# Effect of cytochrome CYP2C19 metabolizing activity on antidepressant response and side effects: meta-analysis of data from genome-wide association studies

Chiara Fabbri[1,2], Katherine E. Tansey[3], Roy H. Perlis[4], Joanna Hauser[5], Neven Henigsberg[6], Wolfgang Maier[7], Ole Mors[8], Anna Placentino[9], Marcella Rietschel[10], Daniel Souery[11], Gerome Breen[2], Charles Curtis[2], Sang-Hyuk Lee[2], Stephen Newhouse[2], Hamel Patel[2], Michael O'Donovan[12], Glyn Lewis[13], Gregory Jenkins[14], Richard M. Weinshilboum[15], Anne Farmer[2], Katherine J. Aitchison[16], Ian Craig[2], Peter McGuffin[2], Koen Schruers[17], Joanna M. Biernacka[14,18], Rudolf Uher[19], Cathryn M. Lewis[2]

1 Department of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, Italy

2 Institute of Psychiatry, Psychology and Neuroscience, King's College London, United Kingdom

3 College of Biomedical and Life Sciences, Cardiff University, Cardiff, United Kingdom

4 Department of Psychiatry, Center for Experimental Drugs and Diagnostics, Massachusetts General Hospital, Boston, USA

5 Laboratory of Psychiatric Genetics, Department of Psychiatry, Poznan University of Medical Sciences, Poznan, Poland

6 Croatian Institute for Brain Research, Medical School, University of Zagreb, Zagreb, Croatia

7 Department of Psychiatry, University of Bonn, Bonn, Germany

8 Centre for Psychiatric Research, Aarhus University Hospital, Risskov, Denmark

9 Biological Psychiatry Unit and Dual Diagnosis Ward, Istituto Di Ricovero e Cura a Carattere Scientifico, Centro San Giovanni di Dio, Fatebenefratelli, Brescia, Italy

10 Division of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Mannheim, Germany

11 Laboratoire de Psychologie Médicale, Université Libre de Bruxelles and Psy Pluriel—Centre Européen de Psychologie Médicale, Brussels, Belgium

12 MRC Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, United Kingdom

13 Division of Psychiatry, University College London (UCL), London, United Kingdom

14 Department of Health Sciences Research, Mayo Clinic, Rochester, MN, United States

15 Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN, USA

16 Department of Psychiatry, University of Alberta, Edmonton, AB, Canada

17 School of Mental Health and Neuroscience, Department of Psychiatry and Neuropsychology, Maastricht University, Maastricht, The Netherlands

18 Department of Psychiatry & Psychology, Mayo Clinic, Rochester, MN, United States

19 Department of Psychiatry, Dalhousie University, Halifax, Canada

**Running title**: CYP2C19 in citalopram/escitalopram treatment

**Corresponding author**:

Prof. Cathryn Lewis

Social, Genetic and Developmental Psychiatry Centre

Institute of Psychiatry, Psychology and Neuroscience - PO80

De Crespigny Park,

Denmark Hill, London,

United Kingdom, SE5 8AF

Email address: cathryn.lewis@kcl.ac.uk

**Abstract**

Cytochrome (CYP) P450 enzymes have a primary role in antidepressant metabolism and variants in these polymorphic genes are targets for pharmacogenetic investigation. This is the first meta-analysis to investigate how *CYP2C19* polymorphisms predict citalopram/escitalopram efficacy and side effects.

*CYP2C19* metabolic phenotypes comprise poor metabolizers (PM), intermediate and intermediate+ metabolizers (IM; IM+), extensive and extensive+ metabolizers (EM [wild type]; EM+) and ultra-rapid metabolizers (UM) defined by the two most common *CYP2C19* functional polymorphisms (rs4244285 and rs12248560) in Caucasians. These polymorphisms were genotyped or imputed from genome-wide data in four samples treated with citalopram or escitalopram (GENDEP, STAR*D, GenPod, PGRN-AMPS). Treatment efficacy was assessed by standardized percentage symptom improvement and by remission. Side effect data were available at weeks 2-4, 6 and 9 in three samples. A fixed-effects meta-analysis was performed using EM as the reference group.

Analysis of 2558 patients for efficacy and 2037 patients for side effects showed that PMs had higher symptom improvement (SMD=0.43, CI=0.19-0.66) and higher remission rates (OR=1.55, CI=1.23-1.96) compared to EMs. At weeks 2-4, PMs showed higher risk of gastro-intestinal (OR=1.26, CI=1.08-1.47), neurological (OR=1.28, CI=1.07-1.53) and sexual side effects (OR=1.52, CI=1.23-1.87; week 6 values were similar). No difference was seen at week 9 or in total side effect burden. PMs did not have higher risk of dropout at week 4 compared to EMs. Antidepressant dose was not different among *CYP2C19* groups.

*CYP2C19* polymorphisms may provide helpful information for guiding citalopram/escitalopram treatment, despite PMs being relatively rare among Caucasians (~2%).

**Keywords**: *CYP2C19*; gene; antidepressant; response; side effects; major depression

**1. Introduction**

Major depressive disorder (MDD) is a leading cause of disability-adjusted life years worldwide (GBD 2015 Disease and Injury Incidence and Prevalence Collaborators, 2016). Although anti-depressant drugs can be an effective therapy, remission rates are disappointing, largely as a consequence of high variability in efficacy among individuals combined with early discontinuation or poor compliance due to side effects (Hodgson et al., 2012; Crawford et al., 2014). Genetic variants are considered key modulators of antidepressant efficacy and side effects (Cacabelos et al., 2012). Common variants were estimated to explain approximately 42% of inter-individual variability in antidepressant response (Tansey et al., 2013), confirming the role of genetic polymorphisms as promising markers to provide personalized treatments.

Previous pharmacogenetic studies for antidepressant efficacy and side effects have focused on genes involved in antidepressant mechanisms of action (pharmacodynamics) or in antidepressant transport/metabolism (pharmacokinetics), including the cytochrome P450 genes (CYP450) (Fabbri and Serretti, 2015). These CYP450 genes are included in commercial pharmacogenetic tests (e.g. GeneSight Psychotropic, Genecept Assay™, YouScript Psychotropic (GTR: Genetic Testing Registry, 2017)). They form promising targets for personalizing antidepressant treatment, since they are responsible for antidepressant drug metabolism and their polymorphisms define phenotypic groups with different level of metabolic activity (Porcelli et al., 2011). An association between CYP450 metabolizer status (CYP450 metabolic phenotypes) and metabolite plasma levels has been consistently reported for antidepressants, but the association of CYP450 metabolic phenotypes with antidepressant efficacy and side effects is more controversial (Porcelli et al., 2011).

*CYP2C19* is the primary CYP450 isoform responsible for the metabolism of citalopram and escitalopram, two commonly prescribed SSRIs (selective serotonin reuptake inhibitors) (Hicks et al., 2015). Elevated drug concentrations have been observed in *CYP2C19* poor metabolizers (PMs), which may increase the risk of adverse drug reactions, while *CYP2C19* ultrarapid metabolizers (UMs) may have lower exposure to these drugs leading to treatment failure. *CYP2C19*-adjusted doses for citalopram and escitalopram have been estimated, but these were based on observed differences in drug pharmacokinetics, not differences in clinical outcomes of efficacy and side effects (Hicks et al., 2015).

