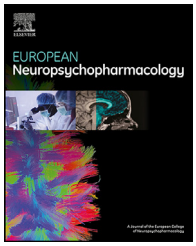
European Neuropsychopharmacology 59 (2022) 68–81



www.elsevier.com/locate/euroneuro

Evaluating study designs and treatment outcomes of antidepressant pharmacogenetic clinical trials - Challenges and future perspectives. A critical review

Alessandra Minelli[a,b,*], **Stefano Barlati**[c,d], **European College of Neuropsychopharmacology (ECNP) Pharmacogenomics & Transcriptomics Network**[1], **Bernhard T Baune**[e,f,g]

[e] *Department of Psychiatry and Psychotherapy, University of Münster, Münster, Germany*
[f] *Department of Psychiatry, Melbourne Medical School, University of Melbourne, Parkville, VIC, Australia*
[g] *The Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Parkville, VIC, Australia*
[a] *Department of Molecular and Translational Medicine, Biology and Genetic Division, University of Brescia, Viale Europa 11, Brescia 25123, Italy*
[b] *Genetics Unit, IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy*
[c] *Department of Clinical and Experimental Sciences, University of Brescia, Brescia, Italy*
[d] *Department of Mental Health and Addiction Services, ASST Spedali Civili of Brescia, Brescia, Italy*

Received 10 June 2021; received in revised form 6 April 2022; accepted 8 April 2022

KEYWORDS
Pharmacogenetic test;
Antidepressant;
Personalized medicine;

Abstract
Several data indicate that the success of pharmacological treatment in major depressive disorder (MDD) is still unsatisfactory. The reasons for the low response and remission rates are multiple and depend on environmental and biological factors intrinsic to the disease and drug treatments. Pharmacogenetic (PG) tests have the potential to increase efficacy predicting outcome and to reduce antidepressant discontinuation due to side effects. Several studies investi-

* Corresponding author at: Department of Molecular and Translational Medicine, Biology and Genetic Division, University of Brescia, Viale Europa 11, Brescia 25123, Italy.
E-mail address: alessandra.minelli@unibs.it (A. Minelli).

[1] European College of Neuropsychopharmacology (ECNP) Pharmacogenomics & Transcriptomics Network list of collaborators: Bernhard T Baune, Alessandra Minelli, Catherine Harmer, Alessandro Serretti, Mara Dierssen, Eduard Maron, Marie-Claude Potier, Massimo Gennarelli, Roos van Westrhenen, Alessio Squassina, David Stacey, Divja Mehta, Joost G E Janzing, Bjarke Ebert, Chiara Fabbri, Lara Nonell, Pietro Lio', Filip Rybakowski, Laura Xicota, Claudia Pisanu.

https://doi.org/10.1016/j.euroneuro.2022.04.007

Major depressive disorder;
Randomized controlled trial;
Remission

gated the utility of PG tests for antidepressants in MDD with interesting but contrasting results. To date most of them are observational studies with no comparator group, and few are randomized controlled trials (RCTs). The aim of this review is to provide an evaluation of the state of art on clinical methodologic features of RCTs with PG tests for antidepressant drugs in MDD, offering suggestions and favoring new insights that could be useful in the implementation of future trials. Several limitations concerning study design, generalization of results, duration of trials, patients group studied, and cost-effectiveness ratio were found, and a number of barriers have been noted in the adoption of PG tests into clinical practice. Despite some preliminary positive results, there is the need for larger and longer-term RCT studies, with the goal to capture the real impact of PG tests, also with stratified analysis concerning MDD features in terms of severity and antidepressant treatment failures in different ethnicity cohorts.

1. Introduction

Major depressive disorder (MDD) is the most common psychiatric disease worldwide and represents a leading cause of years lived with disability, leading to an enormous socioeconomic impact (Hasin et al., 2018). The most common therapeutic strategy for moderate to severe MDD is pharmacological treatment. In spite of the advances in antidepressant options, many patients fail to benefit from pharmacotherapy with low response and remission rates (Rush et al., 2006; Thase et al., 2010) as well as low adherence due to side effects (Cipriani et al., 2018; Sharma et al., 2019). This leads to a long unremitted disease, worse long-term prognosis, and significant medical, social and economic burden (Mrazek et al., 2014). The reasons are multiple and depend on environmental and biological factors intrinsic to the disease and drug treatments (Fabbri and Serretti, 2020; Gratten et al., 2014).

In this context, pharmacogenetic (PG) tests may be a valuable decision support tool for the management of pharmacological treatment in MDD, because they could have the potential to increase efficacy predicting treatment outcome, along with the reduction of antidepressant discontinuation decreasing side effects (Tanner et al., 2018).

Along with the growing availability of commercial PG tests for antidepressant drugs, there has been an equally growing concern about their utility. Several studies have been performed to investigate the impact of PG testing on antidepressant outcome in MDD (Bousman et al., 2019; Fabbri et al., 2018; Rosenblat et al., 2018) contributing to a general evaluation of their effectiveness and applicability, with interesting but contrasting results. To date most of them are observational studies with no comparator group, and few are randomized controlled trials (RCTs).

2. Aim of the review

Previous reviews on PG tests for antidepressants have mainly taken into consideration the limitations concerning the PG test mechanisms in terms of choices of genetic variants in accordance with drug labels and international guidelines (Bousman et al., 2019; Fabbri et al., 2018; Fabbri and Serretti, 2020; Zanardi et al., 2021a). None has been focused on clinical and assessment methodological characterization as possible sources of errors.

On this basis, the goal of this narrative and critical review is to provide an evaluation on the state of art concerning clinical methodologic features of RCTs with PG tests for antidepressant drugs in MDD. This structured analysis aims to provide insights and suggestions that could be useful for the clinical implementation of future trials.

3. Methods

In order to achieve the aims of the present review we focused on RCTs on PG tests for antidepressant drugs performed in MDD patients.

Electronic searches were performed using MEDLINE/PubMed and Scopus databases combining the following keywords/search terms: “Pharmacogenetics”, “Pharmacogenomics”, “test”, “genes”, “antidepressant(s)”, “response”, “remission”, “side effects”, “randomized controlled trial”, “depression”, “major depressive disorder”, “MDD”. Two of the authors (SB, AM) independently reviewed the database to avoid mistakes in the selection of articles. The reference list of the studies, meta-analyses and reviews on this issue were also reviewed in order to detect further publications. All RCT studies, meta-analyses, and review articles on PG tests in MDD, published until March 2022 were included. Studies were selected if they met the following criteria: (a) being an RCT on a PG test for antidepressant drugs performed in MDD, (b) being in English language, and (c) being an original paper published in a peer-reviewed journal.

4. Findings

Seven studies were identified (Bradley et al., 2018; Greden et al., 2019; Han et al., 2018; Pérez et al., 2017; Perlis et al., 2020; Shan et al., 2019; Tiwari et al., 2022). Three studies (Singh, 2015; Thase et al., 2019; Winner et al., 2013) were not included for the following reasons: the Winner et al. (2013) study represents a small pilot study with the same study design of the Greden et al. (2019), the Thase et al. (2019) reported further data of Greden “GUIDE trial” RCT (Greden et al., 2019), the Singh (2015) utilized PG test only for antidepressant dosing suggestion and not as an antidepressant choice decision tool.

In the next paragraphs we will present a detailed description of the characteristics of the included studies concerning experimental design, inclusion and exclusion criteria used, sample size, demographic features of the patients

recruited, assessment, outcomes and main findings. These features are synthetized and displayed in Table 1.

