

Adopting a societal perspective and using the self-rated Quick Inventory of Depressive Symptomatology as the effectiveness outcome

When the cost of the generic drug was used, a societal costing perspective was adopted, and the QIDS-SR was used as the effectiveness outcome, results of bootstrapping (*Figure 55*) showed that the most likely scenario with 75.9% of the re-samples was that quetiapine resulted in higher societal costs and an improvement in QIDS-SR score compared with lithium. The probability that quetiapine was the cost-effective option dropped to < 0.7 for a WTP threshold of £20,000 (*Figure 56*).

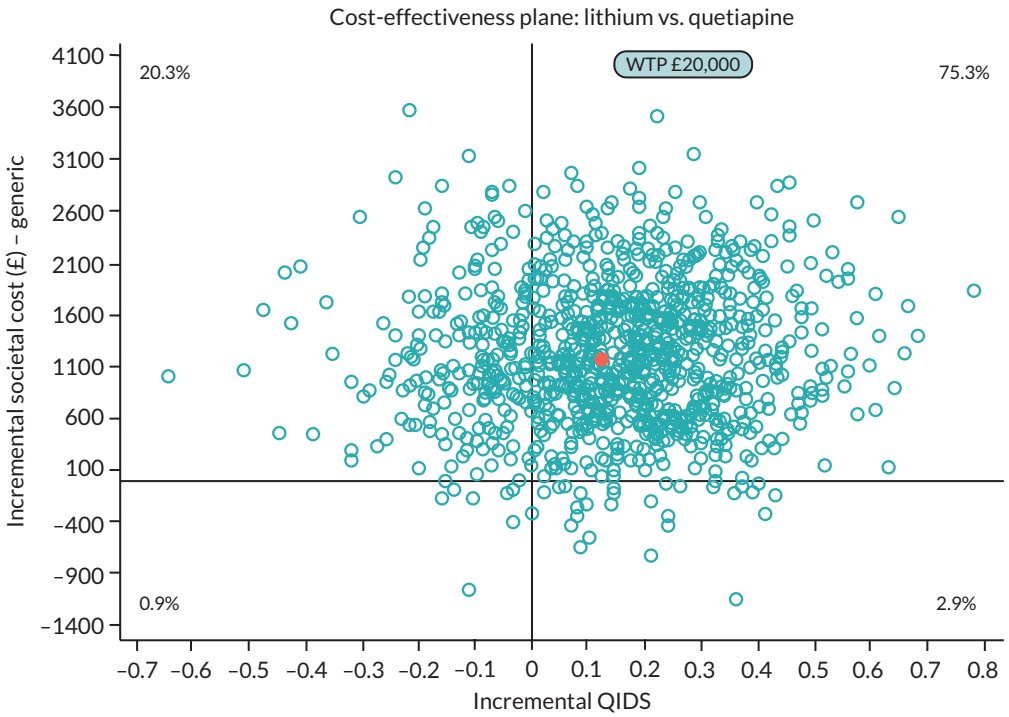

FIGURE 55 Cost-effectiveness plane of societal cost and QIDS-SR differences – using generic drug costs.

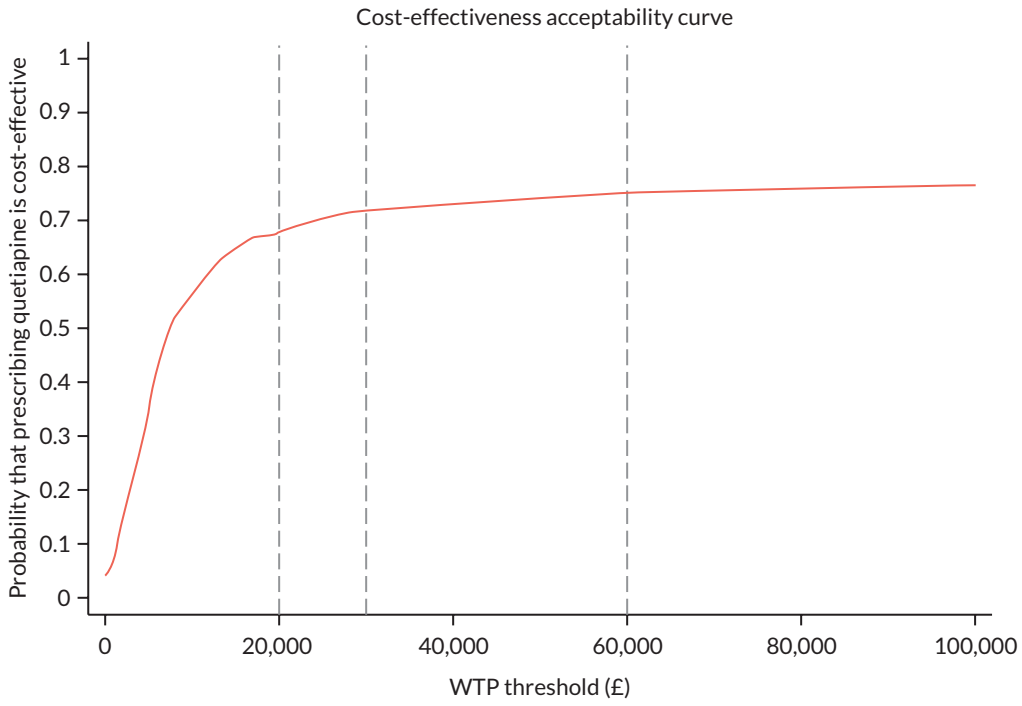

FIGURE 56 Cost-effectiveness acceptability curves showing the probability that quetiapine was the most cost-effective option at different WTP thresholds for improvement in QIDS-SR score – societal perspective, generic drug cost.

Sensitivity analysis 2: testing alternative assumptions of missing data (covariate dependent missing completely at random)

Primary analysis

Results of the cost-effectiveness analysis assuming CD-MCAR and using the cheapest cost of the trial drug are presented in *Table 67*. Of the resamples, 46.8% showed higher costs and higher QALYs gain for the quetiapine arm while 40.1% showed lower costs and a higher QALY gain (*Figure 57*). The probability that quetiapine was the cost-effective option was around 0.8 at NICE's WTP threshold of £20,000 per additional unit of QALY (*Figure 58*).