Inconsistent associations between *CYP2C19* metabolic phenotypes and citalopram/escitalopram outcomes have been observed, and several factors may have led to the contradictory results (Peters et al., 2008; Mrazek et al., 2011; Hodgson et al., 2014; Hodgson et al., 2015):

1) Only a weak correlation exists between SSRI dose and efficacy. Drug plasma levels may not be associated with either efficacy or side effects, at least not linearly, and power to detect this

association is limited by several factors (e.g. difference between plasma and brain drug concentration) (Jakubovski et al., 2016; Hodgson et al., 2014; Hodgson et al., 2015; Florio et al., 2017);

2) Pharmacodynamic mechanisms may modulate the association between *CYP2C19* metabolic phenotypes and citalopram/escitalopram efficacy and some side effects, confounding the association between pharmacokinetic parameters and treatment outcomes (Jukić et al., 2016);

3) *CYP2C19* PM are rare, and studies may have lacked power to detect a pharmacogenetic association with this metabolic phenotype.

In this study, we present the first meta-analysis to investigate association between *CYP2C19* metabolic phenotypes and citalopram/escitalopram efficacy and side effects. This large study aimed to identify a link between *CYP2C19* metabolic phenotypes and treatment outcomes and to determine whether dose adjustments based on *CYP2C19* metabolic phenotypes should be part of personalized medicine for antidepressant treatment.

## 2. Experimental procedures

### 2.1. Samples

Four samples were included in this meta-analysis; all patients had a diagnosis of MDD and they were treated with citalopram or escitalopram for 8 weeks (PGRN-AMPS) or 12 weeks (GENDEP, STAR*D, GenPod). Clinical-demographic characteristics are reported in Supplementary Table 1, and show similarity across samples in mean age (~40 years old), percent of females (60-67%) and baseline severity (~22 ± 5 HAMD-17 equivalents).

### 2.1.1. GENDEP

The Genome-Based Therapeutic Drugs for Depression (GENDEP) project was a 12-week partially randomized open-label pharmacogenetic study with two active treatment arms. 867 patients with unipolar depression (ICD-10 or DSM-IV criteria) aged 19–72 years were recruited at nine European centres. Eligible participants were allocated to flexible-dosage treatment with either escitalopram (10–30 mg daily) or nortriptyline. Only 499 patients treated with escitalopram were included in the current meta-analysis. Severity of depression was assessed weekly by the Montgomery-Asberg Depression Rating Scale (MADRS) (Montgomery and Asberg, 1979), Hamilton Rating Scale for Depression (HRSD–17) (Hamilton, 1967) and other measures. Side effects were assessed at baseline and then weekly using the Antidepressant Side-Effect Checklist (ASEC) and UKU Side Effect Rating Scale, with good agreement between them. The ASEC data were analysed for this study, since they have lower rates of missing data (Uher et al., 2009). Detailed information about the GENDEP study has been previously reported (Uher et al., 2010).

### 2.1.2 STAR*D

The Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study was a NIMH-funded study to determine the effectiveness of different treatments for patients with MDD who have not responded to the first antidepressant treatment. A total of 4,041 non-psychotic MDD (DSM-IV criteria) patients aged between 18 and 75 years were enrolled from primary care or psychiatric outpatient clinics from 41 clinical sites around the US. Of these, 2,876 patients met the inclusion criterion of having at least moderate depression severity, among which 1,948 were genotyped using genome-wide arrays (Garriock et al., 2010). Severity of depression was assessed using the 16-item Quick Inventory of Depressive Symptomatology-Clinician Rated (QIDS-C16) (Trivedi et al., 2004) at baseline, weeks 2, 4, 6, 9, and 12. Side effects were measured at the same time points using the Patient-Rated Inventory of Side Effects (PRISE). This study uses data from level 1, where all patients received citalopram (20-40 mg/day). Detailed description of the study design and population are reported elsewhere (Rush et al., 2004).

### 2.1.3. PGRN-AMPS

The Pharmacogenomic Research Network Antidepressant Medication Pharmacogenomic Study (PGRN-AMPS) recruited 529 participants aged between 18 and 84 years with non-psychotic MDD recruited at inpatient and outpatient practices of the Department of Psychiatry and Psychology, Mayo Clinic, Rochester, Minnesota. Included participants were treated with either citalopram (n=167, dose 20-40 mg/day) or escitalopram (n=353, dose 10-20 mg/day) and depressive symptoms were rated using QIDS-C16 as in STAR*D. Side effects were assessed using the PRISE scale at weeks 4 and 8. Further details were reported elsewhere (Ji et al., 2013).

### 2.1.4. GenPod

The GENetic and clinical Predictors Of treatment response in Depression (GenPod) was a multi-centre randomized clinical trial of 601 patients aged between 18-74 years recruited in primary care who had an ICD-10 diagnosis of major depression of at least moderate severity as assessed by the Clinical Interview Schedule-Revised (CIS-R) (Lewis et al., 1992) and the Beck Depression Inventory (BDI) (Beck et al., 1961). No measure of side effects was available. Patients were recruited in three UK centres (Bristol, Birmingham and Newcastle). Individuals were randomly allocated to either reboxetine (4 mg twice daily) or citalopram (20 mg/day) treatment for 12 weeks and symptom severity was assessed using the BDI. 240 patients of European ancestry and treated with citalopram were included in this meta-analysis. Further details about this study can be found elsewhere (Thomas et al., 2008).

### 2.2. Outcomes

### 2.2.1. Treatment efficacy

Treatment efficacy was measured by percentage symptom improvement and by remission at study endpoint (week 12 for GENDEP, STAR*D and GenPod; week 8 for PGRN-AMPS). Continuous measures, such as percentage improvement, capture more information and have higher power than cutoff-based dichotomous measures, however remission has a particular clinical relevance since it is associated with MDD prognosis (Streiner, 2002; Gaynes et al., 2009).

The percentage symptom improvement was corrected for possible confounding variables (age, baseline severity, and center for multi-center studies) and then standardized in each sample to allow comparability across the different scales used to evaluate depressive symptoms.

Remission was defined as a binary variable according to standard definitions (HRSD–17 $\leq$ 7 in GENDEP; QIDS-C16 $\leq$ 5 in STAR*D and PGRN-AMPS; BDI $<$ 10 in GenPod). In GENDEP symptom improvement was calculated using the MADRS scale similarly to previous studies (Uher et al., 2010) while HRSD–17 was used to define remission given the stronger consensus about the threshold to identify remission on this scale in contrast to MADRS, where different definitions of remission have been reported (Li et al., 2016) (Jacobsen et al., 2015).

HRSD–17 and QIDS-C16 missing values at follow-up were imputed using the best unbiased estimate from a mixed-effect linear regression model, with fixed linear and quadratic effects of time and random effects of individual and center of recruitment, following previously reported methods (Uher et al., 2010).

### 2.2.2. Side effects

Measures of side effects were available in GENDEP, STAR*D and PGRN-AMPS. In GENDEP we chose to use the ASEC because data was more complete than the UKU (Uher et al., 2009). In STAR*D and PGRN-AMPS side effects were assessed using the PRISE scale. Both scales use a rating of severity for each side effect (coded 0-3 in ASEC, and 0-2 in PRISE) which was dichotomized (0=absent, 1=present) for the meta-analysis. Side effects were grouped in categories that were assessed in these samples: gastro-intestinal (dry mouth, diarrhea, constipation, nausea or vomiting), cardiovascular (palpitations, dizziness or feeling light-headed on standing), central nervous system (headache, tremor, feeling like the room is spinning), sleep (insomnia, drowsiness or oversleeping) and sexual (loss of desire, trouble achieving orgasm, trouble with erection). These categories were analysed as dichotomous variables (presence of at least one side effect in each category). Study retention at week 4 was compared among *CYP2C19* metabolic phenotypes since patients who did not benefit from treatment or had troubling side effects are expected to be lost from follow-up early in the study. To assess the overall severity of side effects, we summed the number of side effects reported, and dichotomized at the 3rd quartile of the distribution in each sample to study if *CYP2C19* groups were differently distributed among those patients having the

most relevant burden of side effects. We also calculated a standardized continuous measure of all side effects reflecting the whole distribution of side effect severity. Analyses on total side effect severity were performed only in GENDEP and STAR*D because raw data were not available in PGRN-AMPS.

