COMPARATIVE EFFICACY AND ACCEPTABILITY OF FIRST- AND SECOND-
GENERATION ANTIDEPRESSANTS IN THE ACUTE TREATMENT OF MAJOR
DEPRESSION: A MULTIPLE TREATMENTS META-ANALYSIS

Andrea Cipriani, Corrado Barbui, Toshiaki Furukawa, Georgia Salanti, Stefan Leucht,
Eric Ruhe, Sarah Stockton, Julian Higgins, Guy Goodwin, John Geddes

PROTOCOL

VERSION 3.1 – MARCH 2012
**BACKGROUND**

Major depressive disorder is the most prevalent psychiatric disease in the general population, affecting more than 16% of adults during their lifetime (Kessler et al., 2003). In 2000 the economic burden of depressive disorders in the US was estimated to be around 80 billion dollars, with more than 30% of these costs being attributable to direct medical expenses (Greenberg et al., 2003). Pharmacotherapy plays an important role in the management of major depression. Before the late 1980s, pharmacologic treatment was limited to tricyclic antidepressants (TCAs) and monoamine oxidase inhibitors (MAOIs). TCAs and MAOIs sometimes are referred to as traditional or first-generation antidepressants. These drugs are often accompanied by multiple side effects that many patients find intolerable. TCAs tend to cause anticholinergic effects including dry mouth and eyes, urinary hesitancy, and sometimes retention and constipation and MAOIs have the potential to produce hypertensive crises if taken along with certain foods or dietary supplements containing tyramine. However, even though first-generation antidepressants are no longer agents of choice in many circumstances, TCAs are still used worldwide, most of all in low and middle income countries (according to the list of essential medicines issued by the World Health Organization amitriptyline is one of the two available treatment options for major depression, along with an SSRI fluoxetine -WHO 2011). Newer antidepressant treatments include selective serotonin reuptake inhibitors (SSRIs), serotonin and norepinephrine reuptake inhibitors (SNRIs), and other second-generation drugs. The first of the second-generation drugs was introduced to the US market in 1985, when bupropion was approved for the treatment of major depressive disorders. In 1987, the US Food and Drug Administration (FDA) approved the first SSRI, fluoxetine. Since then, five other SSRIs have been introduced in the market between 1991 and 2002: sertraline, paroxetine, citalopram, fluvoxamine and escitalopram. The SNRIs were first introduced in 1993 with the approval of venlafaxine. In 1994, nefazodone, which is essentially an SSRI with additional 5-hydroxytryptamine-2 (5-HT2) and 5-hydroxytryptamine-3 (5-HT3) antagonist properties, was FDA-approved. Mirtazapine, a drug that exhibits both noradrenergic and serotonergic activity with central autoreceptors, was added in 1996 and duloxetine, a serotonin and norepinephrine reuptake inhibitor, was approved for the treatment of MDD (and diabetic peripheral neuropathic pain) in 2004. The latest second-generation antidepressants approved for the treatment of MDD in adults were desvenlafaxine, the major active metabolite of venlafaxine XR, and agomelatine, a melatonergic agonist with 5-HT2 antagonism.

Several systematic reviews have assessed the comparative efficacy and safety of second-generation antidepressants but two recent comparative effectiveness reviews have provided the most comprehensive assessments to date, notwithstanding conflicting interpretation of results (Cipriani et al., 2009; Gartlehner et al., 2011).
Multiple treatments meta-analysis (MTM) is a statistical technique that allows both direct and indirect comparisons to be undertaken, even when two of the treatments have not been directly compared (Salanti et al., 2009). MTM can summarise randomised controlled trials (RCTs) of several different treatments providing point estimates (together with 95% credibility intervals [CIs]) for their association with a given endpoint, as well as an estimate of inconsistency (that is, a measure of how well the entire network fits together, with small values suggesting better internal agreement of the model). MTM has already been used successfully in other fields of medicine (Psaty et al., 2003; Elliott et al., 2007) and psychiatry (Cipriani et al., 2009; Cipriani et al., 2011). The present review will be based on our previous MTM on antidepressants (Cipriani et al., 2009), but we will carry out a different project enlarging the number of antidepressants under investigation, adding new and clinically informative outcome measures and, most of all, including also placebo controlled trials.

**OBJECTIVES**

To compare individual first- and second-generation antidepressants in terms of efficacy and acceptability in the acute treatment of major depression to better inform clinical practice and mental health policies.