Secondary analysis

Adopting a societal perspective

Results of the cost-effectiveness analysis assuming CD-MCAR and using the cheapest cost of the trial drug are presented in *Table 68*. Of the resamples, 64.9% (*Figure 59*) showed higher costs and higher QALYs gain for the quetiapine arm while 22.0% showed lower costs and a higher QALY gain. The probability that quetiapine was the cost-effective option dropped to just over 0.6 at NICE's WTP threshold of £20,000 per additional unit of QALY (*Figure 60*).

TABLE 67 Cost-effectiveness results from the NHS/PSS perspective using the cost of cheapest drug

Intervention	Societal cost (£)	QALY	Incremental cost[a]	Incremental QALY[b]	ICER	INB
Lithium	2292.50	0.441	–	–	Dominated	–
Quetiapine	2214.49	0.540	−12.18	0.045	Dominating	0.045

a Estimated difference is adjusted for arm, EQ-5D baseline index value, baseline cost of the same item, QIDS baseline score, site and failure of antidepressant treatment.
b Estimated difference is adjusted for EQ-5D baseline index value, QIDS baseline score, site and failure of antidepressant treatment.

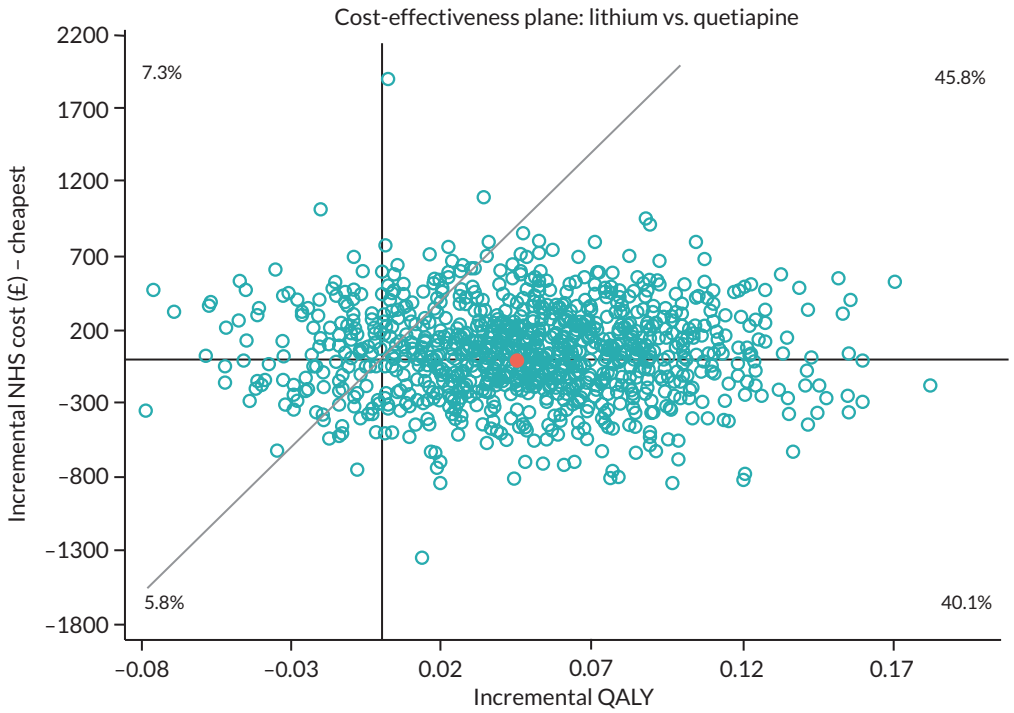

FIGURE 57 Cost-effectiveness plane of NHS and PSS cost and QALY differences, assuming CD-MCAR.

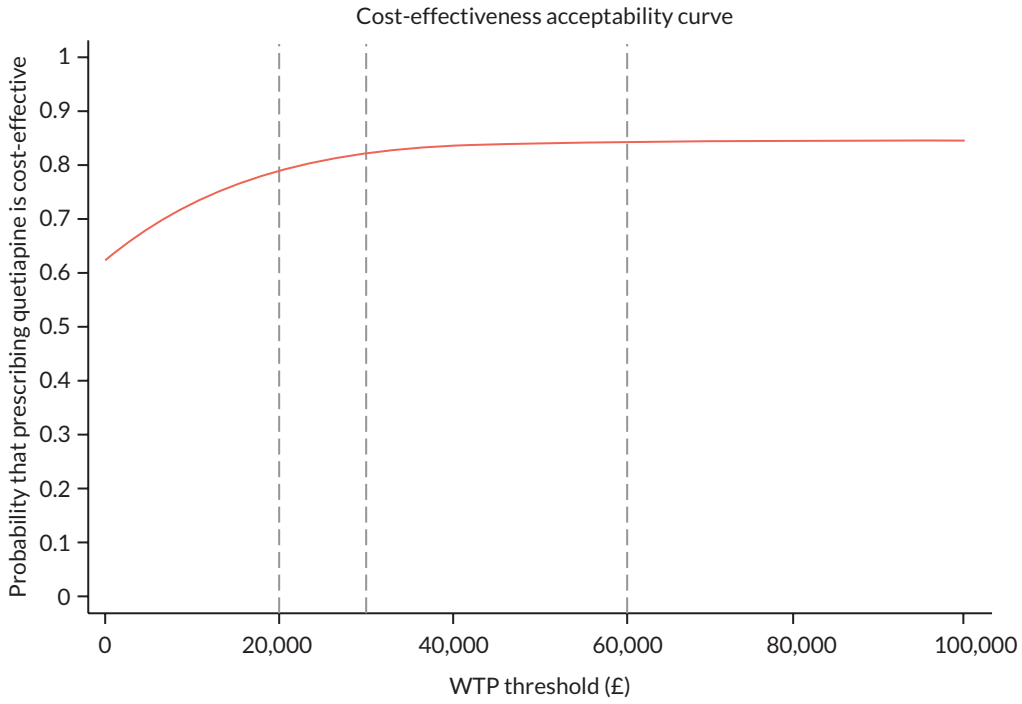

FIGURE 58 Cost-effectiveness acceptability curves showing the probability that quetiapine is the most cost-effective option at different WTP thresholds for improvement in QALYs – NHS and PSS perspective, assuming CD-MCAR.