Antidepressant-induced side effects are more frequent at the beginning of treatment and then decrease (Uher et al., 2009). We therefore meta-analysed side effects at weeks 2-4 (no assessment was performed at week 2 in PGRN-AMPS), week 6 and weeks 8-9 (no assessment was performed at week 8 in STAR*D while in GENDEP we used week 8 data because of lower missing rate compared to week 9).

In GENDEP side effects were common at baseline in medication-free patients (Uher et al., 2009). We therefore performed a sensitivity analysis excluding side effects there were present also at baseline in drug-free GENDEP patients.

## 2.3. Genotyping and definition of *CYP2C19* metabolic phenotypes

*CYP2C19* metabolic phenotypes comprise poor metabolizers (PM), intermediate and intermediate+ metabolizers (IM; IM+), extensive and extensive+ metabolizers (EM [wild type]; EM+) and ultra-rapid metabolizers (UM) defined by the two most common *CYP2C19* functional polymorphisms (rs4244285 and rs12248560) which capture the CYP2C19 *1, *2 and *17 functional alleles (Supplementary Table 2) (Hodgson et al., 2014). These polymorphisms were directly genotyped in GENDEP using the AmpliChip CYP450 test (Hodgson et al., 2014) and they were imputed in the other samples using the Haplotype Reference Consortium (HRC version r1.1 2016) panel as reference and Minimac3. Pre-imputation quality control was performed according to standard criteria (variants with missing rate $\geq$ 5%; monomorphic variants; subjects with genotyping rate < 97%; subjects with gender discrepancies; subjects with abnormal heterozygosity; related subjects (identity by descent (IBD) >0.1875 (Anderson et al., 2010)); population outliers according to Eigensoft analysis of linkage-disequilibrium-pruned genetic data (Price et al., 2006); and non-white subjects). Imputation quality was assessed using $R^2$ (Li et al., 2010) and comparing imputed and genotyped *CYP2C19* metabolic phenotypes in GENDEP.

## 2.4. Statistical analysis

Individual-level phenotypes and genotypes were available for all studies. A fixed-effects meta-analysis was performed with the R package "Netmeta" (https://cran.r-project.org/web/packages/netmeta/index.html). This package has been created for performing network meta-analysis and it was useful for this study since multiple groups needed to be compared to the reference group even if there were not indirect comparisons (i.e. all the studies provided data for each of the considered *CYP2C19* metabolic phenotypes). Phenotypic groups were compared

using the wild-type EM as the reference group. A random-effects meta-analysis was carried out for completeness and comparison of findings. Standardized mean difference (SMD) or odds ratio (OR) with 95% confidence intervals (CI) were calculated. Heterogeneity across studies was assessed using $I^2$ and Cochran's Q (Higgins et al., 2003).

This meta-analysis provided 80% power to identify an effect size (SMD) of $d$=0.40 when comparing PMs (the smallest group, n=51) with EMs (the reference group, n=1049) for a continuous outcome and OR=2.21 for a binary outcome, at a significance level of 0.05 (Faul et al., 2007).

We estimated that a corrected p value of 0.008 would account for the six independent tests that were carried out (improvement and response were correlated and considered as one test; the five side effect outcomes were considered as independent outcomes). Side effects at different weeks are not independent. Each outcome was tested five times comparing EMs with other functional groups, but *CYP2C19* metabolic groups are a consolidated classification with precise functional and biological meaning that represents a useful guide in interpreting findings. A more stringent multiple-testing correction would also take into account the comparisons between EMs and other *CYP2C19* functional groups (0.008/5=0.0016), but we suggest that the known functional interpretation of *CYP2C19* groups may be a valid alternative to strict statistical rigour in this context.


## 3. Results

A description of the clinical-demographic characteristics of the included samples is provided in Supplementary Table 1. There was no difference in mean citalopram or escitalopram dose by *CYP2C19* metabolic phenotypes at study endpoint in GENDEP, STAR*D and PGRN-AMPS (dose information was not available in GenPod). The distribution of phenotypic groups in the analysed samples is reported in Supplementary Table 3A. Imputation quality was high in all samples for both polymorphisms ($R^2$ between 0.95 and 0.99 (Li et al., 2010)). GENDEP participants had 97.6% consistency between genotyped and imputed SNPs (Supplementary Table 3B).


### 3.1. Treatment efficacy

In total, 2558 patients were included in the meta-analysis. The distribution of efficacy outcomes across *CYP2C19* metabolic phenotypes was reported in Supplementary Table 4. Compared to EMs, PMs had higher symptom improvement scores (SMD=0.43, CI=0.19-0.66, p=0.00037) and higher remission rates (OR=1.55, CI=1.23-1.96, p=0.00025), with low or absent heterogeneity ($I^2$ was 11.5% and 0%, respectively). Other *CYP2C19* metabolic phenotypes did not show different

outcomes compared to EMs (Supplementary Table 5 and Figure 1). Results did not change using a random-effects model.

## 3.2. Treatment side effects

Across STAR*D, GENDEP and PGRN-AMPS 2037 patients were included in the analysis. The distribution of side effects across *CYP2C19* metabolic phenotypes was reported in Supplementary Table 6. At weeks 2-4, PMs showed higher risk of gastro-intestinal side effects (OR=1.26, CI=1.08-1.47, p=0.0033), of CNS side effects (OR=1.28, CI=1.07-1.53, p=0.0068) and of sexual side effects (OR=1.52, CI=1.23-1.87, p=0.0001) (Supplementary Table 5 and Figure 2). Considering a corrected p threshold of 0.008, all these side effects were significantly more frequent in PMs. At week 6, PMs showed higher risk of sexual side effects (OR=1.64, CI=1.23-2.17, p=0.0007) but no higher risk of other side effects. For all these comparisons heterogeneity was low ($I^2$ range 0%-24%). No difference was seen at week 8-9 for any side effect, except a weak non-significant trend for sexual side effects. No difference in total side effects burden was observed at any time point when we considered the risk of having side effects in the highest quartile of the distribution in each sample. A non-significant effect of higher total side effects was observed in PMs at week 2 (p=0.03) for the continuous measure of side effect severity (Supplementary Table 5 and Supplementary Figure 1). *CYP2C19* IM+ group was the only metabolic phenotype to show higher risk of drop out at week 4 (OR=1.80, 95% CI=1.08-3.00, p=0.024), but this association did not survive multiple-testing correction. PMs did not show higher risk of dropout at week 4 (OR=1.16, CI=0.38-3.58). Other *CYP2C19* phenotypic groups did not show relevant differences compared to EMs, except lower risk of cardiovascular side effects and sleep side effects in EM+ at weeks 2-4 (OR=0.77, CI=0.64-0.92, p=0.0048) and 6 (OR=0.84, CI=0.75-0.95, p=0.0039), respectively, and higher risk of CNS side effects at week 8 in UMs (OR=1.26, 95% CI=1.04-1.53, p=0.019), but the latter did not survive multiple-testing correction.

