



REVIEW

Treatment-Resistant Depression in Real-World Clinical Practice: A Systematic Literature Review of Data from 2012 to 2022

Albino J. Oliveira-Maia · Ania Bobrowska · Eric Constant ·
Tetsuro Ito · Yerkebulan Kambarov · Hannah Luedke ·
Siobhán Mulhern-Haughey · Christian von Holt

Received: August 25, 2023 / Accepted: September 28, 2023 / Published online: October 26, 2023
© The Author(s) 2023

ABSTRACT

Objective: Real-world evidence in treatment-resistant depression (TRD; commonly defined as non-response to ≥ 2 consecutive treatments at adequate dosage and duration) is lacking. A systematic literature review was conducted to understand disease burden and treatment outcomes for patients with TRD, studied in a real-world setting over the last decade.

Data Sources: A literature search was conducted in May 2022 in MEDLINE, Embase, The

Cochrane Libraries and PsycINFO, comprising studies published from 2012 to 2022. Bibliographies of all relevant identified systematic reviews and relevant conference proceedings from 2020 to 2022 were manually hand-searched.

Study Selection: Real-world studies, including cohort, cross-sectional, case-control, chart review and registry studies, published in English and reporting outcomes in adults with TRD, were included.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12325-023-02700-0>.

A. J. Oliveira-Maia
Champalimaud Research and Clinical Centre,
Champalimaud Foundation, Lisbon, Portugal

A. J. Oliveira-Maia
Faculdade de Ciências Médicas, NOVA Medical
School, NMS, FCM, Universidade NOVA de Lisboa,
Lisbon, Portugal

A. Bobrowska
Costello Medical, Cambridge, UK

E. Constant
Centre Hospitalier Spécialisé Notre-Dame des Anges,
Liège, Belgium

E. Constant
Université Catholique de Louvain, Brussels,
Université de Liège, Liège, Belgium

T. Ito
Janssen EMEA, High Wycombe, UK

Y. Kambarov (✉)
Janssen Pharmaceutica NV, Beerse, Belgium
e-mail: YKambaro@its.jnj.com

H. Luedke
Costello Medical, London, UK

S. Mulhern-Haughey
Janssen EMEA, Dublin, Ireland

C. von Holt
Janssen EMEA, Neuss, Germany

Data Extraction: Extracted data included study and baseline disease characteristics, treatment type, treatment response, clinical outcomes and health-related quality of life.

Results: Twenty studies were included. Criteria for TRD varied, but patients typically experienced long-lasting depression (range 1.4 to 16.5 years). Across studies, mean disease severity scores demonstrated moderate to severe depression, reflecting a high burden of disease at baseline. Remission rates were typically low but generally increased with longer follow-up durations. However, the heterogeneity of interventions, follow-up durations (range 2 weeks to 9.4 years) and assessment tools precluded their quantitative synthesis. Studies were frequently limited by low sample size (range 14 to 411 patients) and health-related quality of life was infrequently assessed.

Conclusions: There is a lack of clinical consensus regarding the definition, assessment and monitoring of TRD in real-world practice. Nevertheless, TRD carries a high burden of illness and there is an unmet need for faster and more effective treatments. To better understand the personal burden of affected patients, future studies would benefit from standardisation of severity assessment and measures of treatment effectiveness, as well as greater consideration of health-related quality of life.

PLAIN LANGUAGE SUMMARY

Many people continue to experience depression even after trying two or more medications. This is called treatment-resistant depression (TRD). Most of the information we have on TRD comes

from clinical trials, which take place under tightly-controlled conditions. It is important to understand the effects of TRD and TRD treatments on people in their day-to-day lives. Researchers studying people's day-to-day lives call this researching in a "real-world setting". We searched for studies carried out in real-world settings in the last 10 years. We found 20 relevant studies. As these studies were in real-world settings, there were many differences between them, including differences in how TRD was diagnosed, the treatments used, how long people were monitored and how results were measured. This made it difficult to compare how successful different treatments were. Most studies included a small number of people and monitored them for a relatively short time. We found people with TRD had usually lived with it for many years and their symptoms were moderate or severe. Only two studies asked people how TRD affected their lives. These two studies found health-related quality of life and work productivity was low. Most studies found lots of people still had symptoms of depression after treatment. However, symptoms typically improved more when studies monitored people for a longer time. To improve our knowledge of TRD, future studies should monitor more people for longer and use the same ways of measuring results. They should also ask how TRD affects people's daily lives.

Keywords: Major depressive disorder; Real-world evidence; Systematic literature review; Treatment-resistant depression

Key Summary Points

A systematic literature review of real-world studies in patients with treatment-resistant depression (TRD) identified 20 studies conducted over the last 10 years.

We found that there was a lack of clinical consensus regarding the definition, assessment and monitoring of TRD in real-world practice.