TABLE 68 Cost-effectiveness results from the societal perspective using the cost of cheapest drug

Intervention	Societal cost (£)	QALY	Incremental cost[a]	Incremental QALY[b]	ICER	INB
Lithium	3556.16	0. 441	–	–	–	–
Quetiapine	4248.06	0. 540	364.53	0.045	8100.66	0.026

a Estimated difference is adjusted for arm, EQ-5D baseline index value, baseline cost of the same item, QIDS baseline score, site and failure of antidepressant treatment.
b Estimated difference is adjusted for EQ-5D baseline index value, QIDS baseline score, site and failure of antidepressant treatment.

Using the self-rated Quick Inventory of Depressive Symptomatology as the effectiveness outcome

When CD-MCAR was assumed and the QIDS-SR was used as the effectiveness outcome, 41.4% of the re-samples showed higher NHS and PSS costs and more improvement in QIDS-SR scores for the quetiapine arm while 37.3% showed lower NHS and PSS costs and more improvement in QIDS-SR scores (*Figure 61*). The probability that quetiapine was the cost-effective option was 0.48 at a WTP threshold of £0 per unit improvement in QIDS-SR score (*Figure 62*). As a unit improvement is valued at higher levels the probability increases.

Adopting a societal perspective and using the self-rated Quick Inventory of Depressive Symptomatology as the effectiveness outcome

When CD-MCAR was assumed, a societal costing perspective was adopted, and the QIDS-SR was used as the effectiveness outcome, 56.8% of the re-samples showed higher societal costs and more improvement in QIDS-SR scores for the quetiapine arm while 22.0% showed lower societal costs and a more improvement in QIDS-SR scores (*Figure 63*). The probability that quetiapine was the more cost-effective option was 0.28 at a WTP threshold of £0 per unit improvement in QIDS-SR score (*Figure 64*). As a unit improvement is valued at higher levels the probability increases.

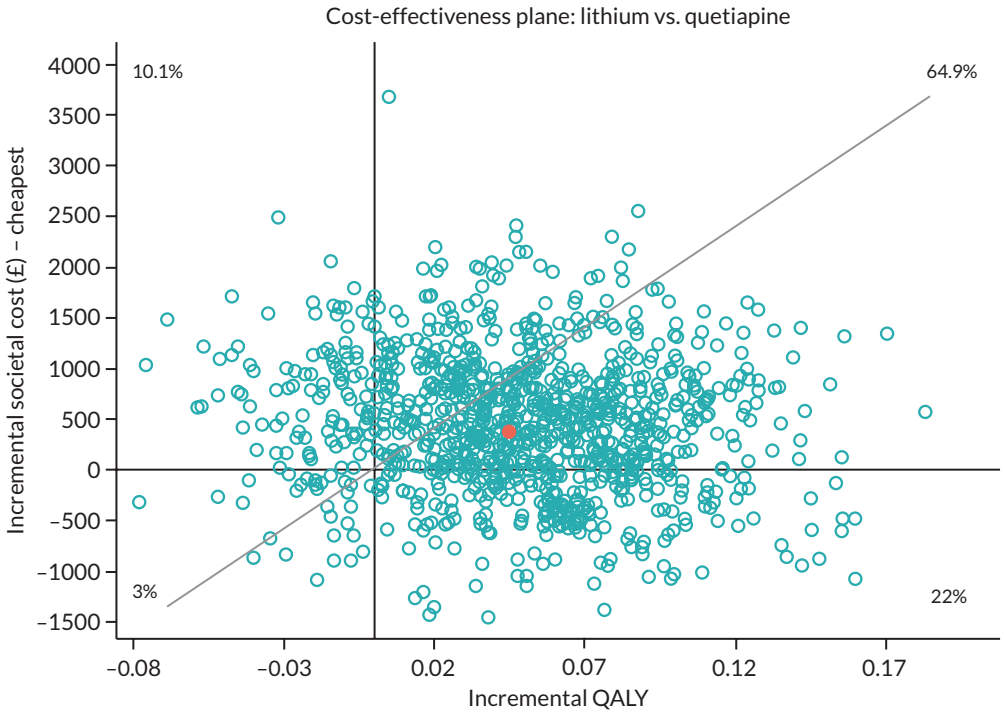

FIGURE 59 Cost-effectiveness plane of societal cost and QALY differences, assuming CD-MCAR.

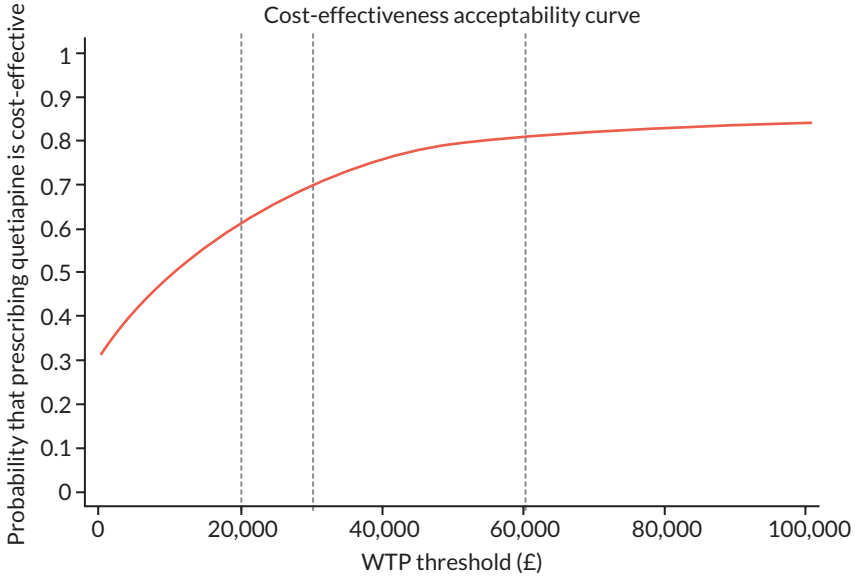

FIGURE 60 Cost-effectiveness acceptability curves showing probability that quetiapine is most cost-effective option at different WTP threshold for improvement in QALYs – Societal perspective, assuming CD-MCAR.