The use of a random-effects model did not change the results.

Excluding those side effects there were already present at baseline in drug-free patients in GENDEP, results did not change, except that PMs showed higher risk of gastro-intestinal side effects also at week 6 (OR=1.47, CI=1.13-1.92, p=0.004). In addition, the trend of higher sexual side effects in PMs at weeks 8-9 was not observed, the lower risk of cardiovascular side effects in EM+ at weeks 2-4 became a non-significant trend (OR=0.82, CI=0.67-0.99) and there was a non-significant trend of higher gastro-intestinal side effects in PMs at weeks 8-9 (OR=1.35, CI=1.01-1.81).

# 4. Discussion

This study shows that *CYP2C19* PMs had higher symptom improvement and higher remission probability compared to EMs during treatment with citalopram or escitalopram (Figure 1). The observed SMD of 0.43 in symptom improvement between PMs and EMs is statistically considered close to a medium effect size (0.50) (Faraone, 2008). Statistical outcomes cannot be equated with clinical relevance and a clinical relevance cutoff of SMD=0.24 was proposed based on the effect size observed for antidepressant drugs (SMD=0.31, CI=0.27-0.35) and psychotherapy (SMD=0.25, CI= 0.14-0.36) in depression (Cuijpers et al., 2014). Other *CYP2C19* metabolic phenotypes, including UMs, showed no differences in efficacy outcomes compared to EMs. In addition to increased treatment efficacy, PMs showed higher risk of gastro-intestinal, CNS and sexual side effects early in treatment (particularly during the first 2-4 weeks), but not later in treatment (weeks 8-9) (Figure 2). The highest effect size was observed for the risk of sexual side effects at weeks 2-4 and 6 (OR=1.52 and 1.64, respectively) and this was the only side effect association that survived a stringent multiple-testing correction (considering the comparisons between EMs and other *CYP2C19* functional groups the multiple testing corrected p-value would be 0.0016). At week 4, PMs did not show a higher burden of total side effects and had no higher risk of dropout. The risk of dropout due to side effects was assumed to be the highest during the first 4 weeks of treatment, but data reporting the observed cause of dropout were not available. Mean antidepressant dose was not different among *CYP2C19* metabolizing groups. These results suggest that although some side effects were more common in PMs in the first weeks of treatment, overall they were not more troubling than in other *CYP2C19* groups and they may be balanced by higher improvement in depressive symptoms.

These findings are consistent with a previous STAR*D study that investigated remission and tolerance to citalopram (Mrazek et al., 2011), where tolerance represents a measure of side effect level. Tolerance was defined as continuation of citalopram treatment after the completion of Level 1 of the STAR*D trial. Previous studies in GENDEP and STAR*D failed to establish association between *CYP2C19* metabolizer status (PM vs. EM) and response, side effects or study retention (Peters et al., 2008; Hodgson et al., 2015; Hodgson et al., 2014), but individual studies would have limited power given the low number of subjects with PM metabolic phenotype (~2% of all patients analysed), particularly in GENDEP which has only six PM subjects. A previous analysis of *CYP2C19* in GENDEP used different definitions of side-effect, investigating each ASEC item and the sum of ASEC items (Hodgson et al., 2015).

No difference in treatment efficacy or side effects was identified between UMs and EMs, except for a non-significant higher risk of CNS side effects only at weeks 8-9 (Figure 2) that was probably the effect of random noise.

The only phenotypic group that showed lower risk of side effects was EM+ (lower risk of sleep side effects at week 6 and of cardiovascular side effects at weeks 2-4), suggesting that weak differences may depend on metabolic level but the UM group may have not provided enough power to observe them (~4-5% of patients were UMs in the analysed samples).

In addition to pharmacokinetic mechanisms, pharmacodynamic mechanisms may be involved in the association between *CYP2C19* and antidepressant response, since *CYP2C19* activity was reported to influence central neurotransmitters and neurotrophins relevant to antidepressant mechanisms of action (Jukić et al., 2016).

Our results conflict with the recommendation, based on pharmacokinetic parameters, of a 50% reduction in the starting dose of citalopram/escitalopram in *CYP2C19* PMs (Hicks et al., 2015), since we showed that a standard dose was associated with greater efficacy without higher drop-out rates. Antidepressant treatment with citalopram/escitalopram may be particularly indicated in *CYP2C19* PMs given the efficacy profile, if appropriate clinical support and monitoring is provided and the patient is informed of potential side effects at the beginning of the treatment. Effective plasma (and brain) drug concentrations may be reached in a higher proportion of PMs than other metabolic phenotypes, at the price of more frequent early side effects. The correlation between escitalopram serum concentration and treatment response was showed to be nearly-asymptotic, with poor antidepressant response at sub-therapeutic plasma concentration and stable response at therapeutic plasma concentration (Florio et al., 2017). When the threshold serum concentration is reached (which corresponds to serotonin transporter occupancy of 80%), further dose increase does not improve response. However, no definite drug plasma concentration threshold was identified in existing studies, thus drug plasma concentration has poor usefulness to guide dose prescription. The good tolerability profile of citalopram/escitalopram implies that side effects are usually not troubling, which may not be true for other antidepressants, such as tricyclic antidepressants (TCAs) or venlafaxine (Cipriani et al., 2012; Cipriani et al., 2009). It should be noted that TCAs and venlafaxine have specific profiles of efficacy and they represent valid alternatives to SSRIs as currently reported in clinical guidelines, but it should not be assumed that the current results referred to *CYP2C19* PMs can be applied to antidepressants different from citalopram and escitalopram.

The limitations and strengths of this study should be considered. This was the first meta-analysis to investigate the role of *CYP2C19* metabolic phenotypes in citalopram/escitalopram efficacy/side

effects, individual level data were available in all samples and the total sample size was the largest ever used for investigating this topic. On the other hand, PMs are rare in the Caucasian population resulting in limited power to identify differences involving this group even in this sample of 2558 patients. Side effect assessment was not available in all samples, and at weeks 6 and 8-9 part of patients dropped from the study and side effects data could not be imputed because it would be unreliable. At weeks 6 and 8-9, respectively, side effects were available in 84.4% and 73.6% of the initial sample in STAR*D, while in 85.9% and 83.3% of the initial sample in GENDEP. In PGRN-AMPS 87% of patients initially included had side effect data at week 4 and 80% at week 8 (sample size for this analysis is reported in Supplementary Table 6). Our findings suggest that *CYP2C19* PMs may benefit from standard doses of citalopram/escitalopram, with a higher response than other metabolic phenotypes. No conclusions could be drawn for UMs since which showed no significant differences in outcomes compared to EMs, and the study was probably under-powered to detect weak effects. EM+ was the only group showing lower risk of some side effects compared to EMs. We observed no to low heterogeneity among studies for both efficacy and side effects. For the former group all samples showed similar better outcome in PMs compared to EMs except GenPod, which included only three PM patients explaining the marginal effect on heterogeneity. Finally, the possible confounding effect of *CYP2C19* enhancers/inhibitors was not assessed, but a previous analysis in GENDEP concluded that the exclusion of subjects with concomitant use of enhancers/inhibitors did not change the pattern of results (Hodgson et al., 2014).