**METHODS**

*Criteria for considering studies for this review*

**Types of studies**

All double-blind RCTs comparing one drug with another within the following group of selected first- and second-generation antidepressants (namely, agomelatine, amitriptyline, bupropion, citalopram, clomipramine, desvenlafaxine, duloxetine, escitalopram, fluoxetine, fluvoxamine, hypericum, milnacipran, mirtazapine, paroxetine, reboxetine, sertraline, trazodone, venlafaxine and vilazodone) or with placebo, as monotherapy, in the acute phase treatment of depression will be included. RCTs in which antidepressants were used as an augmentation strategy will be excluded. Quasi-randomized trials (such as those allocating by using alternate days of the week) will be excluded. For trials which have a crossover design only results from the first randomisation period will be considered. Cluster randomised trials will be included.

**Types of participants**

Patients aged 18 or older, of both sexes with a primary diagnosis of major depression. Studies adopting any standard operationalised criteria to define patients suffering from unipolar major depression will be included, such as Feighner criteria, Research Diagnostic Criteria, DSM-III, DSM-
III-R, DSM-IV and ICD-10. Studies in which less than 20% of the participants may be suffering from bipolar depression will be included. A concurrent secondary diagnosis of another psychiatric disorder will not be considered as exclusion criteria. RCTs in which all participants have a concurrent primary diagnosis of another Axis I or II disorders will be excluded. Studies in which all participants have a diagnosis of resistant depression will be excluded. Antidepressant trials in depressive patients with a serious concomitant medical illness will be excluded. RCTs of women with post-partum depression will be also excluded, because post-partum depression appears to be clinically different from major depression (Cooper & Murray, 1998). Trials which allow rescue medications will be included so long as they are equally provided among the randomised arms.

**Outcome measures**

- **Primary outcomes:**

  1. *Efficacy (as dichotomous outcome) - Response*

     Measured by the total number of patients who had a reduction of at least 50% on the total score between baseline and week 8 (range 4 to 12 weeks) on a standardized rating scale for depression (Hamilton Depression Rating Scale (HDRS) or another standardised rating scale, if HDRS was not used). Any version of HDRS will be accepted.

  2. *Acceptability of treatment*

     Treatment discontinuation (acceptability) is defined as the proportion of patients who leave the study early for any reason during the first 8 weeks of treatment (range 4 to 12 weeks).

- **Secondary outcomes:**

  3. *Efficacy (as continuous outcome)*

     Measured by the mean change on the HDRS or Montgomery-Åsberg Depression Rating Scale (MADRS), if HDRS was not used, after 8 weeks (range 4 to 12 weeks). If none of the former scales was used, we will consider other standardised rating scales.

  4. *Efficacy (as dichotomous outcome) - Remission*

     Measured by the total number of patients who had a remission of depressive symptoms between baseline and week 8 (range 4 to 12 weeks) on a standardized rating scale for depression (HDRS or another standardised rating scale, if HDRS was not used). The remission will usually be defined as =<7 or 8 on the 17-item HDRS or as =<10 or =<11 on the MADRS.

  5. *Tolerability of treatment*

     The proportion of patients who leave the study early due to adverse events during the first 8 weeks of treatment (range 4 to 12 weeks).
Search strategy

All published and unpublished RCTs that compared the efficacy and acceptability of one antidepressant with another (see the list of included antidepressants here above) or placebo in the treatment of major depression will be identified by searching the Cochrane Collaboration CENTRAL register, AMED, CINAHL, EMBASE, LilACS, MEDLINE, UK National Research Register, PSYCINFO and PSYNDEX databases.

Trial databases of the following drug-approving agencies (the Food and Drug Administration in the USA, the Medicines and Healthcare products Regulatory Agency in the UK, the European Medicines Agency in the EU, the Medicines Evaluation Board in the Netherlands, the Medical Products Agency in Sweden, the Pharmaceuticals and Medical Devices Agency (PMDA) in Japan, the Therapeutic Goods Administration (TGA) in Australia) and ongoing trial registers (clinicaltrials.gov in the USA, ISRCTN and National Research Register in the UK, the Netherlands Trial Register, EUDRACT in the EU, the UMIN-CTR, JapicCTI and JMACCT in Japan, the Australian Clinical Trials Registry and the WHO International Clinical Trials Registry Platform) will be hand-searched for published, unpublished and ongoing controlled trials. The National Institute for Clinical Excellence (UK) and the Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (Germany) will also be contacted for additional information. No language restrictions will be applied. The following phrase will be used: [depress* or dysthymi* or adjustment disorder* or mood disorder* or affective disorder or affective symptoms] and combined with a list of all included antidepressants. The list will be supplemented by reference search and personal contacts. All relevant authors will be contacted to supplement the incomplete report of the original papers.