Sensitivity analysis 3: testing the impact of COVID-19

Primary analysis

When the impact of COVID-19 was considered, the costs of both hospital and NHS community-based services decreased after the start of the COVID-19 pandemic at the same rate in both arms, see *Table 69*.

Results of the cost-effectiveness analysis are shown in *Table 70*. Results of bootstrapping (*Figure 65*) showed that the most likely scenario with 83.1% of the re-samples was that quetiapine resulted in lower NHS and PSS costs and higher

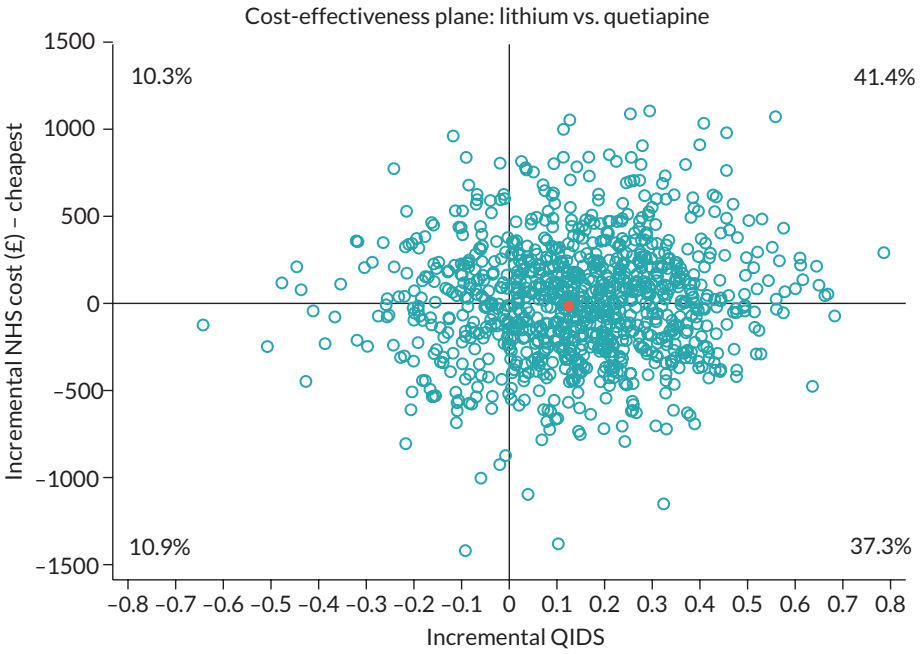

FIGURE 61 Cost-effectiveness plane of NHS and PSS cost and QIDS-SR differences, assuming CD-MCAR.

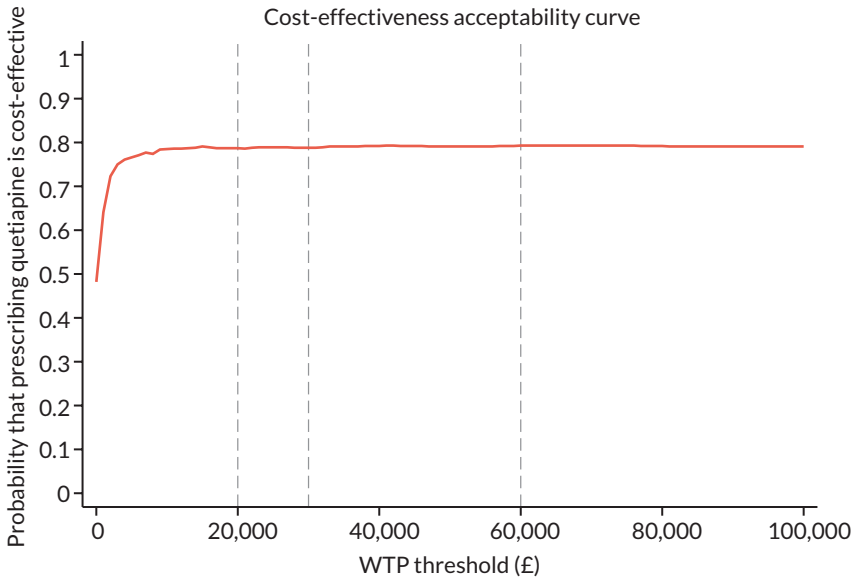

FIGURE 62 Cost-effectiveness acceptability curves showing the probability that quetiapine is the most cost-effective option at different WTP threshold for improvement in QIDS-SR score – NHS and PSS perspective, assuming CD-MCAR.

QALYs gain compared with lithium. The probability that quetiapine was the cost-effective option was 0.96 at NICE's WTP threshold of £20,000 per additional unit of QALY (*Figure 66*).

Secondary analysis

Adopting a societal perspective

The cost-effectiveness results when considering the societal perspective and the effect of COVID-19 are shown in *Table 71*. Results of bootstrapping (*Figure 67*) showed that the most likely scenario with 63.4% of the re-samples was that quetiapine resulted in higher societal costs and higher QALYs gain compared with lithium. The probability that