In conclusion, this meta-analysis shows good efficacy in *CYP2C19* poor metabolisers with citalopram/escitalopram, contrasting previous pharmacokinetic findings (Hicks et al., 2015). Our results show better treatment outcomes in PMs treated with standard doses with no relevant impact on late side effects (after the 6th week of treatment). Careful information for patients and monitoring of side effects during the early phase of treatment are recommended. Other *CYP2C19* metabolic phenotypes, including UMs, did not show differences in efficacy or side-effect outcomes compared to EMs. An interesting implication of this study is the possibility to derive *CYP2C19* metabolic groups from standard genome-wide data with a good level of quality.
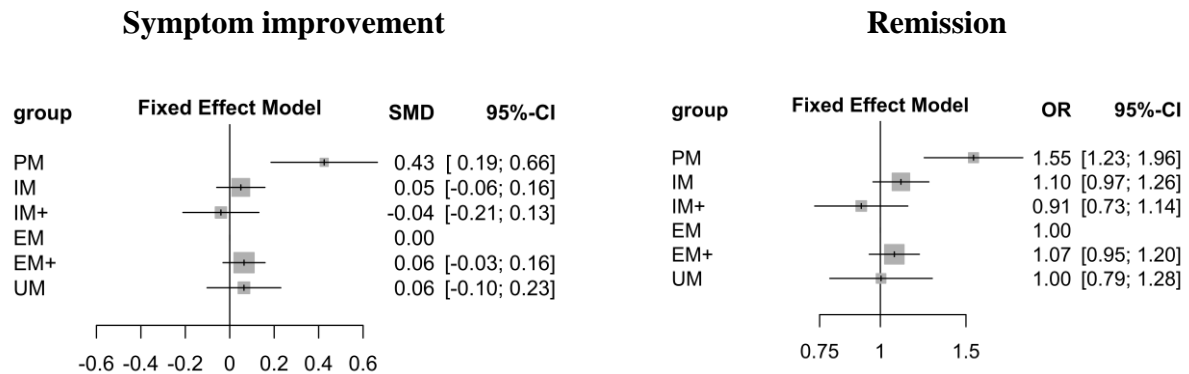
**References**

Anderson, C.A., Pettersson, F.H., Clarke, G.M., Cardon, L.R., Morris, A.P., and Zondervan, K.T., 2010. Data quality control in genetic case-control association studies. Nat. Protoc. 5, 1564–1573.

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., and Erbaugh, J., 1961. An inventory for measuring depression. Arch. Gen. Psychiatry 4, 561–571.

Cacabelos, R., Martinez-Bouza, R., Carril, J.C., Fernandez-Novoa, L., Lombardi, V., Carrera, I., Corzo, L., and McKay, A., 2012. Genomics and pharmacogenomics of brain disorders. Curr. Pharm. Biotechnol. 13, 674–725.

Cipriani, A., Santilli, C., Furukawa, T.A., Signoretti, A., Nakagawa, A., McGuire, H., Churchill, R., and Barbui, C., 2009. Escitalopram versus other antidepressive agents for depression. Cochrane Database Syst. Rev. CD006532.

Cipriani, A., Purgato, M., Furukawa, T.A., Trespidi, C., Imperadore, G., Signoretti, A., Churchill, R., Watanabe, N., and Barbui, C., 2012. Citalopram versus other anti-depressive agents for depression. Cochrane Database Syst. Rev. CD006534.

Crawford, A.A., Lewis, S., Nutt, D., Peters, T.J., Cowen, P., O'Donovan, M.C., Wiles, N., and Lewis, G., 2014. Adverse effects from antidepressant treatment: randomised controlled trial of 601 depressed individuals. Psychopharmacology (Berl). 231, 2921–2931.

Cuijpers, P., Turner, E.H., Koole, S.L., van Dijke, A., and Smit, F., 2014. What is the threshold for a clinically relevant effect? The case of major depressive disorders. Depress. Anxiety 31, 374–378.

Fabbri, C., and Serretti, A., 2015. Pharmacogenetics of Major Depressive Disorder: Top Genes and Pathways Toward Clinical Applications. Curr. Psychiatry Rep. 17,50

Faraone, S.V., 2008. Interpreting estimates of treatment effects: implications for managed care. P&T 33, 700–711.

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A., 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods 39, 175–191.

Florio, V., Porcelli, S., Saria, A., Serretti, A., and Conca, A., 2017. Escitalopram plasma levels and antidepressant response. Eur. Neuropsychopharmacol. 27, 940–944.

Garriock, H.A., Kraft, J.B., Shyn, S.I., Peters, E.J., Yokoyama, J.S., Jenkins, G.D., Reinalda, M.S., Slager, S.L., McGrath, P.J., and Hamilton, S.P., 2010. A Genomewide Association Study of Citalopram Response in Major Depressive Disorder. Biol. Psychiatry 67, 133–138.

Gaynes, B.N., Warden, D., Trivedi, M.H., Wisniewski, S.R., Fava, M., and Rush, A.J., 2009. What did STAR*D teach us? Results from a large-scale, practical, clinical trial for patients with depression. Psychiatr. Serv 60, 1439–1445.

GBD 2015 Disease and Injury Incidence and Prevalence Collaborators, 2016. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 388, 1545–1602.

Hamilton, M., 1967. Development of a rating scale for primary depressive illness. Br. J. Soc. Clin. Psychol. 6, 278–296.

Hicks, J.K., Bishop, J.R., Sangkuhl, K., Müller, D.J., Ji, Y., Leckband, S.G., Leeder, J.S., Graham, R.L., Chiulli, D.L., LLerena, A., et al., 2015. Clinical Pharmacogenetics Implementation Consortium ,CPIC Guideline for CYP2D6 and CYP2C19 Genotypes and Dosing of Selective Serotonin Reuptake Inhibitors. Clin. Pharmacol. Ther. 98, 127–134.

Higgins, J.P.T., Thompson, S.G., Deeks, J.J., and Altman, D.G., 2003. Measuring inconsistency in meta-analyses. BMJ 327, 557–560.

Hodgson, K., Mufti, S.J., Uher, R., and McGuffin, P., 2012. Genome-wide approaches to antidepressant treatment: working towards understanding and predicting response. Genome Med. 4, 52.

Hodgson, K., Tansey, K., Dernovsek, M.Z., Hauser, J., Henigsberg, N., Maier, W., Mors, O., Placentino, A., Rietschel, M., Souery, D., et al., 2014. Genetic differences in cytochrome P450 enzymes and antidepressant treatment response. J. Psychopharmacol 28, 133–141.

Hodgson, K., Tansey, K.E., Uher, R., Dernovšek, M.Z., Mors, O., Hauser, J., Souery, D., Maier, W., Henigsberg, N., Rietschel, M., et al., 2015. Exploring the role of drug-metabolising enzymes in antidepressant side effects. Psychopharmacology (Berl) 232, 2609–2617.

Jacobsen, P.L., Mahableshwarkar, A.R., Serenko, M., Chan, S., and Trivedi, M.H., 2015. A randomized, double-blind, placebo-controlled study of the efficacy and safety of vortioxetine 10 mg and 20 mg in adults with major depressive disorder. J. Clin. Psychiatry 76, 575–582.

Jakubovski, E., Varigonda, A.L., Freemantle, N., Taylor, M.J., and Bloch, M.H., 2016. Systematic Review and Meta-Analysis: Dose-Response Relationship of Selective Serotonin Reuptake Inhibitors in Major Depressive Disorder. Am. J. Psychiatry 173, 174–183.