(a) Experimental design

The RCTs included in this review are prospective multicentre studies lasting 8 (Greden et al., 2019; Han et al., 2018; Perlis et al., 2020; Shan et al., 2019), 12 (Bradley et al., 2018; Pérez et al., 2017) or 36 weeks (Tiwari et al., 2022). The Greden et al. (2019) was followed by 4-week of unblended follow-up and of a further 12-week open-label extension period during which clinicians had access to the PG test report to support treatment decisions for all patients, including those assigned to the TAU group. The Tiwari et al. (2022) study was followed by a 16-week open-label extension period. In five studies the trial was registered at www.clinicaltrials.gov under the following identifier numbers (Bradley NCT02878928; Greden NCT02109939; Perez NCT02529462; Perlis NCT02634177; Tiwari NCT02466477) (Bradley et al., 2018; Greden et al., 2019; Pérez et al., 2017; Perlis et al., 2020; Tiwari et al., 2022) whereas for the other two studies no registration on official trial registers has been done (Han et al., 2018; Shan et al., 2019).

Five studies are partially double-blinded since the prescriber was not blind, while the rater and the patient were blind to the study group the patient was assigned (Bradley et al., 2018; Greden et al., 2019; Perlis et al., 2020; Shan et al., 2019; Tiwari et al., 2022). One is a single-blind study because the rater has never been blinded (Han et al., 2018), whereas in another study both the rater and the prescriber were not blinded, except for the assessment of the Patient Global Impression of Improvement (PGI-I) performed by phone (Pérez et al., 2017).

In all studies, patients in the TAU (treatment as usual) group were treated following the standard of care, thus, they received antidepressant treatment according to the psychiatrist's clinical discretion without the aid of PG testing.

(b) Inclusion and exclusion criteria

All the studies had as main inclusion criteria a primary diagnosis of MDD according to DSM-IV-TR or DSM-5, with the exception of Greden et al. (2019) study for which the diagnosis of depression was made using the 16-item Quick Inventory of Depressive Symptomatology, both Clinician-Rated and Self-report (QIDS-C16 and QIDS-SR16), where it was required to have a score higher or equal to eleven points in both symptom scales. Moreover, in Bradley et al. (2018) study, also patients with anxiety disorders according to DSM-5 as main diagnosis were included. Consequently, the patients were categorized into three different diagnosis categories: MDD, anxiety disorders, and both MDD and anxiety disorders in comorbidity. Finally, five studies (Greden et al., 2019; Han et al., 2018; Perlis et al., 2020; Shan et al., 2019; Tiwari et al., 2022) specified the inclusion of MDD patients only with absence of psychotic symptomatology at least in the current depressive episode. Moreover, three studies (Bradley et al., 2018; Greden et al., 2019; Perlis et al., 2020) excluded subjects with significant risk for suicide. All the studies, with the exception of Pérez et al. (2017), reported to excluded patients with concurrent main psychiatric disorders diagnosis such as bipolar disorder, schizophrenia, personality disorder, obsessive-compulsive disorder, eating disorder.

All the RCTs included patients who failed at least one prior adequate trial with antidepressants for the current depressive episode due to inefficacy or intolerable adverse effects, whereas three studies (Bradley et al., 2018; Pérez et al., 2017; Shan et al., 2019) included also patients who required medication de novo and who had never received psychiatric treatment in their lives.

(c) Sample size

The sample size of the RCT studies is extremely wide-ranging, going from a very small dimension of the initial total group, corresponding to 71 and 100 enrolled patients in the Han et al. (2018), Shan et al. (2019) studies, respectively, to a really large sample size of 1398 subjects recruited in the Greden et al. (2019). There is the same wide variability for the drop-out rate, ranging from high percentages in three studies at 8-week primary outcome endpoint with the loss of about 30% of the initial sample (Han et al., 2018; Shan et al., 2019; Tiwari et al., 2022), to a very low reduction in the Perlis et al. (2020) at the same primary outcome endpoint, with a drop-out rate of about 7%. The two RCTs lasting over 12 weeks had similar drop-out rates, 11.4% and 15.5% in (Bradley et al. (2018), Pérez et al. (2017) study, respectively .

Finally, four studies (Han et al., 2018; Pérez et al., 2017; Perlis et al., 2020; Tiwari et al., 2022) reported a sample size estimation and power according to the study design and analysis, but only in Perlis and colleagues' RCT the number of completers was congruent with the initial sample size planned.

(d) Demographics

The RCT studies included mainly women with a total group mean percentage of 68.7% (range 63.1%—74.9%), with similar percentages both in PG-guided and in TAU groups. The mean age of the total group of patients considering all studies is 43.8 (range 27.7–51.2) and no differences are reported between the two groups. With the exception of Shan et al. (2019) study, none reported the years of education. Concerning the ethnicity, two studies involved only Asian cohorts (Han et al., 2018; Shan et al., 2019), the others instead had a large heterogeneity although they mainly included Caucasian populations (mean percentage of Caucasians: 81.6%).

(e) Assessment

All the studies performed the assessment of depressive symptomatology with one of the most common clinical rating scales that is the Hamilton one (Hamilton Rating Scale for Depression 17 items - HAM-D17 or Structured Interview

Table 1 Characteristics of the randomized controlled trial studies (RCTs) included in the review.

Refs.	Experimental Design	Inclusion/Exclusion Criteria	Sample size	Demographics	Assessment	Outcomes	Main findings
Pérez et al. (2017)	- 12-week, multicenter, prospective, single blinded (patient blinded, double blinded only for the PGI-I scale, RCT - Hospitals and associated mental health centers, Spain - Two arms: PG-guided vs. TAU - 5 Timepoints: Baseline, 4, 6, 8 and 12 weeks	- Age: $\geq$ 18 years - Diagnosis of MDD (DSM-IV-TR) - CGI-S $\geq$ 4 - Dysthymic disorder, other non-specified depressive disorder as main diagnosis and secondary comorbidity of psychiatric and medical illness could be included - Subjects who required medication de novo or were receiving treatment and required substitution or addition of drug treatment with an AD - Exclusion: Other primary psychiatric diagnoses as main diagnosis, pregnancy and breastfeeding, requiring treatment with quinidine, cinacalcet and/or terbinafine (CYP2D6 inhibitors)	316 (PG-guided $n = 155$, TAU $n = 161$) - Completers at 12 weeks: 280 patients (PG-guided $n = 136$, TAU $n = 144$) - Drop-out rate at 12 weeks: 11.4% (PG-guided: 12.3%, TAU: 10.6%)	- Sex (% females): 63.6 (PG-guided: 63.9, TAU: 63.4) - Age (years), mean (SD): 51.2 (12.6), PG-guided: 51.7 (12.0), TAU: 50.7 (13.1) - Ethnicity (%): Caucasian (91.3), Latin American (6.2), other (2.5)	- Clinician-rated: CGI-S, HAM-D17 assessed at baseline, 6 and 12 weeks. Assessors not blinded. - Self-report: SDI and SATMED-Q assessed at baseline 6 and 12 weeks. PGI-I assessed by phone call in a double-blinded manner at 4, 8 and 12 weeks. - Adverse effects: FIBSER assessed at 6 and 12 weeks. Assessors not blinded.	- Primary outcome: Proportion of patients achieving a sustained response (PGI-*I* $\leq$ 2) within the 12 weeks. A sustained response was defined when a patient was a responder on at least two consecutive evaluations, maintaining that status until the final visit of the study. - Secondary outcomes: Response at the end of the 12 weeks (based on a PGI-I score of 2 or less), clinical progression as measured by HDRS-17, severity as measured by CGI-S, tolerability of treatment as measured by FIBSER, patient satisfaction with treatment as measured by SATMED-Q, patient disability as measured by SDI	- No difference in sustained response within the study period as measured by PGI-I (primary outcome) - Higher responder rate at 12 weeks in PG-guided as measured by PGI-I - Better tolerability at 6 and 12 weeks in PG-guided as measured by FIBSER

(*continued on next page*)

Table 1 (*continued*)