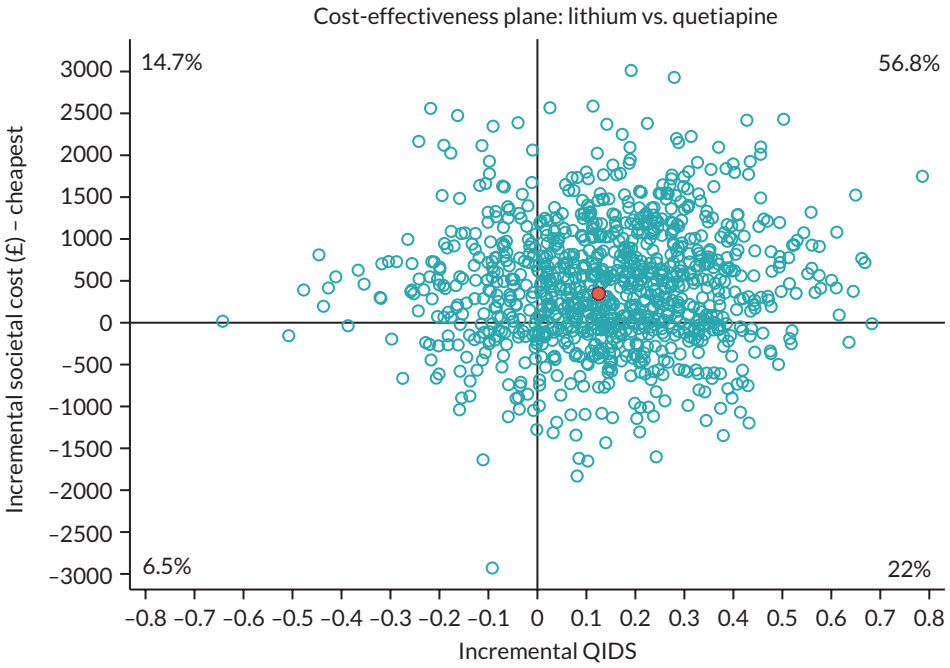

FIGURE 63 Cost-effectiveness plane of societal cost and QIDS-SR differences, assuming CD-MCAR.

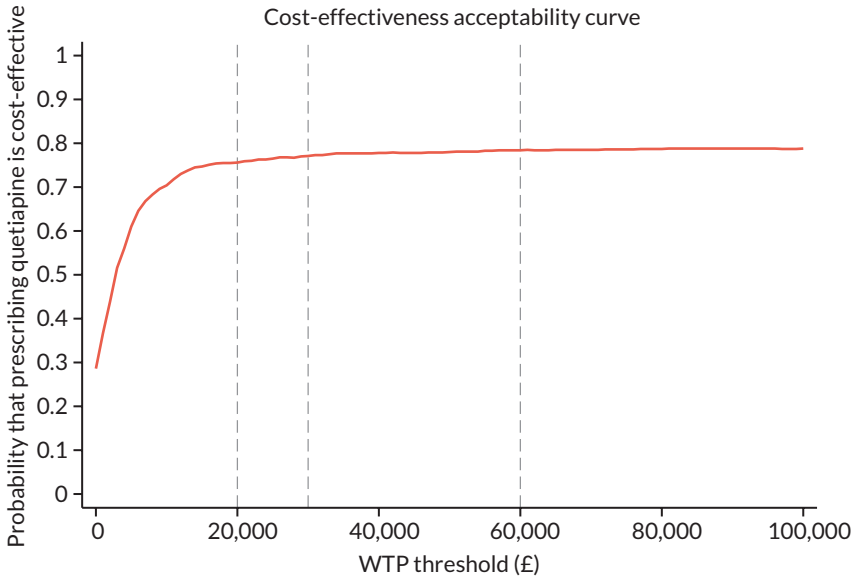

FIGURE 64 Cost-effectiveness acceptability curves showing the probability that quetiapine was the most cost-effective option at different WTP thresholds for improvement in QIDS-SR score – societal perspective, assuming CD-MCAR.

TABLE 69 Service costs before and after the COVID-19 pandemic by trial arm

	Over 52-week							
	Quetiapine (*n* = 107)				Lithium (*n* = 105)			
	Pre COVID		Post COVID		Pre COVID		Post COVID	
Cost item	*N*	Mean (SD), cost	*N*	Mean (SD), cost	*N*	Mean (SD), cost	*N*	Mean (SD), cost
Hospital services	47	1229.79 (1975.91)	29	378.21 (861.34)	38	1108.54 (1371.25)	27	963.18 (1467.25)
NHS Community based services	47	1278.20 (1172.75)	29	809.01 (813.19)	38	1553.83 (1942.73)	27	688.45 (746.07)
Loss of productivity	47	2420.74 (5703.31)	29	2436.08 (5967.12)	38	1368.76 (3544.62)	27	1256.85 (2213.28)

TABLE 70 Cost-effectiveness results from the NHS and PSS perspective using cost of cheapest drug

Intervention	NHS and PSS cost (£)	QALY	Incremental cost[a]	Incremental QALY[b]	ICER	INB
Lithium	3151.46	0.468	–	–	Dominated	–
Quetiapine	2706.77	0.540	−383.57	0.046	Dominating	0.026

a Missing value are multiple imputed and the estimated difference is adjusted for arm, EQ-5D baseline index value, baseline cost of the same item, QIDS baseline score, site, COVID19, and failure of antidepressant treatment.
b Missing values are multiple imputed and the estimated difference is adjusted for EQ-5D baseline index value, QIDS baseline score, site, COVID19, and failure of antidepressant treatment.

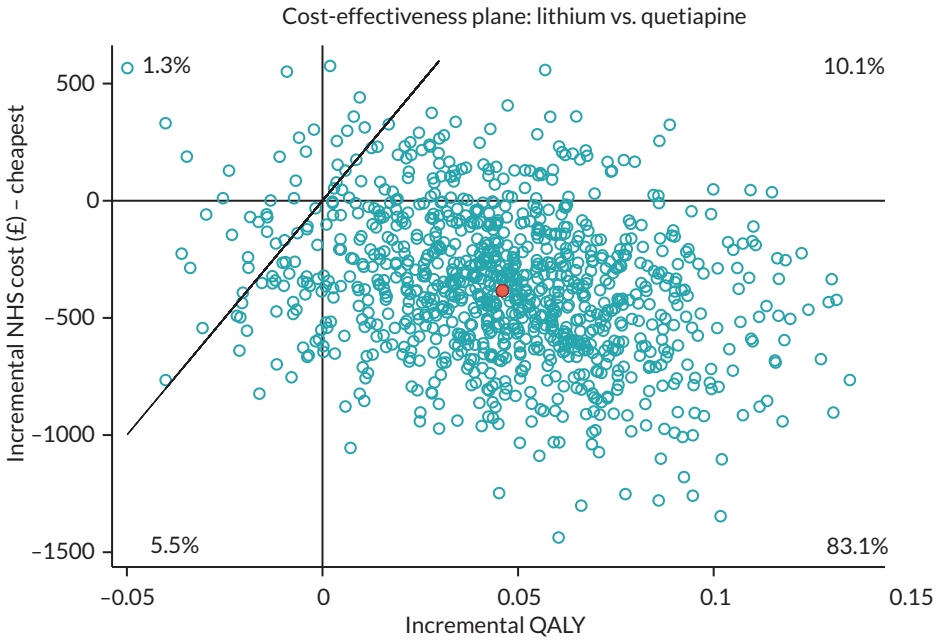

FIGURE 65 Cost-effectiveness plane of NHS and PSS cost and QALYs differences – including whether patient's follow-up overlapped with the COVID-19 pandemic as a covariant in missing data imputation.