At baseline, patients with TRD presented with moderate to severe depression and had typically experienced long-lasting major depressive episodes spanning multiple years; while seldom measured, health-related quality of life was apparently low.

Rates of response to treatment varied greatly between studies but were generally low; few studies investigated long-term outcomes, but those that did typically reported marginally greater rates of remission than those reporting on short-term outcomes.

Future real-world studies would benefit from standardised modalities of assessment, monitoring and reporting of treatment effectiveness, including greater consideration of health-related quality of life outcomes.

INTRODUCTION

Major depressive disorder (MDD) is one of the most common, yet debilitating, psychiatric disorders, characterised by persistently low mood and energy; anhedonia; changes in appetite, weight and sleep; fatigue; and suicidality, among other symptoms [1, 2]. The lifetime prevalence of MDD among the general population is estimated to be ~ 13 to 15%, with first-line treatments consisting of antidepressant medications, behavioural psychotherapy,

or a combination thereof [3, 4]. Many patients with MDD, however, do not experience a sufficient response to initial antidepressant treatments and may develop treatment-resistant depression (TRD) [5]. TRD is most commonly defined as non-response to two or more different pharmacological treatments, taken for an adequate duration and at an adequate dosage [6, 7].

TRD affects approximately one-third of patients with MDD and is associated with functional and physical decline, resulting in diminished health-related quality of life (HRQoL) [3, 8, 9]. Indeed, a considerable proportion of patients living with TRD are reported to be on long-term sick leave or unemployed [10]. The burden of illness is substantially greater, both to the individual and to society, for patients with TRD than it is for patients with MDD who respond to initial treatment [11]. Furthermore, the burden of TRD increases with the duration of the disorder, culminating in rates of hospitalisation for general medical and depression-related causes that are double those reported in patients with treatment-responsive MDD [12, 13]. Even in the absence of treatment resistance, patients who do not achieve remission experience increased risk of relapse and an increased personal burden arising from residual symptoms [14]. Moreover, while long-term remission is the primary goal of antidepressant treatment, the probability of achieving remission after experiencing non-response to two adequate trials of medication decreases with each subsequent treatment, and those who require more treatment steps also demonstrate higher rates of relapse during follow-up [5, 14, 15]. Furthermore, residual symptoms in patients who do not achieve complete remission result in an increased risk of relapse, with lower levels of social and psychological functioning, and greater rates of physical morbidity and mortality [14].

Pharmacological treatment of TRD can employ all approved antidepressant drugs, including selective serotonin- and serotonin and norepinephrine-reuptake inhibitors (SSRIs/SNRIs), tricyclic antidepressants (TCAs), monoamine oxidase inhibitors and other types of antidepressants. Several medications that are

not approved for antidepressant monotherapy in MDD and do not have direct antidepressant activity, such as lithium, thyroid hormone and some atypical antipsychotic drugs, may be used to augment antidepressant treatments [16, 17]. However, advances in the development of specific treatments for TRD have been slow. Currently, in Europe, the only treatment approved specifically for TRD, as it is defined above, is esketamine, an *N*-methyl-*D*-aspartate receptor antagonist, which is administered as a nasal spray in combination with an SSRI/SNRI [18]. In the US, in addition to esketamine, a combination of olanzapine and fluoxetine hydrochloride (Symbyax®) is also approved [19, 20].

Advances in the development of novel treatments for TRD have been slow, with current strategies involving the switching, combining and augmenting of medications approved for the treatment of MDD [10]. Beyond the confines of clinical trials, there is a dearth of evidence assessing the characteristics of, treatment strategies employed for, and outcomes experienced by, patients with TRD in the real world, where populations are more diverse, have more comorbidities and may be less adherent to treatments [16]. Such real-world data are essential to draw a more realistic picture of the treatment landscape, long-term outcomes and the personal burden of disease for patients with TRD. Similarly, it is important to establish how outcomes experienced by patients with TRD are assessed and monitored in real-life clinical practice, and the timescales over which outcomes are reported, in order to support greater comparability of future studies.

The purpose of this systematic literature review (SLR) was therefore to assess real-world evidence in TRD, in order to understand and summarise available evidence regarding current treatment strategies and outcomes. Specifically, the objectives of this SLR were: (1) to further understand the clinical and patient-reported disease burden for patients with TRD in the real-world, and (2) to explore real-world effectiveness of available treatments and the unmet need for better treatment options for these patients. Given the rapid transformation of real-world treatment settings and practices, and in

order to provide data relevant to current real-world practice, this systematic review was restricted to the decade prior to the review date.

METHODS

This article is based on previously conducted studies and does not contain any new studies with human participants or animals performed by any of the authors.

