PANDA ANALYSIS PLAN

PANDA: What are the indications for Prescribing ANtiDepressAnts that will lead to a clinical benefit?

TRIAL DESIGN AND OBJECTIVES

Trial design and methods:	Patients presenting with depression in primary care where there is clinical equipoise whether SSRI treatment would be prescribed will be individually randomised between sertraline and placebo. The primary outcome will be the PHQ9 questionnaire that measures depressive symptoms at 6 weeks. The 12 week outcome will be a secondary outcome to look for more persistent effects.
Trial medication:	Sertraline vs. Placebo
Objectives:	The primary objective is to investigate the severity and duration of the depressive symptoms that are associated with a clinically important response (compared to placebo) to sertraline in people with depression. The secondary objective is to investigate quality of life, the economic cost and whether performance on emotional processing tasks are associated with response to treatment with sertraline.
Trial duration per participant:	12 weeks plus 2 weeks for tapering medication.
Planned trial sites:	Bristol, York, Liverpool, Southampton, London
Main inclusion/exclusion criteria:	Inclusion criteria: 1) Uncertainty of GP and patient about the possible benefits of antidepressants (any severity and duration of depression may be included). 2) aged 18 to 74 years (inclusive) 3) have not been treated with antidepressants in the previous 8 weeks 4) people who are having other interventions such as low intensity IAPT can also take part Exclusion criteria: 1) have other psychiatric disorders, i.e. bipolar disorder, eating disorder or psychosis 2) have major alcohol or substance abuse problem/s 3) are not able to complete the study questionnaires, including registered blind 4) women who are currently pregnant or planning pregnancy or lactating 5) people taking contraindicated medications: monoamine oxidase inhibitors and pimozide.

6) severe hepatic impairment

7) people with bleeding disorders such as haemophilia, Christmas disease and von Willebrands disease, as well as those with past medical history of bleeding gastric or duodenal ulcers or other significant bleeding disorders

PARTICIPANT FLOW

The potential participants will be recruited to a trial site by referral from GPs who will see patients either at consultation or perform a database search.

At consultation. The GP will be asked to refer potential participants by fax (or secure email) to the local trial centre giving the patients' permission for further contact if eligible. The researcher will contact the patient to confirm eligibility for the trial and arrange the baseline assessment visit.

Database searches. We will also ask GPs or their administrative staff to carry out record searches according to a Record Search SOP to identify people in whom the GP has recorded low mood, depression symptoms and are not currently on antidepressants. We will ask the GP to write to these individuals so they can consider joining the study.

At the baseline assessment the participant will complete the computerised Clinical Interview Schedule Revised (CISR) and other clinical and background information (see below). Eligibility will be confirmed and informed consent given.

STRATIFICATION AND RANDOMISATION

The randomisation will be conducted by PRIMENT CTU (http://www.ucl.ac.uk/priment) using a remote computer generated code (Sealed Envelope). The randomisation will be stratified by severity (in three categories: CISR total score <12 ,12-20, > 20) and duration of depression (< 2years, 2 years or more) and centre and with random block lengths; the randomisation list will be held by Sealed Envelope. The delegated member of staff will access a web based interface using a unique user name and password. The delegated member of staff will enter the unique study identification number for the participant and the stratification variables. The random treatment allocation will then be sent to the pharmacy. Blinded confirmation of randomisation will be printed and filed.