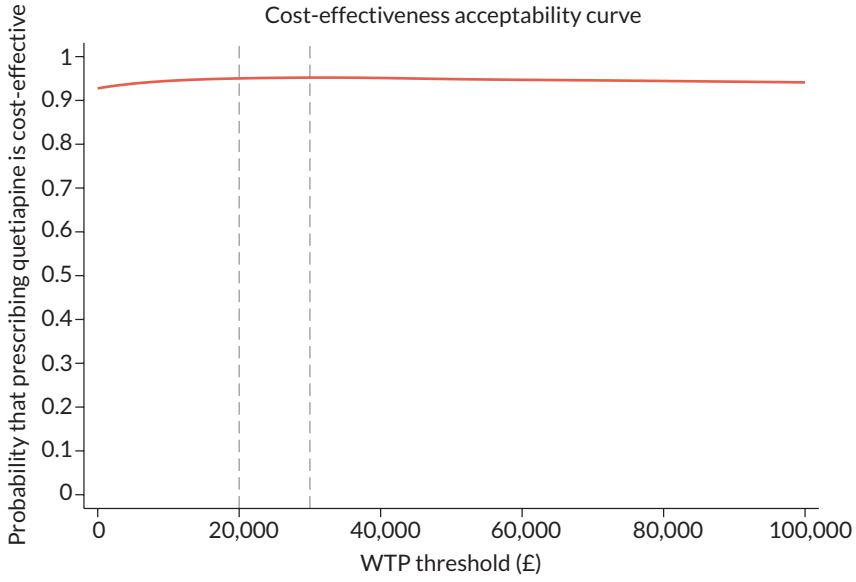

FIGURE 66 Cost-effectiveness acceptability curves showing the probability that quetiapine is the most cost-effective option at different WTP thresholds for improvement in QALYs – including whether patient's follow-up overlapped with the COVID-19.

TABLE 71 Cost-effectiveness results from the societal perspective using cost of cheapest drug

Intervention	NHS and PSS cost (£)	QALY	Incremental cost[a]	Incremental QALY[b]	ICER	INB
Lithium	3151.46	0.468	–	–	–	–
Quetiapine	2706.77	0.540	297.81	0.066	4512.27	0.051

a Missing value are multiple imputed and the estimated difference is adjusted for arm, EQ-5D baseline index value, baseline cost of the same item, QIDS baseline score, site, COVID19, and failure of antidepressant treatment.
b Missing values are multiple imputed and the estimated difference is adjusted for EQ-5D baseline index value, QIDS baseline score, site, COVID19, and failure of antidepressant treatment.

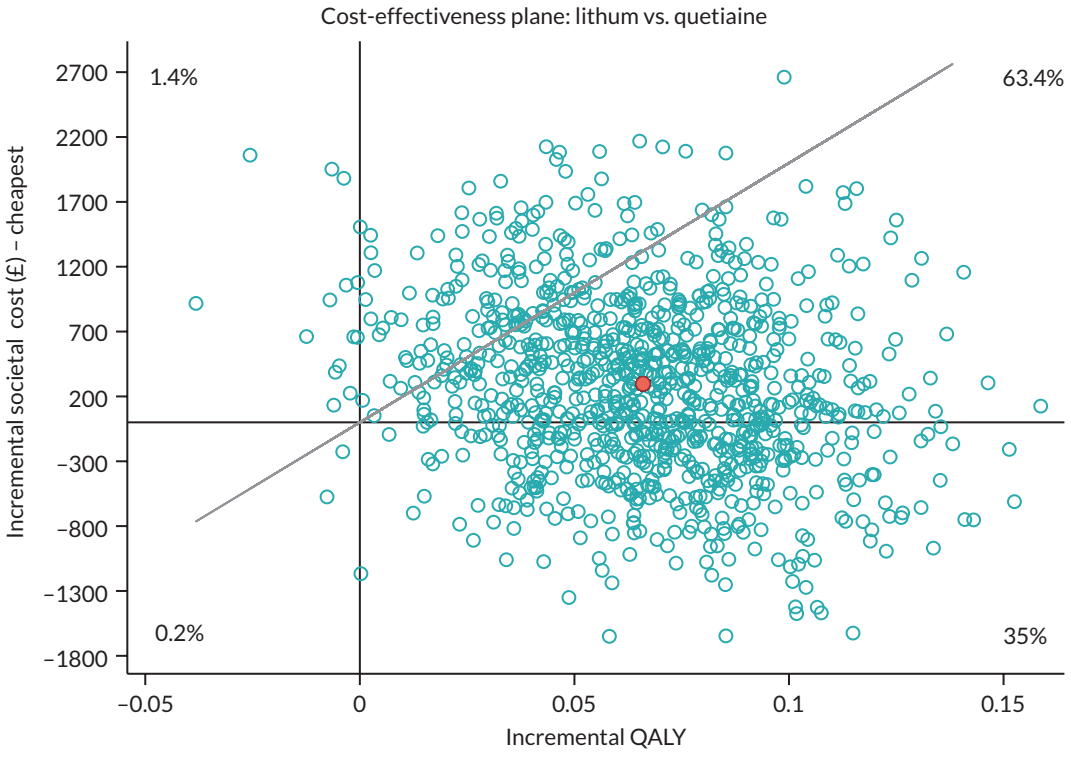

FIGURE 67 Cost-effectiveness plane of societal cost and QALYs differences – including whether participants' follow-up overlapped with the COIVD-19 pandemic as a covariant in missing data imputation.

quetiapine was the cost-effective option was 0.84 at NICE's WTP threshold of £20,000 per additional unit of QALY (*Figure 68*).