Ji, Y., Biernacka, J.M., Hebbring, S., Chai, Y., Jenkins, G.D., Batzler, A., Snyder, K.A., Drews, M.S., Desta, Z., Flockhart, D., et al., 2013. Pharmacogenomics of selective serotonin reuptake inhibitor treatment for major depressive disorder: genome-wide associations and functional genomics. Pharmacogenomics J. 13, 456–463.

Jukić, M.M., Opel, N., Ström, J., Carrillo-Roa, T., Miksys, S., Novalen, M., Renblom, A., Sim, S.C., Peñas-Lledó, E.M., Courtet, P., et al., 2016. Elevated CYP2C19 expression is associated with depressive symptoms and hippocampal homeostasis impairment. Mol. Psychiatry 22, 1155-1163.

Lewis, G., Pelosi, A.J., Araya, R., and Dunn, G., 1992. Measuring psychiatric disorder in the community: a standardized assessment for use by lay interviewers. Psychol. Med. 22, 465–486.

Li, H., Gu, N., Zhang, H., Wang, G., Tan, Q., Yang, F., Ning, Y., Zhang, H., Lu, Z., Xu, X., et al., 2016. Efficacy and safety of quetiapine extended release monotherapy in bipolar depression: a multi-center, randomized, double-blind, placebo-controlled trial. Psychopharmacology (Berl) 233, 1289–1297.

Li, Y., Willer, C.J., Ding, J., Scheet, P., and Abecasis, G.R., 2010. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet. Epidemiol. 34, 816–834.

Montgomery, S.A., and Asberg, M., 1979. A new depression scale designed to be sensitive to change. Br. J. Psychiatry 134, 382–389.

Mrazek, D.A., Biernacka, J.M., O'Kane, D.J., Black, J.L., Cunningham, J.M., Drews, M.S., Snyder, K.A., Stevens, S.R., Rush, A.J., and Weinshilboum, R.M., 2011. CYP2C19 variation and citalopram response. Pharmacogenet. Genomics 21, 1–9.

Peters, E.J., Slager, S.L., Kraft, J.B., Jenkins, G.D., Reinalda, M.S., McGrath, P.J., and Hamilton, S.P., 2008. Pharmacokinetic genes do not influence response or tolerance to citalopram in the STAR*D sample. PloS One 3, e1872.

Porcelli, S., Fabbri, C., Spina, E., Serretti, A., and De Ronchi, D., 2011. Genetic polymorphisms of cytochrome P450 enzymes and antidepressant metabolism. Expert Opin. Drug Metab. Toxicol. 7, 1101–1115.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D., 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38, 904–909.

Rush, A.J., Fava, M., Wisniewski, S.R., Lavori, P.W., Trivedi, M.H., Sackeim, H.A., Thase, M.E., Nierenberg, A.A., Quitkin, F.M., Kashner, T.M., et al., 2004. Sequenced treatment alternatives to relieve depression ,STAR*D: rationale and design. Control. Clin. Trials 25, 119–142.

Streiner, D.L., 2002. Breaking up is hard to do: the heartbreak of dichotomizing continuous data. Can. J. Psychiatry Rev. Can. Psychiatr. 47, 262–266.

Tansey, K.E., Guipponi, M., Hu, X., Domenici, E., Lewis, G., Malafosse, A., Wendland, J.R., Lewis, C.M., McGuffin, P., and Uher, R., 2013. Contribution of Common Genetic Variants to Antidepressant Response. Biol. Psychiatry 73, 679–682.

Thomas, L., Mulligan, J., Mason, V., Tallon, D., Wiles, N., Cowen, P., Nutt, D., O'Donovan, M., Sharp, D., Peters, T., et al., 2008. GENetic and clinical predictors of treatment response in depression: the GenPod randomised trial protocol. Trials 9, 29.

Trivedi, M.H., Rush, A.J., Ibrahim, H.M., Carmody, T.J., Biggs, M.M., Suppes, T., Crismon, M.L., Shores-Wilson, K., Toprac, M.G., Dennehy, E.B., et al., 2004. The Inventory of Depressive Symptomatology, Clinician Rating (IDS-C) and Self-Report (IDS-SR), and the Quick Inventory of Depressive Symptomatology, Clinician Rating (QIDS-C) and Self-Report (QIDS-SR) in public sector patients with mood disorders: a psychometric evaluation. Psychol. Med. 34, 73–82.

Uher, R., Farmer, A., Henigsberg, N., Rietschel, M., Mors, O., Maier, W., Kozel, D., Hauser, J., Souery, D., Placentino, A., et al., 2009. Adverse reactions to antidepressants. Br. J. Psychiatry 195, 202–210.

Uher, R., Perroud, N., Ng, M.Y., Hauser, J., Henigsberg, N., Maier, W., Mors, O., Placentino, A., Rietschel, M., Souery, D., et al., 2010. Genome-wide pharmacogenetics of antidepressant response in the GENDEP project. Am. J. Psychiatry 167, 555-64.
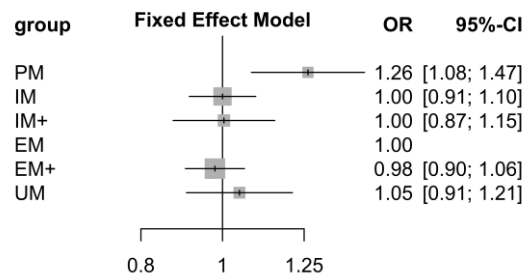
**Figure 1**: meta-analysis results for improvement and remission. PM=poor metabolizers: IM=intermediate metabolizers; IM+= intermediate metabolizers plus; EM=extensive metabolizers; EM+= extensive metabolizers+; UM=ultrarapid metabolizers. EM was taken as reference group. SMD=standardized mean difference. CI=confidence interval.
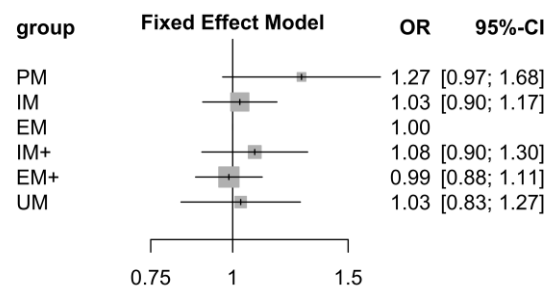


**Symptom improvement**

| group | Fixed Effect Model | SMD | 95%-CI |
|---|---|---|---|
| PM | | 0.43 | [ 0.19; 0.66] |
| IM | | 0.05 | [-0.06; 0.16] |
| IM+ | | -0.04 | [-0.21; 0.13] |
| EM | | 0.00 | |
| EM+ | | 0.06 | [-0.03; 0.16] |
| UM | | 0.06 | [-0.10; 0.23] |

-0.6 -0.4 -0.2 0 0.2 0.4 0.6

Quantifying heterogeneity/inconsistency:
tau^2 = 0.0033; I^2 = 11.5%
Test of heterogeneity/inconsistency:
Q=16.94, df=15, p=0.32

**Remission**

| group | Fixed Effect Model | OR | 95%-CI |
|---|---|---|---|
| PM | | 1.55 | [1.23; 1.96] |
| IM | | 1.10 | [0.97; 1.26] |
| IM+ | | 0.91 | [0.73; 1.14] |
| EM | | 1.00 | |
| EM+ | | 1.07 | [0.95; 1.20] |
| UM | | 1.00 | [0.79; 1.28] |

0.75 1 1.5

Quantifying heterogeneity/inconsistency:
tau^2 = 0; I^2 = 0%
Test of heterogeneity/inconsistency:
Q=12.47, df=15, p=0.64

**Figure 2**: meta-analysis results for side effects. PM=poor metabolizers: IM=intermediate metabolizers; IM+= intermediate metabolizers plus; EM=extensive metabolizers; EM+= extensive metabolizers+; UM=ultrarapid metabolizers. EM was taken as reference group. SMD=standardized mean difference. CI=confidence interval. For each comparison heterogeneity is quantified using tau^2, I^2 and assessed using Q test.