DATA COLLECTION
Flowchart of study assessments

Concept	Measure	Baseline	Week 2	Week 6	Week 12	Week 12 +
Socio-Demographics		X				
Medical History	Bespoke questionnaire	X				
Standardised assessment of depressive symptoms	CISR	X				
Health related quality of life	SF-12	X	X	X	X	
Depression symptoms	BDI-II, PHQ 9	X	X	X	X	
Self-reported improvement	Global Rating question	X	X	X	X	
Emotional processing	Word task(1)	X	X	X		
Anxiety	GAD 7; brief PHQ panic questions	X	X	X	X	
Quality of life	EQ-5D-5L	X	X	X	X	
Resource use	Bespoke questionnaire			X	X	
Resource use	Patients notes					X
Patient experience	Exit questionnaire					X
Adherence and dose	Morisky(2); pill counting question		X	X	X	
Concomitant Medication review	Bespoke questionnaire		X	X	X	
Physical symptoms	Modified Toronto side effects scale(3)	X	X	X	X	

Abbreviations:
CISR   Clinical Interview Schedule Revised (4)
SF12   Short Form 12 item
BDI-II   Beck Depression Inventory (version II)(5)
PHQ9   Patient Health Questionnaire 9 item (6)
GAD7   Generalised anxiety scale 7 item (7)
EQ-5D-5L   Five level version of EQ5D (8)

Baseline measures
Self-administered computerised CISR, PHQ9, BDI-II, SF12, EQ5D, GAD7. Emotional processing tasks that are sensitive to antidepressants and may be a marker of treatment response; questions about the belief in the efficacy of antidepressants; medical history and concomitant medication.

Follow-up assessments
2 weeks PHQ9, BDI-II, SF12, EQ5D, GAD7, Emotional processing tasks, modified Morisky(2) adherence measure pill count, side effects of antidepressant a modified version of the Toronto Side Effects scale as used in GENPOD(3), open ended question about adverse events, concomitant medication.

6 weeks: PHQ9, BDI-II, SF12, EQ5D, GAD7, Emotional processing tasks, modified Morisky adherence measure and pill count; side effects of antidepressant a modified version of the Toronto Side Effects scale as used in GENPOD; open ended question about adverse events, concomitant medication and health service and other resource use.

12 weeks: PHQ9, BDI-II, SF12, EQ5D, GAD7, modified Morisky adherence measure and pill count; side effects of antidepressant a modified version of the Toronto Side Effects scale as used in GENPOD; open ended question about adverse events, concomitant medication, health service and other resource use.

Resource use data
Health service and other resource data will be collected using a set of questions used previously in the CoBalT trial at 6 and 12 week follow ups. Also after 12 weeks resource use data will be extracted from GP electronic health record for economic analysis.

STATISTICAL ANALYSIS

PRELIMINARY ANALYSES

Description of practice recruitment
A summary of the characteristics of participating practices will be given, including details of the practices in the different centres (Bristol, York, Liverpool, London and Southampton): number of patients recruited, practice list size and number of full-time equivalent GPs employed.

Flow of participants through the trial

CONSORT diagrams will be produced separately for each participating centre and for the five centres combined. Separate CONSORT diagrams will be produced illustrating the case-finding process and participant throughput in the follow up phase of the study.

The case-finding CONSORT will summarise each phase of the recruitment and identification of primary care patients. We will report the following information:

Search:
- Number of patients identified by the record search
- Number of patients excluded by GPs from invitation to participate
- Number of invitations sent and a summary of the responses to this invitation (accepted invitation, declined invitation, non-response)
- Number of patients recruited at consultation
- Comparison of the distributions of age and gender variables between those identified by GPs as potential participants and those GPs excluded

Eligibility:
- The number of telephone interviews conducted
- Number of those who did not meet one or more of the screening criteria
- Number of those eligible to proceed to consenting and randomisation stages and number who refused further participation
- Number of those who could not be reached via telephone to confirm their eligibility
- Number of those whose eligibility confirmed by their GP
- Number of those who did not meet eligibility criteria

Randomisation:
- Number of those who consented to take part
- Number of those eligible but did not consent and the reasons
- Number of those who completed the baseline interview assessment
- Number of those who completed the baseline but refused randomisation and their reasons

The RCT CONSORT flow diagram will summarise participant throughput from randomisation. For each follow-up point (2, 6 and 12 weeks) the following information will also be reported:
- Number who withdrew