Refs.	Experimental Design	Inclusion/Exclusion Criteria	Sample size	Demographics	Assessment	Outcomes	Main findings
Bradley et al. (2018)	- 12-week, multicenter, prospective, patient and rater blinded RCT - Clinical sites, psychiatric and other specialization sites, USA - Two arms: PG-guided vs. TAU - 4 Timepoints: Baseline, 4, 8 and 12 weeks	- Age: range 19-87 years - Diagnosis of MDD and/or anxiety (DSM-5) - Subjects who required medication de novo or were receiving treatment and required substitution due to lack of efficacy or treatment discontinuation due to adverse events or intolerability - Exclusion: Concurrent diagnosis of BD, SZ, personality disorder, traumatic brain injury, significant risk for suicide and hospitalization, history of chronic kidney dysfunction, abnormal hepatic function, pregnancy	685 (PG-guided $n = 352$, TAU $n = 333$) - Completers at 12 weeks: 579 patients (PG-guided $n = 297$, TAU $n = 282$) - Drop-out rate at 12 weeks: 15.5% (PG-guided: 15.6%, TAU: 15.3%)	- Sex (% females): 72.5 (PG-guided: 73, TAU: 72) - Age (years), mean (SD): 47.5, PG-guided: 47.8 (14.5), TAU: 47.3 (15.2) - Ethnicity (%): Caucasian (63), African-American (18), Hispanic (16), Other (2), Asian (1)	- Clinician-rated: HAM-D17, HAM-A (only for patients diagnosed with anxiety disorders), assessed at baseline 4, 8 and 12 weeks - Self-report: None - Adverse effects: ADE assessed at baseline 4, 8 and 12 weeks	- Outcomes: Symptom improvement, response and remission rate at 4, 8 and 12 weeks as measured by HAM-D17 and HAM-A	- In MDD patients, response rate and remission rate were higher in PG-guided group - In patients diagnosed with anxiety disorder, higher improvement in HAM-A scores at both 8 and 12 weeks along with in PG-guided group - No difference between groups in terms of adverse drug events
Han et al. (2018)	- 8-week, prospective, single blinded (patient blinded) RCT - Two university based teaching hospitals, Korea - Two arms: PG-guided vs. TAU - 3 Timepoints: Baseline, 4 and 8 weeks	- Age: ≥ 20 years - Diagnosis of MDD (DSM-5) - CGI-$I \geq 3$ - Subjects who required treatment substitution due to lack of efficacy or adverse events or intolerability - Exclusion: patients not currently on AD treatment; pregnancy or nursing; substance abuse or dependence within the past 12 months; unstable medical disorders; a current Axis I diagnosis of delirium, dementia, amnestic or other cognitive disorder, SZ or other psychotic disorder, BD I or II, ED, OCD, PD, or PTSD; a clinically significant current Axis II diagnosis; psychotic symptomatology in the current depressive episode; who received psychotherapy; hospitalization or having ECT within 8 weeks of the first visit	100 (PG-guided $n = 52$, TAU $n = 48$) - Completers at 8 weeks: 69 patients (PG-guided $n = 52$, TAU $n = 48$) - Drop-out rate at 8 weeks: 31.0% (PG-guided: 25.0%, TAU: 37.5%)	- Sex (% females): 74.9 (PG-guided: 76.9, TAU: 72.9) - Age (years), mean (SD): 44.0, PG-guided: 44.2 (16.1), TAU: 43.9 (13.8) - Etnicity (%): Korean (100)	- Clinician-rated: HAM-D17, CGI-S assessed at baseline 4 and 8 weeks - Self-report: PHQ-9/15, GAD-7, SDS assessed at baseline 4 and 8 weeks - Adverse effects: FIBSER, SAFTEE assessed at baseline 4 and 8 weeks.	- Primary outcome: mean change of total score of HAM-D17 from baseline to 8 weeks. - Co-primary outcome: change of total score of FIBSER from baseline to 8 weeks. - Secondary outcomes: response and remission rates at 8 weeks as measured by HAM-D17. Changes of total scores of PHQ-9/15, GAD-7, SDS, CGI-S from baseline to 8 weeks.	- Differences of response rates and symptoms improvement between PG-guided and TAU at week 8 - Differences of mean change in the FIBSER score favoring PG-guided group

(*continued on next page*)

Table 1 (*continued*)

Refs.	Experimental Design	Inclusion/Exclusion Criteria	Sample size	Demographics	Assessment	Outcomes	Main findings
Greden et al. (2019)	- 8-week, multicenter, prospective, patient-and rater-blinded, RCT - 60 academic and community sites including psychiatric and primary care providers, USA - Two arms: PG-guided vs. TAU - 5 Timepoints: Baseline, 4, 8, 12 and 24 weeks	- Age: $\geq$ 18 years - Diagnosis of MDD ($\geq$11 on the QIDS-C16 and self-rated QIDS-SR16) - An inadequate response (lack of clinical improvement or intolerable side-effects reported by the patient or treating clinician) to at least one documented psychotropic treatment for the current episode - Exclusion: A current Axis I diagnosis of delirium, dementia, amnestic or other cognitive disorder, SZ or other psychotic disorder, BD I or II, psychotic symptomatology within the current or prior depressive episodes; suicidal risk; significant substance use disorder; significant unstable medical condition or other significant medical conditions	1398 (PG-guided $n = 681$, TAU $n = 717$) - Completers at 8 weeks: 1167 patients (PG-guided $n = 560$, TAU $n = 607$) - Drop-out rate at 8 weeks: 16.5% (PG-guided: 17.8%, TAU: 15.3%)	- Sex (% females): 70.6 (PG-guided: 71.8, TAU: 69.5) - Age (years), mean (SD): 47.5 (14.5), PG-guided: 46.9 (14.5), TAU: 48.0 (14.5) - Ethnicity (%): Hispanic or Latino (7.9), not Hispanic or Latino (92.1)	- Clinician-rated: HAM-D17, QIDS-C16 assessed (via teleconference) at baseline, 4, 8, 12 and 24 weeks - Self-report: PHQ-9 assessed at baseline, 4, 8, 12 and 24 weeks - Adverse effects: Patient-reported side effects assessed at 8 week	- Primary outcome: Symptom improvement at 8 weeks as measured by HAM-D17 - Secondary outcomes: Response and remission rates at 8 weeks according with HAM-D17. Symptom improvement, response and remission rates at 8 weeks as measured QIDS-C16 and PHQ-9	- No differences in symptoms improvement at 8 weeks as measured by HAM-D17 - Higher response and remission rates at 8 weeks according with HAM-D17 in PG-guided group. - Higher in symptom improvement and response rate in PG-guided group at 8 weeks as measured by PHQ-9 - Higher remission rate in PG-guided group at 8 weeks as measured by QIDS-C16
Shan et al. (2019)	- 8-week, prospective, patient-and rater-blinded, RCT - Department of Psychiatry of the Second Xiangya Hospital, China - Two arms: PG-guided vs. TAU - 4 Timepoints: Baseline, 2, 4 and 8 weeks	- Age: Range 18–51 years - Diagnosis of MDD (DSM-5) - HAMD-17 $\geq$ 17 and the first item of the HAMD-17 (depressive mood) $\geq$ 2; who have never received psychiatric treatment or have interrupted AD medication for more than 2 weeks; no psychotic symptoms - Exclusion: Any other psychiatric diagnoses; any significant physical illness; pregnancy	71 (PG-guided $n = 31$, TAU $n = 40$) - Completers at 8 weeks: 48 patients (PG-guided $n = 21$ TAU $n = 27$) - Drop-out rate at 8 weeks: 32.4% (PG-guided: 32.3%, TAU: 32.5%)	- Sex (% females): 63.1 (PG-guided: 61.2, TAU: 65) - Age (years), mean (SD): 27.7, PG-guided: 26.5 (7.9), TAU: 28.8 (8.9) - Etnicity (%): Asian (Han population) (100)	- Clinician-rated: HAM-D17, HAM-A assessed at baseline and 8 weeks - Self-report: None - Adverse effects: TESS assessed at 8 weeks	- Outcomes: Mean change of total score of HAM-D17 from baseline to 8 weeks. Response and remission rates at 8 weeks as measured by HAM-D17.	- No significant difference in HAMD- 17 total scores, response and remission rates - No significant difference in the HAM-A total scores at each timepoint