We are aware that there are many trials carried out in China (Chakrabarti et al., 2007). However, for many of these studies only incomplete or conflicting information is available and it has been reported many of them do not use appropriate randomisation procedures (Wu et al., 2006). In an effort to avoid the potential biases that may be introduced by including these trials without further information, we will exclude these studies.

Study selection and data extraction

Two persons will independently review references and abstracts retrieved by the search. If both reviewers agree that the trial doesn’t meet eligibility criteria, we will exclude it. We will obtain the full text of all remaining articles and use the same eligibility criteria to determine which, if any, to exclude at this stage. Any disagreements will be solved via discussion with a third member of the reviewing team.

Two reviewers will then independently read each article, evaluate the completeness of the data abstraction, and confirm the quality rating (see details below). We will design and use a structured
data abstraction form to ensure consistency of appraisal for each study. Information extracted will include study characteristics (such as lead author, publication year, journal), participant characteristics (such as diagnostic criteria for depression, age range, setting, diagnosis of bipolar depression), intervention details (such as dose ranges, mean doses of study drugs) and outcome measures. A double-entry procedure will be employed by two reviewers.

**Length of trial**

It is a problem of systematic reviews that usually trials have different durations. Clinically, the assessment of efficacy after 8 weeks of treatment or after 16 to 24 weeks or more may lead to differences in terms of treatment outcome. Clinicians need to know whether (and to what extent) treatments work within a clinically reasonable period of time. Unfortunately, there is no consensus on what the appropriate duration of an acute phase trial is. In the present review, acute treatment will be defined as an 8-week treatment in both the efficacy and acceptability analyses (Bauer et al., 2002). If 8-week data are not available, we will use data ranging between 4 to 12 weeks and the time point given in the original study as the study endpoint will be given the preference. Longer-term studies will be excluded if they do not provide data for the 4-12 weeks period.

**Quality Assessment**

We will assess risk of bias in the included studies using the tool described in the Cochrane Collaboration Handbook as a reference guide (Higgins et al., 2011). Where inadequate details of allocation concealment and other characteristics of trials are provided, the trial authors will be contacted in order to obtain further information. We will not include studies where sequence generation was at high risk of bias and where allocation was clearly not concealed. The quality assessment will be done by two independent raters. If the raters disagree, the final rating will be made by consensus with the involvement (if necessary) of another member of the review group.

**Comparability of dosages**

We will include only studies randomizing patients to drugs within the therapeutic dose (both fixed-dose and flexible-dose designs will be allowed) (Cipriani et al., 2009). There is the possibility that some trials compare one agent at the upper limit of its therapeutic range with another agent at the lower limit of its therapeutic range within the same study. We may look at heterogeneity and then add a variable (yes/no) that report if dosages are comparable and use this information for sensitivity analysis.
STATISTICAL ANALYSIS

Considering that clinical trials of antidepressant drugs are usually small and that data distribution is difficult to assess for studies with small samples, in this review priority will be given to the use and analysis of dichotomous variables both for efficacy and acceptability. When dichotomous efficacy outcomes are not reported but baseline mean and endpoint mean and standard deviation of the depression rating scales (such as HDRS or MADRS) are provided, we will calculate the number of responding patients at 8 weeks (range 4 to 12 weeks) employing a validated imputation method (Furukawa et al., 2005). We are aware that other methods to impute response rate are available and have been investigated (Anzures-Cabrera et al., 2011). Even though these imputation methods are valid and may give odds ratios (ORs) with narrower CIs, they only produce logORs and their variances rather than raw data. As we opt for 2x2 tables to model using the binomial likelihood, the Fukurawa method will be used in our review. We will use for imputation the endpoint scores for the following reasons: (i) standardised mean difference should focus on standard deviation of endpoint scores (standard deviation of change does not represent population variation); (ii) reporting change may represent outcome reporting bias; (iii) we would need to make up more data to impute standard deviation of change scores; (iv) observed standard deviation of change is about the same as observed standard deviation of endpoint. Where outcome data or standard deviations are not recorded, authors will be asked to supply the data. When only the standard error or t-statistics or p values are reported, standard deviations will be calculated according to Altman (Altman, 1996). In the absence of data from the authors, the mean value of known standard deviations will be calculated from the group of included studies according to Furukawa and colleagues (Furukawa et al., 2006). We will check that the original standard deviations are properly distributed, so that the imputed standard deviation represent the average. The continuous efficacy outcome of this review will be the endpoint score of the HDRS (or MADRS). Dichotomous outcomes will be analysed on an intention-to-treat (ITT) basis: drop-outs will be assumed to have had negative outcomes and will always be included in this analysis. When data on drop-outs are carried forward and included in the evaluation (Last Observation Carried Forward, LOCF), they will be analysed according to the primary studies.