- Number who were lost-to-follow-up (including did not attend, unable to contact, change mind and other)
- Number who completed and who did not complete the follow-up assessment

Comparability of groups at baseline
Participants in the two treatment groups will be described separately on the following variables:

Socioeconomic characteristics
Gender
Age
Social support
Recent life events
Financial strain
Previous treatment for depression
CISR total score
ICD10 diagnoses for depression and anxiety from CISR
CISR depression severity score (sum of following sections: concentration, fatigue, sleep, depression, depressive ideas)
Depression severity (PHQ-9 score)
Duration of depressive episode from CISR

Baseline measures of the following outcome measures will also be reported:

Depression severity BDI-II score
Generalised anxiety disorder (GAD-7 score)
Health status (mental and physical component scores of SF-12)
Health quality of life outcome (EQ-5D score)

Numbers (%) will be given for binary and categorical variables, and means and standard deviations (SD) or, medians and inter-quartiles ranges (IQR) for continuous variables, as appropriate. There will be no tests of statistical significance or confidence intervals calculated between randomised groups on any baseline variable.

Description of losses to follow up
The number (%) of participants lost to follow-up at 2, 6 and 12 weeks post-randomisation will be reported for both treatment groups and compared using a chi-square test. Factors associated with 'missingness' of key trial outcomes (e.g. PHQ9 score) within the dataset at all follow-up points will be explored and described.

Description of treatments received
Summaries will be given of the study medication received for both treatment groups during the study period. This will report for both groups the number of weeks on study medication, including any change of type or dosage. Any additional antidepressant or other psychoactive medication will be described. Any specialist mental health care received during the study will also be described.

Analysis of primary outcome
Within our overall objective, we wish to test the hypotheses that severity and duration of depressive episode are related to the response to antidepressants. The primary analysis will concentrate on severity of depression.

We have chosen the PHQ9 at 6 weeks as the primary outcome. The reasons for this are as follows:
- The existing data on antidepressant effectiveness is largely based on the outcome at 6 weeks. To be consistent with existing data we have therefore chosen the 6-week measure as the primary outcome.
- There are a large number of existing scales that can assess depression severity. The Hamilton Rating Scale for Depression (HAMD) is the commonest scale used for the evaluation of antidepressants. We have not chosen to use the HAMD as this is an observer-rated scale and requires training and standardisation. Furthermore it may be susceptible to observer bias.
- Our mapping study found that the PHQ9 had a better sensitivity to change compared with the HAMD and the Beck Depression Inventory, which is the other commonly used scale.
- The PHQ9 is used as a routine outcome measure in the NHS Improving Access to Psychological Therapy (IAPT) service (www.iapt.nhs.uk); therefore results from the study are using a scale which is also used in routine outcome assessment in the NHS.

We will model the data with a linear regression model in which log PHQ9 is used as the outcome. This model will produce an estimate (once exponentiated) of the treatment effect between active and placebo that will indicate a proportionate change. This can model the increase in the treatment effect as baseline severity increases without needing a separate interaction term. This approach will simplify the analysis and moreover will make the power calculation less dependent upon rather uncertain assumptions about the size of the interaction term.

The primary outcome is the PHQ9 score at 6 weeks. The primary analysis will therefore be a linear regression of log PHQ9 score at 6 weeks adjusting for baseline PHQ9 score and the stratification variables (severity assessed by CISR, duration in 2 categories and centre). Given the aim of estimating the relationship between the CISR depression score and treatment effect we will use the continuous CISR in the model, being aware of the possibility of collinearity (we think this is unlikely given previous experience(9)). If collinearity proves to be a problem, robust multivariate techniques will be used.(10)

Testing the assumptions underlying the primary analysis
We will carry out further analyses to assess the sensitivity of any results to the modelling assumptions and the covariance structure. This will be relevant to both the primary and secondary analyses (see below). For the primary analysis we will carry out the analysis without the CISR depression severity measure in order to examine the robustness of the findings. We will also examine interactions between baseline severity and treatment outcome within the model.