(*continued on next page*)

Table 1 (*continued*)

Refs.	Experimental Design	Inclusion/Exclusion Criteria	Sample size	Demographics	Assessment	Outcomes	Main findings
Perlis et al. (2020)	- 8-week, multicenter, prospective, patient-and rater-blinded, RCT - Sites of recruitment not specified, USA - Two arms: PG-guided vs. TAU - 5 Timepoints: Baseline, 2, 4, 6, and 8 weeks	- Age: between 18 and 75 years - Diagnosis of nonpsychotic MDD based on DSM-5 and MINI 7.0 - SIGH-D-17 >18 - Fail of at least one prior adequate trial of AD for the current episode due to inefficacy or intolerable adverse effects - Exclusion: A current DSM-5 diagnosis of neurocognitive disorders, SZ spectrum (lifetime diagnosis) and other psychotic disorders, bipolar and related disorders (lifetime diagnosis), trauma and stress-related disorders, OCD and related disorders, personality disorders, PD; substance related and addictive disorders diagnosed in the last 12 months; history of suicidal behavior within 12 months; four or more failed AD in the current episode, ECT or rTMS or psychotherapy initiated within 90 days; unstable or active medical condition(s); pregnancy (or planning) or nursing	304 (PG-guided $n = 151$, TAU $n = 153$) - Completers at 8 weeks: 281 patients (PG-guided $n = 140$ TAU $n = 141$) - Drop-out rate at 8 weeks: 7.6% (PG-guided: 7.3%, TAU: 7.8%)	- Sex (% females): 71.7 (PG-guided: 70.9, TAU: 72.5) - Age (years), mean (SD): 47.7 (12.2), PG-guided: 47.8 (12.3), TAU: 47.6 (12.0) - Ethnicity (%): White (72.7), black/African American (23.4), America Indian or Alaskan Native (1), Native Hawaiian/Pacific Islander (1), Asian (0.3), other (1.6)	- Clinician-rated: SIGH-D-17, CGI-I, C-SSRS assessed at baseline, 2, 4, 6, and 8 weeks - Self-report: QIDS-SR16 assessed at baseline, 2, 4, 6, and 8 weeks - Adverse effect: FIBSER assessed at baseline, 2, 4, 6, and 8 weeks	- Primary outcome: Change from baseline in SIGH-D-17 at 8 weeks - Secondary outcomes: Response and remission rates at 8 weeks as measured by SIGH-D-17. Changes of total scores of QIDS-SR16, CGI-I, FIBSER from baseline	- No significant differences between PG-guided and TAU at week 8

(*continued on next page*)

Table 1 (*continued*)

Refs.	Experimental Design	Inclusion/Exclusion Criteria	Sample size	Demographics	Assessment	Outcomes	Main findings
Tiwari et al. (2022)	- 36-week, multicenter, prospective, patient-and rater-blinded, RCT - 8 academic and community sites including psychiatric and primary care providers, Ontario, Canada - Three arms: PG-guided and PG-EGEN-guided* vs. TAU - 7 Timepoints: Baseline, 4, 8, 12, 24, 36, and 52 weeks	- Age: ≥ 18 years - Diagnosis of nonpsychotic MDD based on DSM-IV-TR and ≥11 on the QIDS-C16 and self-rated QIDS-SR16 - An inadequate response (lack of clinical improvement or intolerable side-effects reported by the patient or treating clinician) to at least one documented psychotropic treatment for the current episode - Exclusion: A current Axis I diagnosis of delirium, dementia, amnestic or other cognitive disorder, SZ or other psychotic disorder, BD I or II, psychotic symptomatology within the current or prior depressive episodes; suicidal risk; significant substance use disorder; significant unstable medical condition or other significant medical conditions, currently receiving or scheduled to receive ECT, DBS, or TMS during course of study; pregnant or lactating.	276 (PG-guided $n = 90$, PG-EGEN-guided*= 93; TAU $n = 93$) - Completers at 8 weeks: 202 patients (PG-guided $n = 68$, PG-EGEN-guided*= 63; TAU $n = 71$) - Drop-out rate at 8 weeks: 26.8% (PG-guided: 24.4%, PG-EGEN-guided*=32.3%, TAU: 23.7%)	- Sex (% females): 64.5 (PG-guided: 65.6, PG-EGEN-guided*: 64.5, TAU: 63.4) - Age (years), mean (SD): 41.1 (14.1), PG-guided: 40.3 (15.3), PG-EGEN-guided*: 40.7 (12.9), TAU: 42.3 (14.2) - Ethnicity (%): Caucasian (89.2), Asian (7.5), Black (1.1), Latin American (2.2)	- Clinician-rated: HAM-D17 assessed (via telephone) at baseline, 4, 8, 12, 24, 36, and 52 weeks. - Self-report: PHQ-9, QIDS-SR16 assessed at baseline, 4, 8, 12, 24, 36, and 52 weeks.	- Primary outcome: Symptom improvement at 8 weeks as measured by HAM-D17 - Secondary outcomes: Response and remission rates at 8 weeks according with HAM-D17. Symptom improvement, response and remission rates at 24 weeks as measured by HAM-D17	- No differences in symptoms improvement at 8 weeks as measured by HAM-D17 - No differences in response and remission rates at 8 weeks according with HAM-D17. - No differences in symptoms improvement, response and remission rates at 24 weeks according with HAM-D17.

Acronym table

AD: Antidepressant Drugs; ADE: Adverse Drug Events; BD: Bipolar Disorder; CGI-I: Clinical Global Impression-Improvement CGI-S: Clinical Global Impression-Severity; C-SSRS: Columbia-Suicide Severity Rating Scale; DSM: Diagnostic and Statistical manual of Mental disorders; ECT: Electroconvulsive Therapy; ED: Eating Disorder; FIBSER: Frequency, Intensity, and Burden of Side Effects Ratings; GAD-7: General Anxiety Disorder-7 items; HAM-A: Hamilton Rating Scale for Anxiety; HAM-D17: Hamilton Rating Scale for Depression (17 items); MDD: Major Depressive Disorder; MINI: Mini-International Neuropsychiatric Interview; OCD: Obsessive-Compulsive Disorder; PD: Panic Disorder; PG: Pharmacogenetic; PGI-I: Patient Global Impression of Improvement: PHQ-9/15: Patient Health Questionnaire-9/15; PTSD: Post Traumatic Stress Disorder; QIDS-C16: Quick Inventory of Depressive Symptomatology Clinician-rated (16 items); QIDS-SR16: Quick Inventory of Depressive Symptomatology Self-rated (16 items); RCT: Randomized Controlled Trial; rTMS: repetitive Transcranial Magnetic Stimulation; SAFTEE: Systematic Assessment for Treatment Emergent Events-Systematic Inquiry; SATMED-Q: Treatment Satisfaction with Medicines Questionnaire; SD: Standard Deviation; SDI: Sheehan Disability Inventory; SDS: Sheehan Disability Scale; SIGH-D: Structured Interview Guide for the Hamilton Depression Rating Scale; SZ: Schizophrenia; TAU: Treatment As Usual; TESS: Treatment Emergent Symptom Scale.

Note: PG-EGEN-guided*: PG-EGEN report included 6 additional genes shown to have genetic variation associated with antipsychotic-induced weight gain.

Guide for the Hamilton Depression Rating Scale -SIGH-D-17 version). In addition, each study used other clinician-rated scales for the assessment of depression-related symptoms (for details see Table 1).

Five studies (Greden et al., 2019; Han et al., 2018; Pérez et al., 2017; Perlis et al., 2020; Tiwari et al., 2022) used also self-report scales for the evaluation of a wide range of constructs, ranging from the well-being of patients, anxiety symptoms, the disability caused by the disease, treatment satisfaction or patients' general impression of improvement. However, only in the two studies (Perlis et al., 2020; Tiwari et al., 2022) a validated self-report scale for the assessment of depressive symptomatology has been used, in order to also understand patients' perspectives.