Using the self-rated Quick Inventory of Depressive Symptomatology as the effectiveness outcome

When the impact of COVID-19 was considered and the QIDS-SR was used as the effectiveness outcome, results of bootstrapping (*Figure 69*) showed that the most likely scenario with 72.8% of the re-samples was that quetiapine resulted in lower NHS and PSS costs and an improvement in QIDS-SR score compared with lithium. The probability that quetiapine was the cost-effective option was 0.92 at a WTP threshold of £0 per unit improvement in QIDS-SR score (*Figure 70*). As a unit improvement is valued at higher levels the probability decreases but remained higher than 0.79.

Adopting a societal perspective and using the self-rated Quick Inventory of Depressive Symptomatology as the effectiveness outcome

Considering the societal perspective when addressing the impact of COVID-19 and using QIDS-SR as the effectiveness outcome, results of bootstrapping (*Figure 71*) showed that the most likely scenario with 45.8% of the re-samples was

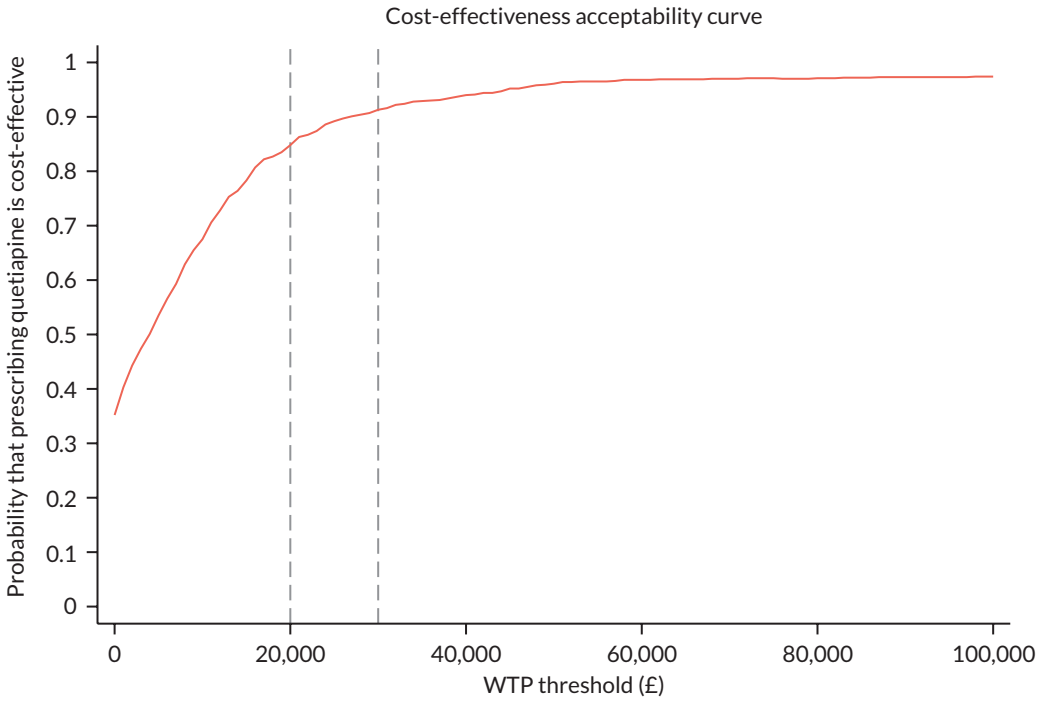

FIGURE 68 Cost-effectiveness acceptability curves showing probability that quetiapine is most cost-effective option at different WTP thresholds for improvement in QALYs – including whether participants' follow-up overlapped with the COIVD-19 pandemic.

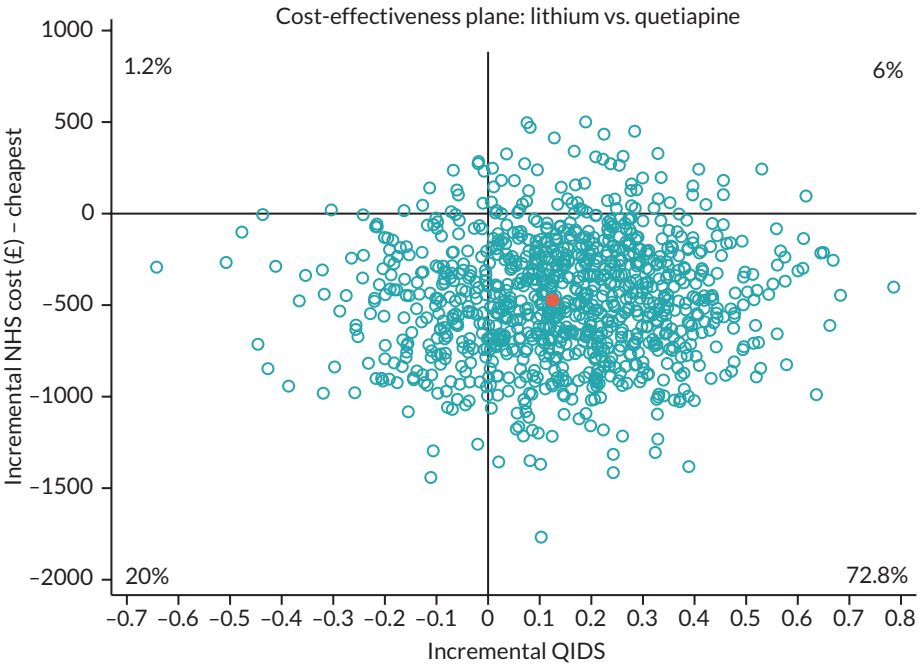

FIGURE 69 Cost-effectiveness plane of NHS and PSS cost and QIDS-SR differences – including whether participants' follow-up overlapped with the COVID-19 pandemic as a covariant in missing data imputation.

that quetiapine resulted in higher societal costs and an improvement in QIDS-SR score compared with lithium. The probability that quetiapine was the cost-effective option was 0.42 at a WTP threshold of £0 per unit improvement in QIDS-SR score (*Figure 72*). As a unit improvement was valued at higher levels the probability increased to 0.77 at a WTP threshold of £20,000.