We will also carry out a sensitivity analysis by adjusting for any variables that are not balanced at baseline.

We will investigate how robust the findings are to the model specification by testing whether the results are consistent with those produced by a Poisson model.

Analysis in relation to duration before randomisation of the depressive episode

7

The duration of the depressive episode is measured before randomisation using the CISR assessment at baseline. The CISR includes a question about duration after each symptom section. After the depressive symptoms section the question asks about the following categories: less than 2 weeks, between 2 weeks and 6 months, between 6 months and 1 year, between 1 and 2 years, between 2 and 5 years, between 5 and 10 years and more than 10 years. In the primary analysis described above we have proposed a method using a linear regression with a logged outcome that will allow for the expected relationship between baseline severity and treatment response without requiring an interaction term. The influence of the duration of the depressive episode will be investigated in the same way, by modelling it as another exposure variable in the model.

If we find evidence to support the hypothesis that a longer duration is associated with a larger treatment response on the proportionate scale, duration will
increase the treatment effect as a proportion irrespective of levels of baseline severity. The proposal to apply the Minimal Clinically Important Difference (MCID) analysis to the results of the trial requires the conversion of the proportionate reduction to the absolute scale. As a result, if duration has an association with outcome in the model this will change the treatment differences on the absolute scale at all levels of severity and will alter the threshold we will recommend for a beneficial effect of treatment. This is explained further below on page 10 headed "Applying the MCID estimates to the results of the trial".

Analysis of secondary outcomes
The following secondary analyses will be conducted adjusting for the baseline measure of the outcome variable, stratification and minimisation variables. We will also adjust for any variables that show imbalance at baseline as a sensitivity analysis:

- PHQ9 score as continuous outcomes at 2, 6 and 12 weeks in a repeated measures analysis
- PHQ9 score as a binary outcome where remission is defined as scoring < 10 on the PHQ9 at 2, 6 and 12 weeks in a repeated measures analysis
- BDI-II score as continuous outcome at 2, 6 and 12 weeks in a repeated measures analysis
- BDI-II score as binary outcome (BDI-II <10) at 2, 6 and 12 weeks in a repeated measures analysis
- GAD7 scores as continuous outcome at 2, 6 and 12 weeks in a repeated measures analysis
- SF12 physical and mental component scores at 2, 6 and 12 weeks in a repeated measures analysis
- Self-reported global improvement at 2, 6 and 12 weeks in a repeated measures analysis
- Change in Emotional processing tasks between baseline and 2 weeks in relation to PHQ9 at 6 and 12 weeks in a repeated measures analysis

Choice of regression model will depend on the outcome. Distributional assumptions appropriate for positive continuous outcomes will be investigated including the lognormal or Poisson distributions. The SF12 is usually modelled using linear regression as this is often a good fit to the data. Logistic regression will be used for the analysis of binary outcome data.

Clustering by practice
Exploration of possible clustering by practice will be carried out by calculating the intra-cluster correlation coefficient across practices using random effects regression models before and after adjustment for baseline PHQ9 score. If there is evidence of practice clustering we will then repeat the

8

primary analysis allowing for practice effects by for example creating a random effect for practice. For repeated measures analysis, the practice effect will be at a higher level than the random effect for individuals. We expect to have well in excess of 100 practices so any impact on the estimates will be small. In previous studies with individual randomisation we have not found any evidence for clustering by practice.(11, 12) The pharmacological intervention will also not vary according to practice. We therefore do not expect to find any clustering and have not taken this into account in the power calculation.