Eligibility Criteria, Selection Process and Outcomes

This SLR was conducted according to a pre-specified protocol. Studies were included if they reported outcomes for adult patients with treatment-resistant MDD [with MDD being diagnosed by the Diagnostic and Statistical Manual of Mental Disorders (DSM) Edition 3 or above, or the International Classification of Diseases (ICD) Edition 9 or above], in which treatment resistance was defined as failure to adequately respond to at least two treatments given at an adequate dose during the same major depressive episode (MDE). Any definition of failure or inadequate response to a treatment as provided by the study authors was deemed eligible for inclusion. Eligible studies included those reporting on both pharmacological and non-pharmacological interventions but excluded non-medical interventions such as traditional Chinese medicine and nutritional supplements. While the search strategy also included studies reporting outcomes related to psychiatric emergencies in patients with MDD (MDD-PE), only articles specific to the TRD population are included within this publication.

Data were collected on prevalence of TRD, treatment types and effectiveness, HRQoL and patient characteristics. Studies reporting the following outcomes for the patient population of interest were included: improvement of depression severity as measured by one of the following validated scales: Montgomery–Åsberg Depression Rating Scale (MADRS), Hamilton Depression Rating Scale (HAM-D), Beck Depression Inventory (BDI), Clinical Global

Impression–Change (CGI-C), Clinical Global Impression–Severity (CGI-S), 9-question Patient Health Questionnaire (PHQ-9); improvement in HRQoL as measured by any valid instrument; and rates of remission and/or response and/or non-remitters and/or non-responders, according to any of the aforementioned validated scales.

Only real-world non-interventional studies were included. This included cohort studies, cross-sectional studies, case–control studies, chart reviews and registry studies, but excluded case reports or case series, due to their high potential for selection bias. Interventional studies, such as randomised controlled trials, were also excluded. The search included studies published between January 2012 and up to and including May 2022, and conference proceedings published between January 2020 and up to and including June 2022. Full eligibility criteria are reported in Table 1.

All titles were reviewed by a single senior reviewer (AB or HL) and 10% of each reviewer's excluded decisions were checked by a second reviewer. All articles included after the title review were reviewed by two independent reviewers at title, abstract and full-text stages. For the reviews at abstract and full-text stage, disagreements were resolved by discussion until a consensus was met. If necessary, a third reviewer made the final decision.

Search Strategy

A search was conducted using the electronic databases in MEDLINE (including MEDLINE In-Process, MEDLINE Daily and MEDLINE Epub Ahead of Print), Embase, The Cochrane Libraries [including Cochrane Database of Systematic Reviews (CDSR) and Cochrane Central Register of Controlled Trials (CENTRAL)] and PsycINFO. MEDLINE, MEDLINE In-Process, MEDLINE Epub Ahead of Print and Embase were searched simultaneously via the Ovid SP platform (11/05/2022). The full list of search terms used for the Ovid SP platform is presented in Supplementary Table S1. CDSR and CENTRAL were searched via The Cochrane Library, via the Wiley Online platform (11/05/2022; Supplementary

Table S2). PsycINFO was searched via the American Psychological Association (APA) website (10/05/2022; Supplementary Table S3).

In addition, the bibliographies of all relevant SLRs identified during the literature review were hand-searched for any additional relevant studies. Furthermore, conference proceedings for 2020 to 2022 from the European College of Neuropsychopharmacology Congress, European Congress of Psychiatry, American Psychiatric Association Annual Meeting, American College of Neuropsychopharmacology Annual Meeting and Psych Congress were also searched.

Data Collection Process and Data Items

Data extractions and quality assessments were performed by a single researcher, with a second researcher independently verifying the extracted information. When necessary, a third individual was enlisted to arbitrate the final decision. The quality of all included studies was assessed using the Alberta Heritage Foundation for Medical Research (AHFMR) tool (Supplementary Figure S1), which was found to be the tool most suited to the heterogeneous nature of the study designs and outcomes collected.

Data were extracted for predefined outcomes. Extracted study characteristics included: the definition of TRD used, patient inclusion and exclusion criteria, total number of patients included and number of patients of relevance (patients with TRD) included, duration of follow-up and investigational treatment type. For baseline participant characteristics, we recorded: age, sex, education level, marital status, employment status, race or ethnicity, disease severity, disease duration, number and type of previous therapeutic interventions that did not result in adequate response and existing treatment within the current MDE. Treatment outcomes were extracted for: change in depression severity and rates of remission and response over time, as measured by MADRS, HAM-D, BDI, CGI-C, CGI-S or PHQ-9 score, as well as change in HRQoL over time, as measured by any validated instrument.