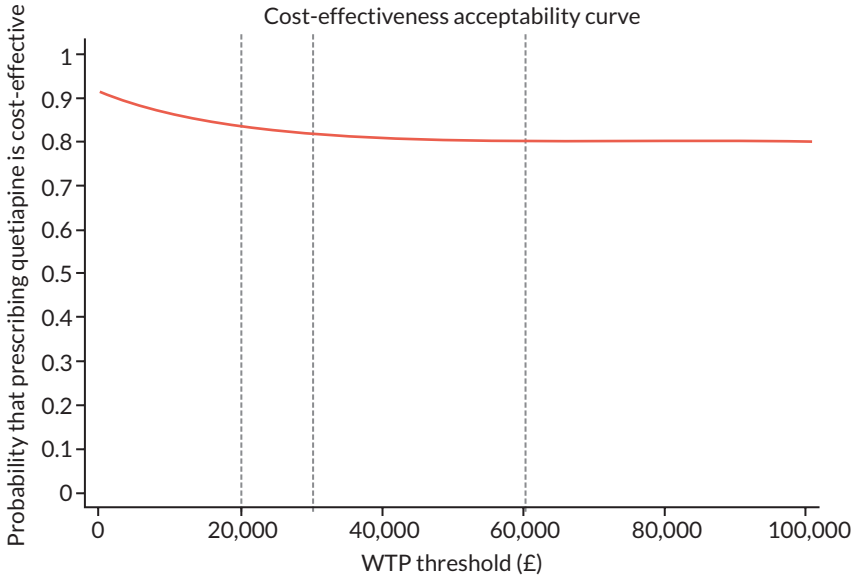

FIGURE 70 Cost-effectiveness acceptability curves showing the probability that quetiapine was the most cost-effective option at different WTP thresholds for improvement in QIDS-SR score – including whether participants' follow-up overlapped with the COVID-19 pandemic as a covariant in missing data imputation – NHS and PSS cost.

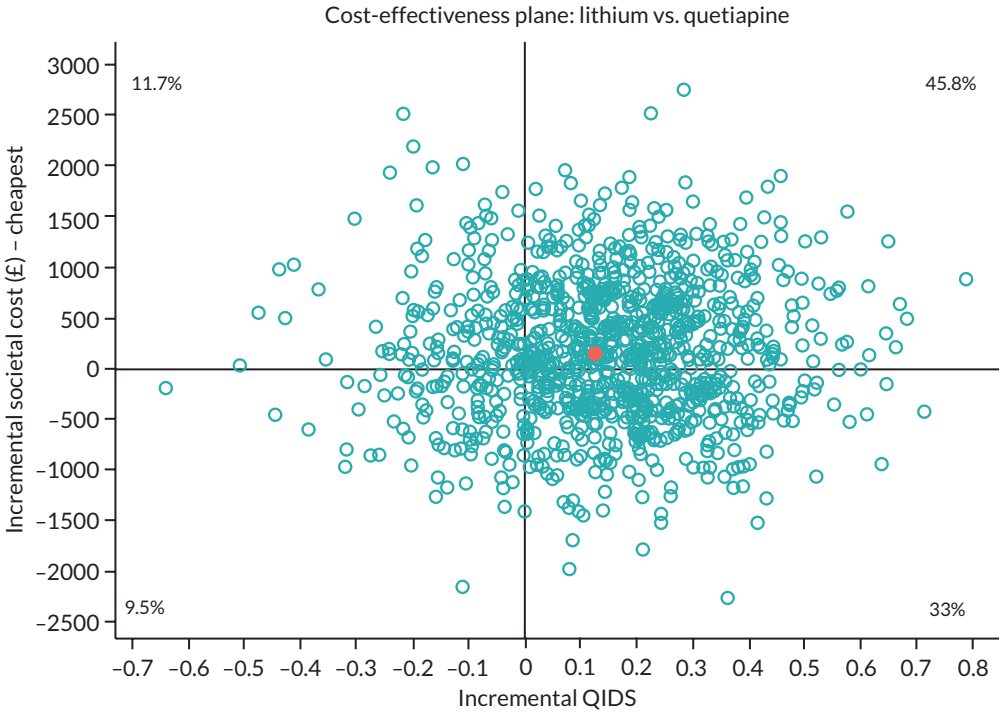

FIGURE 71 Cost-effectiveness plane of societal cost and QIDS-SR differences – including whether participants' follow-up overlapped with the COVID-19 pandemic as a covariant in missing data imputation.

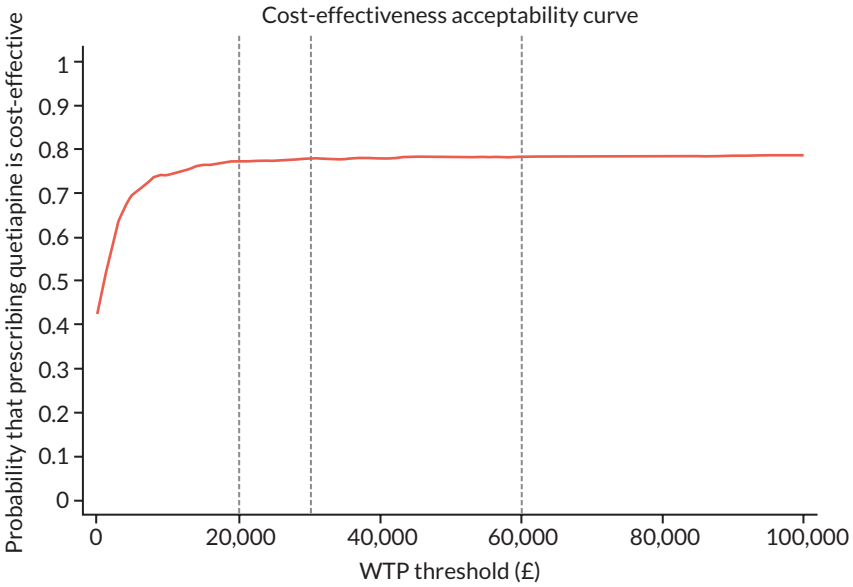

FIGURE 72 Cost-effectiveness acceptability curves showing the probability that quetiapine was most cost-effective option at different WTP thresholds for improvement in QIDS-SR score – including whether participants' follow-up overlapped with the COVID-19 pandemic as a covariant in missing data imputation – societal cost.

EME
HSDR
HTA
PGfAR
PHR

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

Published by the NIHR Journals Library