Regarding the evaluation of side effects, five studies (Bradley et al., 2018; Han et al., 2018; Pérez et al., 2017; Perlis et al., 2020; Shan et al., 2019) assessed adverse events and tolerability using different scales and forms with the exception of the Greden et al. (2019) study in which patients' medical records were used to evaluate the mean number of side effects and the proportion of patients reporting side effects.

(f) Outcomes

Most of the RCT studies specified as primary outcome the symptom improvement at 8 weeks as measured by the mean change of the main clinician-rated scale score (Greden et al., 2019; Han et al., 2018; Pérez et al., 2017; Perlis et al., 2020; Tiwari et al., 2022), whereas all the other evaluations were indicated as secondary or tertiary outcomes. These can include (1) response and remission rates at different timepoints, as well as symptom improvement as measured at different timepoints (excluding 8-weeks as primary outcome) according to main clinician-rated scale used; (2) changes in scores of depressive symptoms, as well as response and remission rates at different timepoints assessed with a self-rating scale, (3) changes in scores of depressive-related symptoms, such as anxiety, at different timepoints; (4) side effects at different time-points. The response is defined as $a \geq 50\%$ decrease of points on a clinical scale of interest at one specific timepoint compared with the baseline. Remission is defined as a score lower of a specific cut-off, defined accordingly by the scoring parameters of the assessment scale used.

(g) Main findings

Concerning the results obtained in relation to the declared primary outcomes, only in Han et al. (2018) study the patients allocated to the PG-guided group showed higher symptom improvement at 8 weeks from the beginning of the treatment, whereas no differences were obtained in the other studies (Greden et al., 2019; Pérez et al., 2017; Perlis et al., 2020; Tiwari et al., 2022). Regarding the other outcomes, several significant results favouring PG-guided groups were found in several RCTs (Bradley et al., 2018; Greden et al., 2019; Han et al., 2018; Pérez et al., 2017). In particular, in three RCTs the response and/or the remission rates assessed with the clinical-rated scale used at 8 weeks, were higher in PG-guided groups (Bradley et al., 2018; Greden et al., 2019; Han et al., 2018). In the Perez's study a higher response rate at 12 weeks in the PG-guided group as measured by PGI-I was reported (Pérez et al., 2017). In addition, the Greden et al. (2019) study reported higher symptom amelioration and response and remission rates at 8 weeks as measured by PHQ-9 and by QIDS-C16, respectively, in favor of PG-guided patients. Two studies (Greden et al., 2019; Tiwari et al., 2022) reported a similar increase in clinicians prescribing congruent medications for patients in the PG-guided arm, but not in the TAU arm. Moreover, Han et al. (2018), Pérez et al. (2017) studies reported better tolerability and lower side effects as measured by the Frequency, Intensity, and Burden of Side Effects Ratings (FIBSER) scale in PG-guided patients. Finally, in Greden et al. (2019), in which patients were evaluated over the full 24-week study period, the outcomes of the PG-guided patient group continued to improve through 24 weeks, showing that the rate of remission nearly doubled from week 8 to week 24 (data reported and extended also in Thase et al. 2019). The authors conclude that this observation supports that PG testing may provide durability in antidepressant treatment effects. However, this result presents with relevant limitations since only PG-guided patients were observed for a longer time interval and no comparison was shown with the TAU group given the unblinded study design between 8 and 24 weeks was applied.

Some post-hoc analyses were carried out. In Pérez et al. (2017) study, post-hoc stratified analyses were performed on the basis of severity of depression episode as well as for the number of previous antidepressant medication failure. The results showed that at 12-week visit, the response rate was higher in the PG-guided group compared to the TAU group for patients diagnosed with severe depression. Moreover, patients having received 1 to 3 previous failed psychiatric treatments in the current episode showed a small clinical benefit compared to TAU as seen by Cohen's d calculated from the change in HDRS-17, whereas drug naïve subjects and those having received 4 or more medication trials did not. Furthermore, among subjects with 1 to 3 treatment failures, statistically significant differences were identified at 12 weeks in the response rate based on the PGI-I score, and on the HDRS-17 score both at 6 weeks and 12 weeks, in favor of the PG-guided patient group.

In Bradley et al. (2018) study, post-hoc analyses were performed stratified for the severity of the current depressive episode. Both at 8 and 12-week follow-up, the response and remission rates were higher in the PG-guided group compared to the TAU patients diagnosed with severe depression. Similarly, when both moderate and severe patients were included in the analysis at 8 and 12-week response rates remained significantly higher for the PG-guided group of patients. No significant improvements were found in patients with mild depression.

In Perlis et al. (2020) study, a post-hoc exploratory analysis showed that at 8 weeks, significantly more patients had failed to improve or worsen (by at least one point on the SIGH-D-17) in the TAU group compared to the PG-guided one. Unfortunately, only one study (Pérez et al., 2017) applied multiple-testing correction in their analysis.

5. Discussion

The RCTs on PG testing for antidepressants included in this review showed several biases that should be overcome with well-designed future trial studies on this issue. Although all the selected studies described above are RCTs with interesting data, a series of methodological limitations reduce, at least in part, the relevance and the generalization of the results achieved.

First, all RCTs except two studies (Han et al., 2018; Shan et al., 2019) have been conducted by PG test manufacturers, leading to a significant industry bias. Consequently, the Korean (Han et al., 2018) and the Chinese (Shan et al., 2019) studies can be considered the first two non-industry sponsored trials.

Second, concerning the study methodology, the seven RCTs have, overall, low risks of bias regarding the random sequence generation (selection bias), the allocation concealment (selection bias) and incomplete outcome data (attrition bias), while they present high risk of bias regarding: (i) blinding of participants, participating personnel (performance bias), where patients were blind to the study group, but treating clinician were not blinded (Bradley et al., 2018; Greden et al., 2019; Han et al., 2018; Pérez et al., 2017; Shan et al., 2019; Tiwari et al., 2022); (ii) blinding of outcome assessors (detection bias), where the clinical assessment was performed by an unblinded rater (Pérez et al., 2017) or by the treating clinician (Han et al., 2018); (iii) selective outcome reporting (reporting bias), where results for remission were reported only in a subset of the sample, rather than for the entire sample (Bradley et al., 2018); and (iv) recruitment bias, where patients were recruited by the treating clinicians (Bradley et al., 2018; Greden et al., 2019; Han et al., 2018; Pérez et al., 2017; Perlis et al., 2020; Shan et al., 2019; Tiwari et al., 2022), rather than by site-independent investigators.

Concerning the study design, almost all of the studies declared a double-blinded design, but often they were only partially double-blinded, where for example the prescriber was not blind, while the rater and the patient were blind to the study group (Bradley et al., 2018; Greden et al., 2019; Han et al., 2018; Pérez et al., 2017; Perlis et al., 2020; Shan et al., 2019; Tiwari et al., 2022). In another study both the prescriber and the rater were not blinded for the assessment (Pérez et al., 2017). The lack of prescriber and / or rater blindness does not allow the exclusion of a possible influence on the outcome assessment (performance and / or detection bias), therefore strategies to enable complete blinding of patients, raters, and prescribers deserve further investigation in order to minimize observer-expectancy effects.

Another relevant point to be taken into account is the clinical evaluation time point, since the studies are not always homogeneous regarding time assessment. It is relevant to consider clinical outcomes after 8 weeks, the typical duration of acute phase depression treatment, and to extend the evaluation for 12 or 24 weeks whenever possible, since clinical contexts may change considerably over time and response duration and sustained remission are important issues to be evaluated (Frieden, 2017). To date, only one RCT (Tiwari et al., 2022) showed a longer blinded observation until 36 weeks, however, data published from this study are concerning a time period of 24 weeks without any significant main results.