Table 1 Inclusion and exclusion criteria

Study parameter	Inclusion criteria	Exclusion criteria
Patient population	Adults (≥ 18 years) with major depressive disorder (MDD) ^a diagnosed by The Diagnostic and Statistical Manual of Mental Disorders (DSM)-3 or above, or International Classification of Diseases (ICD)-9 or above, who have not adequately responded (any definition of failure or inadequate response as provided by study authors) to at least two treatments given at an adequate dose during the same major depressive episode (MDE)	Studies in children or adolescents (< 18 years) Postnatal depression (as defined by study authors) MDD that is not treatment-resistant (i.e. have not experienced failure or inadequate response to at least two treatments given at an adequate dose during the same MDE) Patients with psychiatric disorders other than treatment resistant depression (TRD)
Study intervention/comparator	Any pharmacological treatment Any non-pharmacological medical treatment, including psychotherapies, electroconvulsive therapy and repetitive transcranial magnetic stimulation Any comparator or no comparator	Supplements, over the counter (OTC) medication or traditional Chinese medicine Non-pharmacological behavioural treatments other than psychotherapies (such as mindfulness or yoga)
Relevant outcomes	Improvement in depression symptoms using one of the following validated scales: Montgomery–Åsberg Depression Rating Scale (MADRS), Hamilton Depression Rating Scale (HAM-D), Beck Depression Inventory (BDI), Clinical Global Impression–Change (CGI-C), Clinical Global Impression–Severity (CGI-S), Patient Health Questionnaire 9-item (PHQ-9) Improvement in health-related quality of life using any validated instrument [e.g. European Quality of Life Group, 5-Dimension Scale (EQ-5D), Sheehan Disability Scale (SDS), Work Productivity and Activity Impairment questionnaire (WPAI)] Rate of remission or response (any definition using a validated scale) Absolute number of non-remitters/non-responders (as opposed to rate; any definition using a validated scale)	Improvement in depression symptoms reported only using measures other than MADRS, HAM-D, BDI, CGI-C, CGI-S or PHQ-9 Outcomes not reported separately for the population of interest

Table 1 continued

Study parameter	Inclusion criteria	Exclusion criteria
Study design	Observational/non-interventional studies, including: Cohort studies, case–control studies, cross-sectional studies and chart reviews	Randomised control trials (RCTs) or non-randomised interventional studies Economic evaluations Non-systematic or narrative reviews Editorials, notes or comments Case reports/case studies In vitro/preclinical/in silico/animal studies Systematic literature reviews (SLRs)/ (network) meta-analysis ([N]MAs)
Language	Abstract or full text in English	Abstract or full text in any other language
Publication date	Studies published within the last 10 years (2012 to 2022) Conference abstracts published within the last 2 years (2020 to 2022)	Studies published prior to 2012 Conference abstracts published prior to 2020

BDI Beck Depression Inventory, *CGI-C* Clinical Global Impression–Change, *CGI-S* Clinical Global Impression–Severity, *DSM* The Diagnostic and Statistical Manual of Mental Disorders, *EQ-5D* European Quality of Life Group, 5-Dimension Scale, *HAM-D* Hamilton Depression Rating Scale, *ICD* International Classification of Diseases, *MADRS* Montgomery–Åsberg Depression Rating Scale, *MDD* major depressive disorder, *MDE* major depressive episode, *PHQ-9* Patient Health Questionnaire 9-item, *(N)MA* (network) meta-analysis, *OTC* over the counter, *RCT* randomised controlled trial, *SDS* Sheehan Disability Scale, *SLR* systematic literature review, *TRD* treatment resistant depression, *WPAI* Work Productivity and Activity-Impairment questionnaire

^aTreatment resistance is defined as failure to adequately respond to at least two treatments given at an adequate dose, during the same episode

RESULTS

Included Studies

A total of 8,030 records were identified through database searches, with a further 5,296 identified through supplementary searches. Following title, abstract and full-text review, 22 publications were included in the SLR (Fig. 1). The 22 publications reported on 20 unique studies, including 13 prospective cohort studies, 5 retrospective cohort studies, 1 chart review and 1 case–control study.

The key characteristics of the included studies reporting on the TRD population are detailed in Table 2. Non-pharmacological

treatments consisted of transcranial magnetic stimulation (TMS; $n = 7$) [21–27], and electroconvulsive therapy (ECT; $n = 1$) [28]. Studies of specific pharmacological treatments comprised ketamine and/or esketamine ($n = 3$) [29–31], onabotulinum toxin ($n = 1$) [32], valproate ($n = 1$) [33], pramipexole ($n = 1$) [34] and tranylcypromine and amitriptyline ($n = 1$) [35]. Other studies employed combinations of multiple pharmacological and/or non-pharmacological treatments ($n = 5$) [16, 36–40]. The majority of included studies (13/20) were prospective cohort investigations and reported a wide range of follow-up durations (range: 2 weeks to 9.4 years). The number of patients of relevance included in each study ranged from 14 to 411 (Table 2). Just over half of the