### Gastro-intestinal side effects at weeks 2-4

Fixed Effect Model

| group | OR | 95%-CI |
|---|---|---|
| PM | 1.26 | [1.08; 1.47] |
| IM | 1.00 | [0.91; 1.10] |
| IM+ | 1.00 | [0.87; 1.15] |
| EM | 1.00 | |
| EM+ | 0.98 | [0.90; 1.06] |
| UM | 1.05 | [0.91; 1.21] |

tau^2 = 0; I^2 = 0%
Q=5.54, df=10, p=0.85

### Gastro-intestinal side effects at week 6

Fixed Effect Model

| group | OR | 95%-CI |
|---|---|---|
| PM | 1.27 | [0.97; 1.68] |
| IM | 1.03 | [0.90; 1.17] |
| EM | 1.00 | |
| IM+ | 1.08 | [0.90; 1.30] |
| EM+ | 0.99 | [0.88; 1.11] |
| UM | 1.03 | [0.83; 1.27] |

tau^2 = 0; I^2 = 0%
Q=1.96, df=5, p=0.85

### Gastro-intestinal side effects at weeks 8-9

Fixed Effect Model

| group | OR | 95%-CI |
|---|---|---|
| PM | 1.27 | [0.95; 1.71] |
| IM | 1.02 | [0.89; 1.18] |
| IM+ | 1.14 | [0.94; 1.38] |
| EM | 1.00 | |
| EM+ | 1.03 | [0.92; 1.16] |
| UM | 1.04 | [0.83; 1.31] |

tau^2 = 0; I^2 = 0%
Q=7.6, df =10, p=0.67

### Cardiovascular side effects at weeks 2-4

Fixed Effect Model

| group | OR | 95%-CI |
|---|---|---|
| PM | 1.40 | [0.96; 2.05] |
| IM | 0.94 | [0.78; 1.14] |
| IM+ | 1.00 | [0.75; 1.32] |
| EM | 1.00 | |
| EM+ | 0.77 | [0.64; 0.92] |
| UM | 1.14 | [0.85; 1.53] |

tau^2 = 0; I^2 = 0%
Q=7.48, df=10, p=0.68

### Cardiovascular side effects at week 6

Fixed Effect Model

| group | OR | 95%-CI |
|---|---|---|
| PM | 1.12 | [0.55; 2.30] |
| IM | 0.92 | [0.70; 1.20] |
| IM+ | 1.11 | [0.77; 1.60] |
| EM | 1.00 | |
| EM+ | 0.83 | [0.65; 1.06] |
| UM | 1.22 | [0.83; 1.78] |

tau^2 = 0; I^2 = 0%
Q=2.05df=5, p=0.84

### Cardiovascular side effects at weeks 8-9

Fixed Effect Model

| group | OR | 95%-CI |
|---|---|---|
| PM | 1.64 | [0.93; 2.90] |
| IM | 0.92 | [0.68; 1.23] |
| IM+ | 1.32 | [0.93; 1.88] |
| EM | 1.00 | |
| EM+ | 0.90 | [0.70; 1.16] |
| UM | 1.38 | [0.94; 2.04] |

tau^2 = 0; I^2 = 0%
Q= 3.29, df=10, p=0.97

**CNS side effects at weeks 2-4**

| group | Fixed Effect Model | OR | 95%-CI |
|---|---|---|---|
| PM | | 1.28 | [1.07; 1.53] |
| IM | | 1.01 | [0.91; 1.12] |
| IM+ | | 0.91 | [0.77; 1.08] |
| EM | | 1.00 | |
| EM+ | | 1.05 | [0.96; 1.14] |
| UM | | 0.94 | [0.78; 1.13] |

0.75  1  1.5

tau^2 = 0.0011; I^2 = 5.2%
Q=10.55, df=10, p=0.39
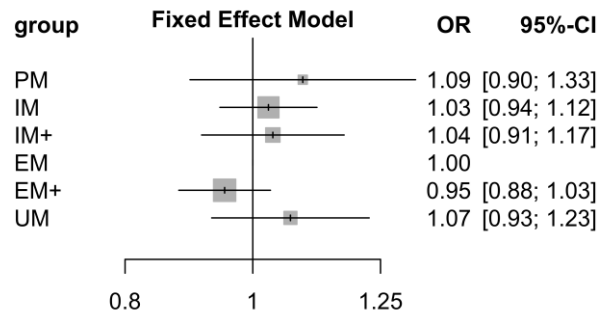
**CNS side effects at week 6**

| group | Fixed Effect Model | OR | 95%-CI |
|---|---|---|---|
| PM | | 1.16 | [0.80; 1.67] |
| IM | | 1.01 | [0.87; 1.17] |
| IM+ | | 1.12 | [0.92; 1.38] |
| EM | | 1.00 | |
| EM+ | | 0.89 | [0.78; 1.02] |
| UM | | 1.01 | [0.79; 1.29] |

0.75  1  1.5

tau^2 = 0; I^2 = 0%
Q=0.68, df=5, p=0.98

**CNS side effects at weeks 8-9**

| group | Fixed Effect Model | OR | 95%-CI |
|---|---|---|---|
| PM | | 1.19 | [0.86; 1.65] |
| IM | | 0.97 | [0.84; 1.13] |
| IM+ | | 1.16 | [0.96; 1.41] |
| EM | | 1.00 | |
| EM+ | | 1.01 | [0.89; 1.15] |
| UM | | 1.26 | [1.04; 1.53] |

0.75  1  1.5
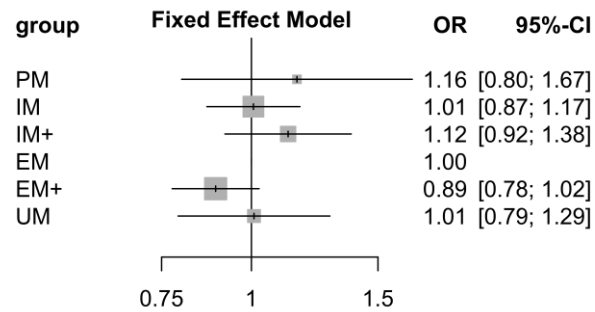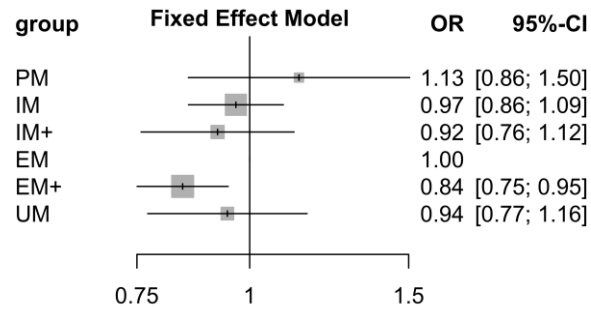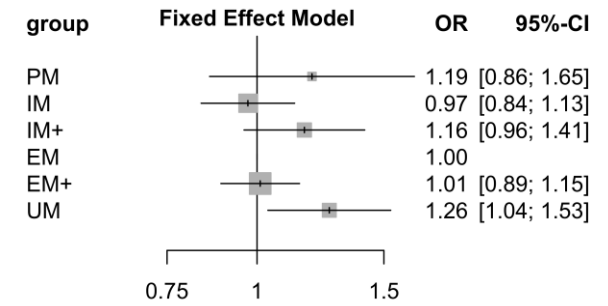
tau^2 = 0.0018; I^2 = 4.8%
Q=10.51, df=10, p=0.40