Approach to dealing with missing outcome data

We will carry out sensitivity analyses to investigate the possible impact of missing data.(13) The two main approaches will be to adjust for baseline variables associated with missing outcome data and also to use multiple imputation. We will of course be making strenuous efforts to minimise the amount of missing data, and in our power calculations have estimated that there will be up to 10% missing data at 6 weeks. In GENPOD we obtained 91% follow up at 6 weeks.(14)

Economic evaluation

Unit costs will be obtained from published national sources where possible, including Curtis (www.pssru.ac.uk), Department of Health reference costs (www.dh.gov.uk) and the BNF (www.bnf.org). Cost will be expressed in £ sterling, valued at current prices, adjusted for the GDP deflator where necessary. The cost-effectiveness analysis will evaluate the efficiency of sertraline according to baseline severity and duration of symptoms. We will calculate net monetary benefit (NMB)(15) of treatment for individual patients, based on standard NICE willingness to pay thresholds. The advantage of using the NMB approach rather than a more conventional incremental cost effectiveness ratio is that we can use net benefit regression to explore the relationship between symptom severity and duration and economically important differences in patient costs and outcomes over the 12 week period. In this circumstance we will use linear regression including an interaction term with severity as the NMB outcome is best considered on an absolute scale. We will investigate the assumptions underlying the model and consider other relationships if appropriate.

The economic analysis will provide a threshold for when it becomes economically appropriate to use antidepressants, based on the current willingness to pay threshold. The recommended threshold will correspond to the point that the relationship between baseline severity (or duration) and the net benefit outcome is zero. This threshold may well be different from that estimated below using the data on Minimal Clinically Important Difference (MCID).

Power Calculation

We have revisited the power calculation in the light of our results from the Value of Information study and PANDA cohort study.

As we have decided to use a linear model with a log link function, we can now analyse the data without using an interaction term. The results from previous meta-analyses suggest that the effect size is about an 11% reduction in HAMD score (calculated as part of the Value for Information study). In the Fournier meta-analysis, the reduction is about 17% (personal communication, Jay Fournier). We have therefore taken the 11% estimate as the more conservative option in order to inform our power calculation. Our best estimate of the Minimal Clinically Important Difference (MCID) from the PANDA cohort study is that this corresponds to a 14 percentage points (95% CI 10 to 17 percentage points) reduction in score on the PHQ9. Therefore giving the power for effect sizes of 11 and 14 percentage points is reasonable, and if

anything conservative in the light of the confidence limits and the previous results from the systematic review.

An 11 percentage point difference corresponds to a difference in proportions of 0.89 and this is -0.117 on the natural logarithm scale. Similarly, a 14% difference corresponds to 0.86 and -0.15 on the natural logarithm scale. We have estimated the SD of the logarithm of PHQ9 scores from existing data from the PANDA cohort study in a Poisson regression in which the follow-up PHQ9 scores are the outcome and the baseline scores are an offset variable. This led to an estimate of SD of 0.32 to 0.34 for the log PHQ9. Given the uncertainty in estimating SDs we have also included estimates assuming an SD of 0.4. The following table gives an estimate of required total sample sizes according to the assumptions in the table in order to provide 90% power at the two sided 5% significance level.

Percentage point reduction	Natural log of reduction	SD	Total Sample size	Allowing for 10% attrition
11	-0.117	0.34	366	407
14	-0.150	0.34	216	240
11	-0.117	0.4	492	547
14	-0.150	0.4	300	333

Given the lack of information on duration, it is difficult to estimate power for the analysis with duration. This analysis is likely to have similar power as the severity analysis since it will also not rely on an interaction term in the regression model. In any case, duration is a secondary analysis so we propose to have a target of 547 participants given the above calculations.

Applying the Minimal Clinically Important Difference (MCID) estimates to the results of the trial
We propose to perform the analysis of the trial and will then use information on the MCID to make inferences about the likely threshold above which antidepressants are likely to be effective.