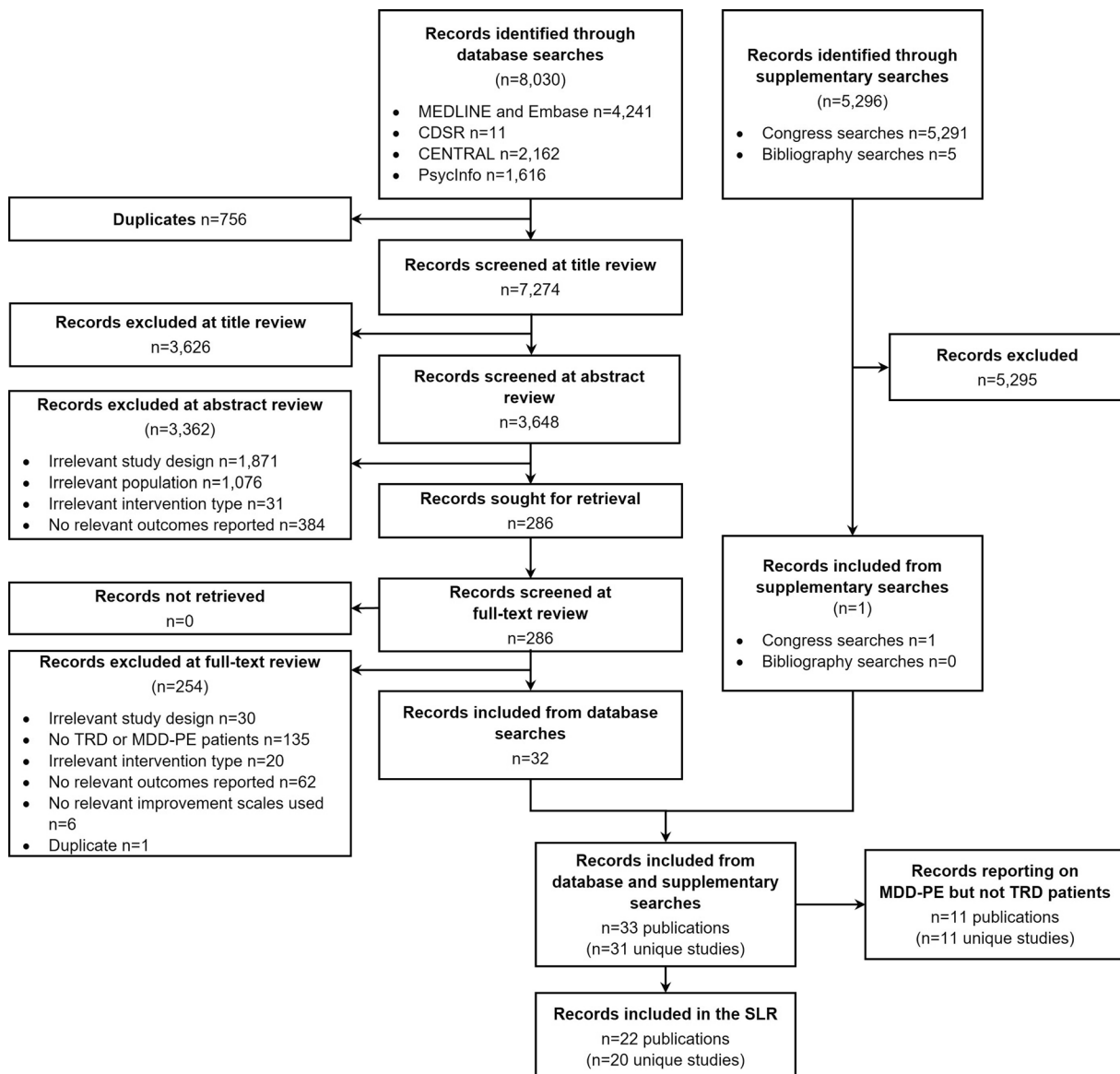


Fig. 1 Flowchart of studies included and excluded in the systematic review process. *CDSR* Cochrane Database of Systematic Reviews, *CENTRAL* Cochrane Central

Register of Controlled Trials, *MDD-PE* major depressive disorder-psychiatric emergency, *TRD* treatment resistant depression, *SLR* systematic literature review

included studies (12/20) had less than 50 patients with TRD, with only three studies including more than 100 patients.

Of the 19 studies reporting the number of sites, most were undertaken at a single site (14/20), with the remaining five being multicentre studies. The majority of studies were conducted in North America (8/20), Europe (5/20) or Asia (5/20), with only one study reporting data from

multiple countries [16]. The most common countries of study location were Canada (4/20), the United States (4/20), India (3/20) and Italy (3/20).