**Sleep side effects at weeks 2-4**

| group | Fixed Effect Model | OR | 95%-CI |
|---|---|---|---|
| PM | | 1.09 | [0.90; 1.33] |
| IM | | 1.03 | [0.94; 1.12] |
| IM+ | | 1.04 | [0.91; 1.17] |
| EM | | 1.00 | |
| EM+ | | 0.95 | [0.88; 1.03] |
| UM | | 1.07 | [0.93; 1.23] |

0.8  1  1.25

tau^2 = 0; I^2 = 0%
Q=4.24, df=10, p=0.94

**Sleep side effects at week 6**

| group | Fixed Effect Model | OR | 95%-CI |
|---|---|---|---|
| PM | | 1.13 | [0.86; 1.50] |
| IM | | 0.97 | [0.86; 1.09] |
| IM+ | | 0.92 | [0.76; 1.12] |
| EM | | 1.00 | |
| EM+ | | 0.84 | [0.75; 0.95] |
| UM | | 0.94 | [0.77; 1.16] |

0.75  1  1.5

tau^2 = 0; I^2 = 0%
Q=3.24, df=5, p=0.66

**Sleep side effects at weeks 8-9**

| group | Fixed Effect Model | OR | 95%-CI |
|---|---|---|---|
| PM | | 0.93 | [0.63; 1.37] |
| IM | | 0.99 | [0.86; 1.13] |
| IM+ | | 1.10 | [0.92; 1.32] |
| EM | | 1.00 | |
| EM+ | | 0.92 | [0.81; 1.04] |
| UM | | 1.06 | [0.86; 1.31] |

0.75  1  1.5

tau^2 = 0; I^2 = 0%
Q=4.57, df=10, p=0.92

**Sexual side effects at weeks 2-4**

| group | Fixed Effect Model | OR | 95%-CI |
|---|---|---|---|
| PM | | 1.52 | [1.23; 1.87] |
| IM | | 1.07 | [0.94; 1.22] |
| IM+ | | 1.14 | [0.95; 1.37] |
| EM | | 1.00 | |
| EM+ | | 0.92 | [0.81; 1.04] |
| UM | | 1.18 | [0.96; 1.45] |

0.75  1  1.5
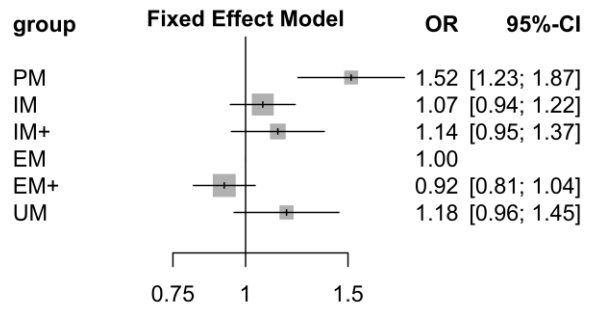
tau^2 = 0; I^2 = 0%
Q=6.89, df=10, p=0.74

**Sexual side effects at week 6**

| group | Fixed Effect Model | OR | 95%-CI |
|---|---|---|---|
| PM | | 1.64 | [1.23; 2.17] |
| IM | | 1.15 | [0.98; 1.34] |
| IM+ | | 1.01 | [0.77; 1.31] |
| EM | | 1.00 | |
| EM+ | | 0.88 | [0.75; 1.03] |
| UM | | 1.12 | [0.87; 1.45] |

0.5  1  2
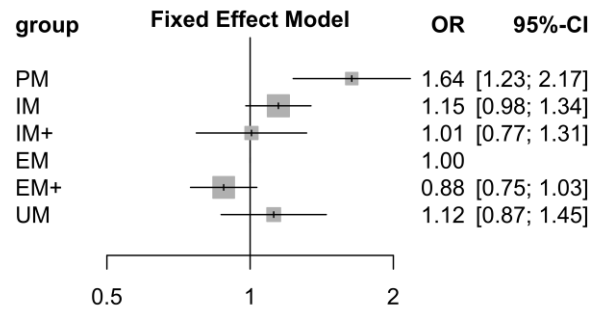
tau^2 = 0.0152; I^2 = 24%
Q=6.58, df=5, p=0.25

**Sexual side effects at weeks 8-9**

| group | Fixed Effect Model | OR | 95%-CI |
|---|---|---|---|
| PM | | 1.42 | [1.01; 1.98] |
| IM | | 1.04 | [0.88; 1.22] |
| IM+ | | 1.10 | [0.87; 1.40] |
| EM | | 1.00 | |
| EM+ | | 0.94 | [0.81; 1.09] |
| UM | | 1.06 | [0.81; 1.39] |

0.75  1  1.5

tau^2 = 0; I^2 = 0%
Q=9.95, df=10, p=0.45

## Contributors

Chiara Fabbri designed this study, performed the analyses and wrote the first draft of the paper. Katherine E. Tansey contributed to the analyses. Gregory Jenkins, Richard M. Weinshilboum and Joanna M. Biernacka contributed to the preparation of data and revision of the manuscript. Charles Curtis, Sang-Hyuk Lee, Stephen Newhouse, Anne Farmer, Hamel Patel contributed to the preparation of data. Katherine J. Aitchison provided suggestions about the method for phenotype classification. Rudolf Uher provided statistical suggestions. Cathryn M. Lewis supervised the whole process leading to the final paper (data preparation, analyses and paper writing). The other authors contributed to data collection, data preparation and/or revision of the manuscript.

## Conflict of interest

N Henigsberg participated in clinical trials sponsored by pharmaceutical companies including GlaxoSmithKline and Lundbeck. D Souery is serving in national advisory boards or consulting for Janssen, TEVA, Glaxo Smith Kline. His center is receiving unrestricted financial support from Lundbeck and Fondation René de Spoellberghe. W Maier, KJ Aitchison, AE Farmer and P McGuffin have received consultancy fees and honoraria for participating in expert panels from pharmaceutical companies including Lundbeck and GlaxoSmithKline and Roche Diagnostics. Dr Perlis reported serving on scientific advisory boards or consulting for Genomind LLC, Healthrageous, Pfizer, Perfect Health, Proteus Biomedical, PsyBrain Inc, and RID Ventures LLC and reported receiving royalties through Massachusetts General Hospital from Concordant Rater Systems (now Bracket/Medco). N Perroud received honoraria for participating in expert panels from pharmaceutical companies including Lundbeck. G Bondolfi is a member of a national advisory board for Bristol-Myer Squibb and Pfizer and has received research funding from GlaxoSmithKline, Wyeth-Lederle, Bristol-Myers-Squibb and Sanofi Aventis. M O'Donovan's department received £2000 in lieu of an honorarium to M O'Donovan from Lilly as a result of his participation in sponsored symposia in 2012. Those symposia were unrelated to the contents of this manuscript. The other authors declare no conflict of interest.

**Supplementary Material**

**Supplementary Material**

**Supplementary Material**

**Supplementary Material**

**Supplementary Material**

**Supplementary Material**

**Supplementary Material**