Synthesis of results

We will generate descriptive statistics for trial and study population characteristics across all eligible trials, describing the types of comparisons and some important variables, either clinical or methodological (such as year of publication, age, severity of illness, sponsorship, clinical setting). For each pair-wise comparison between antidepressants, the standardized mean difference (SMD) or Hedges’s adjusted g will be calculated as the effect size for continuous outcomes and the OR will be
calculated for dichotomous outcomes, both with a 95% CI. We will first perform pair-wise meta-
analyses by synthesizing studies that compare the same interventions using a random effects model
(DerSimonian & Laird, 1986) to incorporate the assumption that the different studies are estimating
different, yet related, treatment effects (Higgins & Green, 2006). Visual inspection of the forest plots
will be used to investigate the possibility of statistical heterogeneity. This will be supplemented
using, primarily, the I² statistic. This provides an estimate of the percentage of variability due to
heterogeneity rather than a sampling error (Higgins et al., 2003). 95% confidence intervals will be
calculated for I², and a p value from a standard Q-test for heterogeneity will be used to assess
evidence of its presence.

We will conduct a MTM. MTM is a method of synthesizing information from a network of trials
addressing the same question but involving different interventions. For a given comparison, say A
versus B, direct evidence is provided by studies that compare these two treatments directly. However, indirect evidence is provided when studies that compare A versus C and B versus C are
analyzed jointly. The combination of the direct and indirect into a single effect size can increase
precision while randomization is respected. The combination of direct and indirect evidence for any
given treatment comparison can be extended when ranking more than three types of treatments
according to their efficacy: every study contributes evidence about a subset of these treatments. We
will perform MTM within a Bayesian framework (Ades et al., 2006). This enables us to estimate the
probability for each intervention of being the best for each positive outcome, given the results of the
MTM. The analysis will be performed using WinBUGS (MRC Biostatistics Unit, Cambridge, U.K.,
http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml). MTM should be used with caution, and
the underlying assumptions of the analysis should be investigated carefully. Key among these is that
the network is consistent, meaning that direct and indirect evidence on the same comparisons agree.
Joint analysis of treatments can be misleading if the network is substantially incoherent, i.e., if there is
disagreement between indirect and direct estimates. To evaluate statistically the assumption of
consistency we will employ two approaches. As a first step, we will calculate the difference between
indirect and direct estimates in each closed loop formed in the network and we will subsequently
examine whether there are any material discrepancies (Salanti 2009). To evaluate the assumption in
the network as whole (rather than testing each closed loop) we will employ the design-by-treatment
interaction (White 2011). In case of significant inconsistency, we will investigate possible sources of it.
Inconsistency may result as an uneven distribution of effect modifiers across groups of trials that
compare different treatments. Therefore, we will investigate the distribution of clinical and
methodological variables that we suspect may be potential sources of either heterogeneity or
inconsistency in each comparison-specific group of trials.
Sensitivity analyses and meta-regression analysis

The following effect modifiers will be examined in a meta-regression analysis or in sensitivity analyses in order to see if they are responsible for heterogeneity and/or inconsistency, if any, and if they modify the obtained effect estimates (Khan et al., 2004; Posternak & Zimmerman, 2007): baseline depression severity, number of randomised arms, use of placebo comparator, sponsorship, use of imputed values. A sensitivity analysis will address whether unbalanced doses affected the results and we will apply a similar approach as that used in our previous MTM on antidepressants (Cipriani et al. 2009) to exclude studies with unfair dose comparisons. Further analyses to address dose effects or other factor possibly related to treatment effect will be performed if necessary. Subgroup analyses are not planned.
REFERENCES