Definition of TRD

There was considerable variation in the definition of TRD used within the inclusion criteria of

Table 2 Characteristics of included studies in the TRD patient population

Author(s), Year	Study design	MDD diagnosis tool (s)	Definition of TRD used	No. patients of relevance to review	Summary of current treatment (s)	Duration of follow- up	Baseline depression severity rating tool (s)	Health- related quality of life rating tool (s)
Furrado et al. (2013) [21]	Prospective cohort	DSM Fourth Edition, Text Revision (DSM- 4-TR)	Non-responsiveness to at least two courses of antidepressant (AD) medications for at least 6 weeks (Stage II, Thase and Rush definition)	29	Transcranial magnetic stimulation (TMS)	Mean 12.45 ± 2.66 weeks	HAM-D- 17, BDI- II	–
Feffer et al. (2017) [22]	Case–control	DSM-4-TR	Could not tolerate or failed to respond to at least two AD trials of adequate dose and duration during the current episode	32	TMS	4–6 weeks	HAM-D- 21	–
Oliveira- Maia et al. (2017) [23]	Prospective cohort	DSM-4	Failure of at least three trials of adequate psychopharmacology treatment	51	TMS	2 weeks	HAM-D- 17, BDI	–
Feffer et al. (2018) [24]	Retrospective cohort	DSM Fifth Edition (DSM-5)	Resistance or intolerance to at least two adequate medication trials in the current episode	42	TMS	3 weeks	–	–
Verma et al. (2018) [25]	Retrospective cohort	ICD-10	Failing to respond to at least two AD medication trials given for at least 6–8 weeks at adequate dose	22	TMS	3 weeks	MADRS, HAM-D	–
Fukuda et al. (2020) [26]	Prospective cohort	DSM-5	Insurance coverage of TMS for TRD, defined by a documented history of nonresponse to multiple pharmacotherapy trials and a course of psychotherapy in the current depressive episode	15	TMS	9 weeks	PHQ-9	–
Toth et al. (2022) [27]	Prospective cohort	DSM-5	At least two AD medication trials with demonstrated inadequacy or intolerance	21	TMS	6 weeks	PHQ-9	–

Table 2 continued

Author(s), Year	Study design	MDD diagnosis tool (s)	Definition of TRD used	No. patients of relevance to review	Summary of current treatment (s)	Duration of follow- up	Baseline depression severity rating tool (s)	Health- related quality of life rating tool (s)
Phillips et al. (2012) [36]	Prospective cohort	DSM-4	Failure to remit following treatment with at least two ADs at adequate dosage for at least 6 weeks each (determined by retrospective chart review)	27	Various pharmacological treatments	Approximately 1 year	HAM-D- 17	–
Gobbi et al. (2018) [37]	Prospective cohort	DSM-5	Detailed history of lack of response to at least two adequate AD mono or combination trials (at therapeutic dosage and for at least three weeks) operationally defined using the Antidepressant Treatment History Form	86	Various pharmacological treatments	3 months	MADRS, HAM- D-17, CGI-S	–
Heerlein et al. (2021) [16, 38]	Prospective cohort	DSM-5 or ICD-10	Failing to respond to two or more different oral ADs ($\leq 25\%$ improvement on best day), given at an adequate dose and for a sufficiently long period (≥ 6 weeks) in the same MDE, as determined using the Massachusetts General Hospital- Antidepressant Treatment Response Questionnaire (MGH-ATRQ)	411	Various pharmacological, psychotherapeutic and neuromodulation treatments	12 months	MADRS	SDS, EQ- 5D-5L, EuroQol- visual analogue scales (EQ- VAS)
Kavita et al. (2021) [39]	Prospective cohort	DSM-4	Depression may be considered resistant to treatment when at least two trials with ADs from different pharmacologic classes (adequate in dose, duration, and compliance) fail to produce a significant clinical improvement	52	Various pharmacological treatments	16 weeks	HAM-D	–
Perugi et al. (2021) [40]	Prospective cohort	DSM-5 or ICD-10	Treatment failure on at least two different oral ADs, given at adequate dose and duration, as evaluated by the MGH- ATRQ	124	Various pharmacological treatments	12 months	MADRS, CGI-S	EQ-5D-5L utility, EQ-VAS

Table 2 continued

Author(s), Year	Study design	MDD diagnosis tool (s)	Definition of TRD used	No. patients of relevance to review	Summary of current treatment (s)	Duration of follow- up	Baseline depression severity rating tool (s)	Health- related quality of life rating tool (s)
Slupski et al. (2020) [29]	Prospective cohort	DSM-5	Insufficient response to at least two trials of treatment	35	Ketamine	4 weeks	–	–
Nguyen et al. (2022) [30]	Chart review	DSM-5	History of at least four AD monotherapy trials and two augmentation strategies	110	Ketamine/ esketamine	8 weeks	PHQ-9	–
Samalin et al. (2022) [31]	Prospective cohort	DSM-5	Non-responsive after two or more different ADs of two different classes given at adequate dose and duration during the current depressive episode	66	Esketamine	6 months	MADRS	–
Tokusu et al. (2013) [28]	Retrospective cohort	ICD-10	Failure to respond to adequate trials of two distinctly different classes of ADs (stage II Thase and Rush definition)	42	Electroconvulsive therapy (ECT)	1 year	HAM-D- 17	–
Ghabrash et al. (2016) [33]	Prospective cohort	DSM-5	Detailed past history of lack of response to at least two adequate AD combination trials defined using the Antidepressant History Form	14	Valproate	7 months	MADRS, CGI-S	–