Regarding the inclusion and exclusion criteria, RCTs should be more homogeneous and they should take into account some relevant aspects in the decision-making process of sampling subjects with MDD. In particular, the presence or absence of the following aspects should be investigated and verified: the depressive symptom profile, the clinical subtypes (with melancholic, atypical or mixed features, or with anxious distress, or with psychotic features), the seasonality, the episode severity (mild, moderate or severe), the suicidality, the clinical staging of depression (first episode, residual phase, recurrent and chronic MDD), the presence of personality traits or a full-blow diagnosis of a personality disorder, an antecedent and / or concomitant psychiatric comorbidity, including alcohol and / or substance abuse, physical comorbidities, early and / or recent adverse life events, a family history of a homotypic and / or heterotypic psychiatric condition (Maj et al., 2020). Despite all but one (Greden et al., 2019) of the selected RCTs used the DSM-IV-TR or DSM-5 criteria for diagnosing MDD, most of them have utilized different inclusion and exclusion criteria, leading to an increased population heterogeneity and to a decrease of results generalization. In particular, no studies have considered the depressive symptom profile, the depression subtype, the clinical staging of depression, the seasonality, the psychiatric family history (taken into account only by Han et al. 2018), early and / or recent adverse life events, while other parameters, such as clinical severity, the psychiatric and physical comorbidity, and suicidality were taken into account differently among studies. For example, Pérez et al. (2017) reported not to have excluded comorbidities including anxiety disorders (also included in Bradley et al. 2018), post-traumatic stress disorders, substance abuse disorders, but did not consider comorbidities in relation to the findings in the different treatments group.

It is also interesting to consider the differences among the studies regarding the subjects who could be included, with respect to whether or not they were on antidepressant treatment and with respect to the number of previous antidepressant therapy failures. More in detail, some studies allowed the participation of subjects who required medication de novo (Bradley et al., 2018; Pérez et al., 2017; Shan et al., 2019), while others included only subjects who were under treatment and required substitution due to lack of efficacy or treatment discontinuation due to adverse events or intolerability (Greden et al., 2019; Han et al., 2018; Perlis et al., 2020; Tiwari et al., 2022). In this perspective, a number of studies (Greden et al., 2019; Han et al., 2018; Pérez et al., 2017; Perlis et al., 2020; Tiwari et al., 2022) had at least two or three previous failed antidepressant treatments for the current MDD episode in their enrolled patients, therefore a percentage of the depressed population could be regarded as having treatment resistant depression (TRD) (Fava, 2003). This recruitment modality does not lead to solving the doubt, in terms of finding a good compromise for cost-benefit, whether genetic testing should be reserved for patients with treatment resistance/sensitivity or if could be better to perform the PG test prior to the beginning of the first antidepressant trial (Zeier et al., 2018).

Concerning the sample size, one study had a small sample size and could have been underpowered (Shan et al., 2019), while in four others the sample power was calculated (Han et al., 2018; Pérez et al., 2017; Perlis et al., 2020; Tiwari et al., 2022), although only in Perlis et al. (2020) study the completers and the attrition rate were taken into account. It should also be noted that, although some studies show a high drop-out rate (Han et al., 2018; Shan et al., 2019; Tiwari et al., 2022), it amounted to a similar level both in the PG-guided and in the TAU groups.

Regarding the demographic characteristics, the majority of studies included individuals of Caucasian origin, with African, American, Hispanic/Latin American and Asian individuals representing less than 20% of the population analysed, one study was based on a Korean population (Han et al., 2018), another was based on a Chinese population (Shan et al., 2019), which limits the generalizability of the findings. On the other hand, an additional evaluation of more diverse populations would be beneficial and replication studies with patients of diverse ancestral origins can further increase confidence in the broader utility of the findings (Rosenblat et al., 2018).

About the assessment and outcomes, they are not standardized and homogeneous among the different RCTs and relevant real-world clinical outcomes, such as cognitive symptoms or psychosocial impairment (Maj et al., 2020; McIntyre et al., 2015), were not considered at all. Moreover, the side effects were not evaluated in all studies with validated and structured tools. Finally, not in all RCTs, self-report scales for depressive symptoms were used, losing the patients' perspective that is extremely important in order to achieve a real functional remission of depressive symptomatology.

6. Conclusions and future directions

Personalization of psychiatric treatments using pharmacogenetic information is emerging as a valuable tool to identify in advance which medications will be more effective, which ones will require dose adjustments or which ones may cause meaningful adverse reactions (Serretti and Fabbri, 2020; Shalimova et al., 2021). The growing number of arrays of PG tests poses clinical implementation challenges. Indeed, besides a clinically demonstrated effect on efficacy and tolerability, PG tests need to represent a real benefit for MDD patients. To achieve that goal and in order to ensure a real-world utility and applicability, PG tests needs to be evaluated in high quality RCT studies with adequate control groups and blinded ratings that have the potential to support generalizability of the results and to evaluate the economic cost-benefit ratio in healthcare systems.

Concerning the above discussed points, some critical notes need to be added. In all the reviewed RCT studies the treating clinician was not blinded to the study arm. This was necessitated by the ethical issues of mandating prescribed medications in order to blind clinicians. To mitigate this limitation, it is extremely important that all the RCTs should be rigorous concerning the blindness of patients and raters for all the evaluation scales used. In this regard, a multi-centric, double-blinded RCT design, in terms of patient and rater blinded design, should represent the gold standard for evidence generation, while observational studies, even though having the advantage of being carried out in a naturalistic scenario, present several limitations typical of this kind of design.

The majority of the larger cohorts studied in RCTs was Caucasian and this ethnic bias represents a strong limit to generalize the results to all populations, since there is a large variability in the variants frequencies of the genes generally included in PG tests. Consequently, larger RCT studies on different ethnicities are needed.

Although sustained clinical remission is the ideal objective of treatment for patients with MDD, most RCTs on PG tests have ended after 8 weeks, the typical duration of acute phase treatment. In the RCTs lasting 12 weeks (Bradley et al., 2018; Pérez et al., 2017) significant effects in favor of the PG-guided patients in terms of amelioration of symptoms were found at this timepoint. Moreover, in the Greden et al. (2019) study, the remission rate doubled from week 8 to week 24 among patients in the PG-guided arm, but this study period was unblinded. These data suggest that improved patient outcomes achieved with PG testing could be durable in the maintenance of therapy settings. Larger and longer-term RCT studies need to be performed to capture the impact of PG tests, and also to aid in combining results of randomized trials with those of longer-term cost-effectiveness investigations.

Another key point to be commented concerns the data sampling used in the analysed RCTs. Indeed, the results obtained cannot be generalized to the entire population of MDD patients that need to be treated. Most patients in the TAU arm were prescribed medications that were congruent with the PG test report. Moreover, in most RCTs the PG tests did not improve the outcomes investigated. So, the prescription of psychiatrists on the basis of their clinical knowledge was sufficient. However, in RCTs that stratified the analysis for patients with a large number of unsuccessful medication attempts, most of them could probably be classified as having treatment resistant depression, a modest but higher rates of response and remission for the PG-guided group were reported, and the largest effect sizes emerged from a post-hoc analysis of the subset of patients with severe depression (Pérez et al., 2017; Bradley et al., 2018; Perlis et al., 2020). Instead, no effects were shown for drug-naïve patients. This highlights the significant clinical challenges in this difficult-to-treat population. It can be proposed that PG testing could be predominantly useful for those patients who carry functional variants ("pharmacogenes") related to a greater vulnerability to develop treatment resistance and/or drugs adverse effects compared to those without. Therefore, PG tests may be a useful and viable treatment tool option for such difficult-to-treat MDD patients since they pursue a precision medicine strategy that maximizes the benefits but minimizes the cost of the use of complex pharmacogenomic analyses that would be justified if the traditional first lines of treatment fail. This is extremely important to understand because these individuals are usually more likely to consume health care resources. Future studies should attempt to assess this hypothesis.