Even though we will be modelling our data from the trial within a proportionate framework, for this phase we will express the treatment differences on an absolute PHQ9 scale according to the baseline severity estimated using the CISR assessment. As mentioned previously, we envisage that any guidelines generated by this research will require standardised information such as that provided by the CISR, rather than information from an unstructured clinical assessment. We therefore plan to carry out this phase of the work using the CISR depression severity score as the baseline measure of depression severity. We will use the estimates from the regression and convert these to the absolute treatment difference (in PHQ9) according to baseline severity (in CISR).

From the cohort (and previous) data we have found that the value of MCID varies according to the baseline severity of depression. Our results from the PANDA cohort study can be used to plot the absolute value of MCID in PHQ9 points according to the baseline severity in CISR. Our results indicate that even if the MCID is modelled as a proportion, there is an increase in this proportion as the baseline severity reduces. We can then use these absolute values of the MCID from the cohort study and apply them to the results from the trial. When the MCID is the same size as the treatment difference then this value of baseline severity will correspond to the point above which treatment will be regarded as beneficial. We can compute the treatment difference and MCID in relation to the CISR depression

severity measure from both the trial and the cohort study so the recommended threshold would be in terms of the CISR depression severity measure.

The stakeholder meeting feedback report suggested that it may be easier to use categories of baseline severity rather than treating severity as a continuous variable. We do not think this will be an advantage and propose to use baseline severity as a continuous measure. The continuous analysis will allow for the threshold for predicting benefit to be anywhere along the spectrum of severity, rather than constrained *a priori* by the choice of a limited number of severity categories.

If we confirm our hypothesis that duration also affects treatment response we would need to repeat this exercise at a selection of values of duration. The results of this would be a combination between a severity score and duration that would correspond to the point above which treatment would be beneficial. If duration is a main effect in the regression model, this would also affect the value of the treatment difference at different durations, once it is converted to absolute differences in PHQ9 score. We would therefore also use this information to calculate recommended thresholds at different durations. If these recommendations were to be implemented, the software that administers the CISR could also compute a recommendation using both the baseline severity and duration.

This approach could also, in the future, allow other factors related to prognosis to contribute towards recommendations about antidepressant use. The use of 'multiplicative' relationships to combine risk factors, followed by estimation of absolute differences in risk, is similar to that used in targeting medication in the prevention of cardiovascular disease.(16)

What was learnt from the first phases of the programme

The stakeholder panel also asked us to illustrate what we had learnt from the early phases of the programme that has influenced our approach towards the trial.

1)   As noted above, we have estimated the MCID and how it varies with baseline severity from the cohort study. This result has been used to guide our power calculation and will also be used after the analysis of the trial in order to provide guidance on the likely threshold of benefit.
2)   We have investigated the likely advantages of modelling the results of the trial using a proportionate model rather than one that estimates the absolute difference in scores. This will simplify the analysis.
3)   We have recalculated the statistical power of the study based on our proposed method of analysis and informed by our estimates of the MCID.
4)   The qualitative results have supported the use of the self-reported global improvement question in addition to the PHQ9, as it highlights different aspects of the patient experience. We have therefore included this as a secondary outcome measure in the RCT. Preliminary quantitative data on the mismatch between self-reported global improvement and the PHQ9 indicate that the percentage of disagreement was relatively modest, and even then some of it could be explained by measurement error.
5)   The qualitative studies have illustrated that the concept of an MCID is an 'average' and individuals are making a complex judgement in deciding on improvement that incorporates a wide variety of information that takes account their own personal context and history.
6)   We have developed further experience of working across several sites in preparation for the RCT fieldwork. We have established effective working relationships with the local research networks and

with research-oriented primary care teams, who are keen to continue to work with us during the next (trial) phase of PANDA.

Feasibility and implementation of standardised assessments in primary care
The panel noted there were concerns that the implementation of guidance based on standardised assessments would be a departure from current clinical practice. The computerised CISR assessment we are using in this study provides more detailed and clinically relevant information than the brief questionnaires. If the results from the CISR prove useful in guiding prescription of antidepressants, we envisage that GPs would then ask patients to complete the self-administered computerised CISR outside the consultation and then return to discuss the results at a future appointment. The computerised assessment would then be used like a blood test in order to guide appropriate recommendations by the doctor.