Table 2 continued

Author(s), Year	Study design	MDD diagnosis tool (s)	Definition of TRD used	No. patients of relevance to review	Summary of current treatment (s)	Duration of follow- up	Baseline depression severity rating tool (s)	Health- related quality of life rating tool (s)
Chugh et al. (2018) [32]	Prospective cohort	ICD-10, major depressive disorder, single episode, severe without psychotic features (F32.2) or major depressive disorder, recurrent severe without psychotic features (F33.2)	Treated consecutively with at least two ADs at a sufficient dosage (according to the prescribing information for the respective drug) and duration (at least 4 to 6 weeks), if the respective drug was tolerated, but no clinical improvement and remaining severe depression	42	Onabotulinumtoxin A (Allergan) + current treatment	3 weeks	MADRS, HAM- D-17, BDI	–
Ferreira- Garcia et al. (2018) [35]	Retrospective cohort	DSM Third Edition, Revision (DSM-3-R), DSM-4-TR	Resistance to at least four adequate AD pharmacological trials (an effective dose of an AD, including potentiation strategies, for at least 6 weeks), including one trial of a tricyclic antidepressant (TCA) and one trial of tranlycypromine	31	Tranlycypromine + amitriptyline	Mean 9.37 ± 3.86 years	HAM-D, CGI-S	–
Tundo et al. (2022) [34]	Retrospective cohort	DSM-5	Failure of at least two trials with two ADs of separate classes	79	Pramipexole + current treatment	24 weeks	HAM-D- 21	–

AD antidepressant, *BDI* Beck Depression Inventory, *CGI-S* Clinical Global Impression-Severity, *DSM-3-R* The Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revision, *DSM-4-TR* The Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision, *DSM-5* The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, *ECT* electroconvulsive therapy, *EQ-5D-5L* European Quality of Life Group, 5-Dimension 5-Level Scale, *EQ-VAS* EuroQol-visual analogue scales, *F32.2* major depressive disorder, single episode, severe without psychotic features, *F33.2* major depressive disorder, recurrent severe without psychotic features, *HAM-D* Hamilton Depression Rating Scale, *ICD* International Classification of Diseases, *MADRS* Montgomery–Åsberg Depression Rating Scale, *MDD* major depressive disorder, *MDE* major depressive episode, *MGH-ATRQ* Massachusetts General Hospital–Antidepressant Treatment Response Questionnaire, *PHQ-9* Patient Health Questionnaire 9-item, *SDS* Sheehan Disability Scale, *TCA* tricyclic antidepressant, *TMS* transcranial magnetic stimulation, *TRD* treatment resistant depression

studies (Table 2). The minimum number of previously failed treatments for classification as treatment-resistant ranged from at least two (as per the study inclusion criteria for this SLR) to at least four. Only some studies (7/20) specified the necessary minimum duration of treatment administration considered to be adequate; this ranged from 'at least 3 weeks' to 'at least 6 to 8 weeks'. Only one study explicitly reported a quantitative value to define inadequate improvement with treatment ($\leq 25\%$ improvement on best day Massachusetts General Hospital–Antidepressant Treatment Response Questionnaire [MGH-ATRQ] score) [16]. Other studies employed different levels of responsiveness, with variations in the language used to define treatment failure, including 'failure to remit', 'insufficient response' and 'demonstrated inadequacy'.

Baseline Characteristics

The level of descriptive characteristic data captured at baseline varied between studies (Table 3). Included patients with TRD were typically middle-aged (range of mean age: 41.2 to 64.5 years) and overall were approximately balanced for sex (range of percentage female: 27.3 to 66.1%).

Disease severity was reported by 18/20 studies, with the HAM-D scale being used most frequently ($n = 11$). Other rating scales used included MADRS ($n = 7$), CGI-S ($n = 4$), BDI ($n = 3$) and the PHQ-9 ($n = 3$). All studies that included a measure of disease severity at baseline reported mean scores that could be classified as either moderate, moderate to severe, or severe, according to previously defined thresholds for the HAM-D [41], MADRS [42], BDI [43] and PHQ-9 [44] scales (Fig. 2). Mean disease duration was reported by 9/20 studies, ranging from 1.4 to 16.5 years in the overall cohorts of included studies, with 6/9 of these studies reporting mean durations > 10 years. The duration of the current depressive episode was reported by 6/20 studies, with mean durations ranging from 0.8 to 12.5 years. Three of the included studies reported the mean number of previously failed treatments during the current

depressive episode, ranging from 2.9 to 5.9 in the main study cohorts, with one study reporting a mean of 6.4 previously failed treatments in a subset of patients who did not respond to TMS treatment [21].