A further point is related to the tolerability and safety of antidepressants, which is a global challenge for psy-

chiatry, since it has been linked to poorer adherence and symptom improvement and other disease-related outcomes (Cipriani et al., 2018; Sharma et al., 2019). To date, most genetic variants included in all types of PG tests have been selected mainly due to significant associations with reduction of side effects rather than an increase in efficacy of antidepressant treatments (Fabbri and Serretti, 2020; Zeier et al., 2018). For this reason, a correct interpretation of PG testing may result in reducing the risk of adverse events and consequently may improve adherence. However, the RCTs performed show a lack of generalizable results concerning this issue and a gold standard in terms of objective assessment of side effects through validated scales should be addressed. Indeed, there is still a strong debate about the clinical utility of PG testing, in large part due to the lack of evidence in reducing side effects.

In addition, negative clinical predictors of response such as severity, suicide ideation, anxiety symptoms, previous drugs failure attempts, presence of cognitive symptoms or functional impairment, should be investigated and analysed in all RCTs focusing on the PG test for antidepressants in MDD.

Finally, it is necessary to pay attention to the education of mental healthcare professionals for being able to help educate patients about PG testing. Indeed, many mental healthcare members may not be fully aware of what pharmacogenetics could offer in their healthcare setting, or how to use the results if they are available. Both healthcare professionals and patients should be well informed about the PG testing process and its limitations, especially with respect to evaluating the evidence supporting the genes, indication of specific tests, how to interpret the test, and how to integrate its results into practice in conjunction with clinical expertise. Despite the potential of being able to tailor medication to a patient's genetic profile is a widespread notion both among clinicians and patients, the acceptance of PG data varies among physicians themselves, mainly because many of them still express lower levels of confidence and knowledge of the process behind a PG test (Vest et al., 2020; Zanardi et al., 2021b). The knowledge gained by the physicians can help reassure patients by addressing their concerns regarding PG testing.

In summary, a number of barriers have been noted for the widespread adoption of PG tests for antidepressants into clinical care for the treatment of MDD patients. Indeed, only seven RCTs have investigated a possible relationship between PG testing and outcomes in terms of antidepressants efficacy and reduction of adverse effects. The quality of the study designs in these RCTs is poor, resulting in weak methodology and limited scope that do not allow us to establish strong and conclusive evidence at this stage. Moreover, although the results of some RCTs indicate a relationship between the use of the PG testing and a reduction of side effects, the findings are preliminary and further exploration is required. However, some positive results coming from analyses in subgroups of patients, such as more benefits using PG testing for patients with severe MDD episode and a greater amelioration in longer period of observation, indicate new perspectives to develop further RCTs. Furthermore, it would be important to have the opportunity to carry out an individual meta-analysis on existing data. In this way, studies that contain the same or overlapping sets of participants can be identified and cluster analyses can be performed at different timepoints. In addition, algorithms that integrate different assessment scales through the selection of specific items are available and might be used to increase the statistical power for some depression-related symptoms outcomes. Finally, to date no studies focus on the subgroups of patients who carry some functional variants of genes of poor or extensive metabolisers.

Efforts to better understand the subset of individuals who may derive benefit from the tests, the time course over which such benefits may be identified, and for which kind of real-world outcomes the tests may be applied, represent important next steps for MDD PG test RCT studies for antidepressants, in term of efficacy, increase of adherence, cost-effective and cost-saving strategy.

Role of funding source

No funding.

Contributors

Author AM and SB contributed designing the study and the search protocol, screening the references, extracting data, assessing methodological quality, drafting and finalizing the manuscript. Author BTB contributed designing the study, drafting and finalizing the manuscript.

Declaration of Competing Interest

Bernhard T Baune has consulting roles with the National Health and Medical Research Council, Australia. He has received honoraria from AstraZeneca, Bristol-Myers Squibb, Lundbeck, Pfizer, Servier, Wyeth, Otsuka; Research or grants from private industries or non-profit funds from AstraZeneca, Sanofi-Synthélabo; and research grants from the National Health and Medical Research Council, Australia. He has served on advisory boards for Janssen-Cilag, Lundbeck and received funds from the Fay Fuller Foundation, and James & Diana Ramsay Foundation, Adelaide.

Acknowledgement

None.

References

Bousman, C.A., Arandjelovic, K., Mancuso, S.G., Eyre, H.A., Dunlop, B.W., 2019. Pharmacogenetic tests and depressive symptom remission: a meta-analysis of randomized controlled trials. Pharmacogenomics 20, 37–47. doi:10.2217/pgs-2018-0142.

Bradley, P., Shiekh, M., Mehra, V., Vrbicky, K., Layle, S., Olson, M.C., Maciel, A., Cullors, A., Garces, J.A., Lukowiak, A.A., 2018. Improved efficacy with targeted pharmacogenetic-guided treatment of patients with depression and anxiety: a randomized clinical trial demonstrating clinical utility. J. Psychiatr. Res. 96, 100–107. doi:10.1016/j.jpsychires.2017.09.024.

Cipriani, A., Furukawa, T.A., Salanti, G., Chaimani, A., Atkinson, L.Z., Ogawa, Y., Leucht, S., Ruhe, H.G., Turner, E.H., Higgins, J.P.T., Egger, M., Takeshima, N., Hayasaka, Y., Imai, H., Shinohara, K., Tajika, A., Ioannidis, J.P.A., Geddes, J.R., 2018.

Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. Lancet 391, 1357–1366. doi:10.1016/S0140-6736(17)32802-7.

Fabbri, C., Serretti, A., 2020. Genetics of treatment outcomes in major depressive disorder: present and future. Clin. Psychopharmacol. Neurosci. 18, 1–9. doi:10.9758/cpn.2020.18.1.1.

Fabbri, C., Zohar, J., Serretti, A., 2018. Pharmacogenetic tests to guide drug treatment in depression: comparison of the available testing kits and clinical trials. Prog. Neuropsychopharmacol. Biol. Psychiatry 86, 36–44. doi:10.1016/j.pnpbp.2018.05.007.

Fava, M., 2003. Diagnosis and definition of treatment-resistant depression. Biol. Psychiatry 53, 649–659. doi:10.1016/s0006-3223(03)00231-2.

Frieden, T.R., 2017. Evidence for health decision making - beyond randomized, controlled trials. N. Engl. J. Med. 377, 465–475. doi:10.1056/NEJMra1614394.

Gratten, J., Wray, N.R., Keller, M.C., Visscher, P.M., 2014. Large-scale genomics unveils the genetic architecture of psychiatric disorders. Nat. Neurosci. 17, 782–790. doi:10.1038/nn.3708.

Greden, J.F., Parikh, S.V., Rothschild, A.J., Thase, M.E., Dunlop, B.W., DeBattista, C., Conway, C.R., Forester, B.P., Mondimore, F.M., Shelton, R.C., Macaluso, M., Li, J., Brown, K., Gilbert, A., Burns, L., Jablonski, M.R., Dechairo, B., 2019. Impact of pharmacogenomics on clinical outcomes in major depressive disorder in the GUIDED trial: a large, patient- and rater-blinded, randomized, controlled study. J. Psychiatr. Res. 111, 59–67. doi:10.1016/j.jpsychires.2019.01.003.

Han, C., Wang, S.M., Bahk, W.M., Lee, S.J., Patkar, A.A., Masand, P.S., Mandelli, L., Pae, C.U., Serretti, A., 2018. A pharmacogenomic-based antidepressant treatment for patients with major depressive disorder: results from an 8-week, randomized, single-blinded clinical trial. Clin. Psychopharmacol. Neurosci. 16, 469–480. doi:10.9758/cpn.2018.16.4.469.

Hasin, D.S., Sarvet, A.L., Meyers, J.L., Saha, T.D., Ruan, W.J., Stohl, M., Grant, B.F., 2018. Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States. JAMA Psychiatry 75, 336–346. doi:10.1001/jamapsychiatry.2017.4602.

Maj, M., Stein, D.J., Parker, G., Zimmerman, M., Fava, G.A., De Hert, M., Demyttenaere, K., McIntyre, R.S., Widiger, T., Wittchen, .H..-U., 2020. The clinical characterization of the adult patient with depression aimed at personalization of management. World Psychiatry 19, 269–293. doi:10.1002/wps.20771.