During this study we would plan to carry out brief telephone-based qualitative interviews with GPs in order to get some preliminary data on the likely feasibility and acceptability of our recommended approach. After this study has finished, and contingent upon promising results, we would seek additional funds to carry out further feasibility work around implementation. However, we also recognise that GPs tend to embrace change as long as they are convinced it will benefit patients. Therefore one of the key elements in successful implementation will be to provide good evidence regarding any such benefits based on the RCT.

REFERENCES

1.   Harmer CJ, Shelley NC, Cowen PJ, Goodwin GM. Increased positive versus negative affective perception and memory in healthy volunteers following selective serotonin and norepinephrine reuptake inhibition. AmJPsychiatry. 2004;161(7):1256-63.
2.   George CF, Peveler RC, Heiliger S, Thompson C. Compliance with tricyclic antidepressants: the value of four different methods of assessment. British journal of clinical pharmacology. 2000;50(2):166-71.
3.   Crawford AA, Lewis S, Nutt D, Peters TJ, Cowen P, O'Donovan MC, et al. Adverse effects from antidepressant treatment: randomised controlled trial of 601 depressed individuals. Psychopharmacology (Berl). 2014;231(15):2921-31.
4.   Lewis G, Pelosi AJ, Araya R, Dunn G. Measuring psychiatric disorder in the community: a standardised assessment for use by lay interviewers. Psychological Medicine. 1992;22:465-86.
5.   Beck AT, Ward CH, Mendelsohn M, Mock J, Erbaugh J. An inventory for measuring depression. Archives of General Psychiatry. 1961;4:561-71.
6.   Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. J Gen Intern Med. 2007;22(11):1596-602.
7.   Spitzer RL, Kroenke K, Williams JB, Lowe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med. 2006;166(10):1092-7.
8.   Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2011;20(10):1727-36.
9.   Wiles NJ, Mulligan J, Peters TJ, Cowen PJ, Mason V, Nutt D, et al. Severity of depression and response to antidepressants: GENPOD randomised controlled trial. Br J Psychiatry. 2012;200(2):130-6.
10.   White IR, Thompson SG. Adjusting for partially missing baseline measurements in randomized trials. Stat Med. 2005;24(7):993-1007.

11.	Chalder M, Wiles NJ, Campbell J, Hollinghurst SP, Haase AM, Taylor AH, et al. Facilitated physical activity as a treatment for depressed adults: randomised controlled trial. BMJ. 2012;344:e2758.
12.	Wiles N, Thomas L, Abel A, Ridgway N, Turner N, Campbell J, et al. Cognitive behavioural therapy as an adjunct to pharmacotherapy for primary care based patients with treatment resistant depression: results of the CoBalT randomised controlled trial. Lancet. 2013;381:375-84.
13.	White IR, Kalaitzaki E, Thompson SG. Allowing for missing outcome data and incomplete uptake of randomised interventions, with application to an Internet-based alcohol trial. Statistics in Medicine. 2011;30(27):3192-207.
14.	Lewis G, Mulligan J, Wiles N, Cowen P, Craddock N, Ikeda M, et al. Polymorphism of the 5-HT transporter and response to antidepressants: randomised controlled trial. Br J Psychiatry. 2011;198:464-71.
15.	Zethraeus N, Johannesson M, Jonsson B, Lothgren M, Tambour M. Advantages of using the net-benefit approach for analysing uncertainty in economic evaluation studies. Pharmacoeconomics. 2003;21(1):39-48.
16.	Blood Pressure Lowering Treatment Trialists C, Sundstrom J, Arima H, Woodward M, Jackson R, Karmali K, et al. Blood pressure-lowering treatment based on cardiovascular risk: a meta-analysis of individual patient data. Lancet. 2014;384(9943):591-8.