McIntyre, R.S., Lee, Y., Mansur, R.B., 2015. Treating to target in major depressive disorder: response to remission to functional recovery. CNS Spectr 20 (Suppl 1), 20–30. doi:10.1017/S1092852915000826, quiz 31.

Mrazek, D.A., Biernacka, J.M., McAlpine, D.E., Benitez, J., Karpyak, V.M., Williams, M.D., Hall-Flavin, D.K., Netzel, P.J., Passov, V., Rohland, B.M., Shinozaki, G., Hoberg, A.A., Snyder, K.A., Drews, M.S., Skime, M.K., Sagen, J.A., Schaid, D.J., Weinshilboum, R., Katzelnick, D.J., 2014. Treatment outcomes of depression. J. Clin. Psychopharmacol. 34, 313–317. doi:10.1097/JCP.0000000000000099.

Pérez, V., Salavert, A., Espadaler, J., Tuson, M., Saiz-Ruiz, J., Sáez-Navarro, C., Bobes, J., Baca-García, E., Vieta, E., Olivares, J.M., Rodriguez-Jimenez, R., Villagrán, J.M., Gascón, J., Cañete-Crespillo, J., Solé, M., Saiz, P.A., Ibáñez, Á., de Diego-Adeliño, J., Collaborative Group, AB-GEN, Menchón, J.M., 2017. Efficacy of prospective pharmacogenetic testing in the treatment of major depressive disorder: results of a randomized, double-blind clinical trial. BMC Psychiatry 17, 250. doi:10.1186/s12888-017-1412-1.

Perlis, R.H., Dowd, D., Fava, M., Lencz, T., Krause, D.S., 2020. Randomized, controlled, participant- and rater-blind trial of pharmacogenomic test-guided treatment versus treatment as usual for major depressive disorder. Depress. Anxiety doi:10.1002/da.23029.

Rosenblat, J.D., Lee, Y., McIntyre, R.S., 2018. The effect of pharmacogenomic testing on response and remission rates in the acute treatment of major depressive disorder: a meta-analysis. J. Affect. Disord. 241, 484–491. doi:10.1016/j.jad.2018.08.056.

Rush, A.J., Trivedi, M.H., Wisniewski, S.R., Nierenberg, A.A., Stewart, J.W., Warden, D., Niederehe, G., Thase, M.E., Lavori, P.W., Lebowitz, B.D., McGrath, P.J., Rosenbaum, J.F., Sackeim, H.A., Kupfer, D.J., Luther, J., Fava, M., 2006. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. Am. J. Psychiatry 163, 1905–1917. doi:10.1176/ajp.2006.163.11.1905.

Serretti, A., Fabbri, C., 2020. The search for personalized antidepressant treatments: what have we learned and where are we going. Pharmacogenomics 21, 1095–1100. doi:10.2217/pgs-2019-0086.

Shalimova, A., Babasieva, V., Chubarev, V.N., Tarasov, V.V., Schiöth, H.B., Mwinyi, J., 2021. Therapy response prediction in major depressive disorder: current and novel genomic markers influencing pharmacokinetics and pharmacodynamics. Pharmacogenomics 22, 485–503. doi:10.2217/pgs-2020-0157.

Shan, X., Zhao, W., Qiu, Y., Wu, H., Chen, J., Fang, Y., Guo, W., Li, L., 2019. Preliminary clinical investigation of combinatorial pharmacogenomic testing for the optimized treatment of depression: a randomized single-blind study. Front. Neurosci. doi:10.3389/fnins.2019.00960.

Sharma, T., Guski, L.S., Freund, N., Meng, D.M., Gøtzsche, P.C., 2019. Drop-out rates in placebo-controlled trials of antidepressant drugs: a systematic review and meta-analysis based on clinical study reports. Int. J. Risk Saf. Med. 30, 217–232. doi:10.3233/JRS-195041.

Singh, A.B., 2015. Improved antidepressant remission in major depression via a pharmacokinetic pathway polygene pharmacogenetic report. Clin. Psychopharmacol. Neurosci. 13, 150–156. doi:10.9758/cpn.2015.13.2.150.

Tanner, J.-.A., Davies, P.E., Voudouris, N.C., Shahmirian, A., Herbert, D., Braganza, N., Gugila, A., Dechairo, B.M., Kennedy, J.L., 2018. Combinatorial pharmacogenomics and improved patient outcomes in depression: treatment by primary care physicians or psychiatrists. J. Psychiatr. Res. 104, 157–162. doi:10.1016/j.jpsychires.2018.07.012.

Thase, M.E., Nierenberg, A.A., Vrijland, P., van Oers, H.J.J., Schutte, A.J., Simmons, J.H., 2010. Remission with mirtazapine and selective serotonin reuptake inhibitors: a meta-analysis of individual patient data from 15 controlled trials of acute phase treatment of major depression. Int. Clin. Psychopharmacol. 25, 189–198. doi:10.1097/YIC.0b013e328330adb2.

Thase, M.E., Parikh, S.V., Rothschild, A.J., Dunlop, B.W., DeBattista, C., Conway, C.R., Forester, B.P., Mondimore, F.M., Shelton, R.C., Macaluso, M., Li, J., Brown, K., Jablonski, M.R., Greden, J.F., 2019. Impact of pharmacogenomics on clinical outcomes for patients taking medications with gene-drug interactions in a randomized controlled trial. J. Clin. Psychiatry 80. doi:10.4088/JCP.19m12910.

Tiwari, A.K., Zai, C.C., Altar, C.A., Tanner, J.A., Davies, P.E., Traxler, P., Li, J., Cogan, E.S., Kucera, M.T., Gugila, A., Braganza, N., Emmerson, H., Zai, G., Müller, D.J., Levitan, R., Kloiber, S., Daskalakis, Z.J., Frey, B.N., Bowen, J.M., Tarride, J.E., Tytus, R., Chandrasena, R., Voudouris, N., Taylor, V.H., Tempier, R., Sharma, V., Vasudev, A., Dzongowski, P., Pliamm, L., Greenspoon, T., Dechairo, B.M., Kennedy, J.L., 2022. Clinical utility of combinatorial pharmacogenomic testing in depression: a Canadian patient- and rater-blinded, randomized, controlled trial. Transl. Psychiatry 12, 101. doi:10.1038/s41398-022-01847-8.

Vest, B.M., Wray, L.O., Brady, L.A., Thase, M.E., Beehler, G.P., Chapman, S.R., Hull, L.E., Oslin, D.W., 2020. Primary care and mental health providers' perceptions of implementation of pharmacogenetics testing for depression prescribing. BMC Psychiatry 20, 518. doi:10.1186/s12888-020-02919-z.

Winner, J.G., Carhart, J.M., Altar, C.A., Allen, J.D., Dechairo, B.M., 2013. A prospective, randomized, double-blind study assessing the clinical impact of integrated pharmacogenomic testing for major depressive disorder. Discov. Med. 16, 219–227.

Zanardi, R., Manfredi, E., Montrasio, C., Colombo, C., Serretti, A., Fabbri, C., 2021a. Pharmacogenetic-guided treatment of depression: real-world clinical applications, challenges and perspectives. Clin. Pharmacol. Ther. doi:10.1002/cpt.2315.

Zanardi, R., Prestifilippo, D., Fabbri, C., Colombo, C., Maron, E., Serretti, A., 2021b. Precision psychiatry in clinical practice. Int. J. Psychiatry Clin. Pract. 25, 19–27. doi:10.1080/13651501.2020.1809680.

Zeier, Z., Carpenter, L.L., Kalin, N.H., Rodriguez, C.I., McDonald, W.M., Widge, A.S., Nemeroff, C.B., 2018. Clinical implementation of pharmacogenetic decision support tools for antidepressant drug prescribing. Am. J. Psychiatry 175, 873–886. doi:10.1176/appi.ajp.2018.17111282.