Clinical Outcomes

Responsiveness to treatment was assessed most frequently by a version of the HAM-D scale ($n = 10$), followed by MADRS ($n = 9$), CGI-S ($n = 4$), BDI ($n = 3$), CGI-C ($n = 2$) and PHQ-9 ($n = 2$). Studies typically reported follow-up data across a relatively short period, with 11/20 studies featuring a follow-up period of 12 weeks or less and only 5/20 studies reporting follow-up data over a period of 12 months or more. Of studies reporting HAM-D scores, 7/10 studies reported the number of patients experiencing at least one level of response (remission, response, partial response or non-response), 8/10 reported absolute scores and 6/10 reported either an absolute or relative change from baseline. Of studies reporting MADRS scores, 6/9 studies reported the number of patients experiencing at least one level of response, 7/9 reported absolute scores and 4/9 reported either an absolute or relative change from baseline.

For studies reporting absolute or relative changes in a depression severity rating scale, a summary of the mean change in depression severity score from baseline to the final pre-specified timepoint is presented in Table 4. Studies reporting absolute mean HAM-D (8/20), MADRS (7/20) or BDI (2/20) scores over time are presented alongside previously described thresholds for severity classification and remission in Fig. 3 [41–43, 45, 46].

In studies reporting the proportion of patients achieving remission and/or response, no single treatment type exhibited a marked pattern of higher rates of treatment responsiveness. However, remission and response rates broadly increased over time.

In the acute setting (≤ 8 weeks), explicitly reported remission rates using HAM-D (2/20) or MADRS (1/20) scores ranged from 0 to 18% in the overall study populations, while rates of response without remission (only reported

Table 3 Baseline characteristics of patients of relevance in included studies of TRD

Author(s), Year	Summary of current treatment(s)	Age, mean (SD), years	No. patients of relevance	% female (patients of relevance)	Disease severity, mean (SD)	Disease duration, mean (SD), years	Age at diagnosis, mean (SD), years	Duration of current episode, mean (SD), years	Number of previously failed treatments, mean (SD)
Furtado et al. (2013) [21]	TMS	44.0 (10.4)	29	–	HAM-D-17: 25.34 (4.29) BDI-II: 38.86 (9.46)	16.0 (11.5)	30.9 (13.0)	5.6 (10.6)	5.9 (3.5)
Feffer et al. (2017) [22]	TMS	47.6 (16.5)	32	–	HAM-D-21: 20.97 (6.43)	–	30.8 (17.5)	–	–
Oliveira- Maia et al. (2017) [23]	TMS	46.5 (12.4)	51	57	HAM-D-17: 31.3 (7.5) BDI-II: 29.9 (9.8)	–	–	–	–
Feffer et al. (2018) [24]	TMS	41.6 (12.9)	42	57	–	–	–	–	–
Verma et al. (2018) [25]	TMS	46.1 (15.2)	22	27	HAM-D: 21.79 (5.36) MADRS: 38.25 (8.53) PHQ-9: 20.07 (5.27)	–	–	–	–
Fukuda et al. (2020) [26]	TMS	50.7 (15.9)	15	–	–	–	–	–	–
Toth et al. (2022) [27]	TMS	42.4 (16.6) ^a , 52.6 (12.6) ^b	11 ^a , 10 ^b	–	PHQ-9: 17.4 (6.4) ^a , 18.2 (5.3) ^b	–	–	–	–

Table 3 continued

Author(s), Year	Summary of current treatment(s)	Age, mean (SD), years	No. patients of relevance	% female (patients of relevance)	Disease severity, mean (SD)	Disease duration, mean (SD), years	Age at diagnosis, mean (SD), years	Duration of current episode, mean (SD), years	Number of previously failed treatments, mean (SD)
Phillips et al. (2012) [36]	Various pharmacological treatments	47.5 (10.6) ^c 44.3 (10.2) ^d	12 ^c , 15 ^d	–	HAM-D-17: 22.5 (4.2) ^c , 24.9 (4.7) ^d	–	31.3 (14.2) ^c , 29.7 (14.0) ^d	–	–
Gobbi et al. (2018) [37]	Various pharmacological treatments	51.4 (14.3) ^c , 48.9 (12.6) ^f	36 ^c , 50 ^f	–	MADRS: 27.5 (1.4) ^{cg} , 33.6 (1.2) ^{fg} HAM-D-17: 20.4 (1.0) ^{cg} , 25.6 (0.9) ^{fg} CGI-S: 4.5 (0.2) ^{cg} , 5.4 (0.2) ^{fg}	–	–	10.9 (9.5) ^c , 12.5 (11.6) ^f	2.9 (1.8) ^c , 4.0 (2.7) ^f
Heerlein et al. (2021) [16, 38]	Various pharmacological, psychotherapeutic and neurostimulatory treatments	51.0 (10.6)	411	62	MADRS: 31.8 (6.0)	13.7 (11.2)	37.2 (13.1)	2.6 (3.9)	–
Kavita et al. (2021) [39]	Various pharmacological treatments	41.2 (10.3)	52	54	HAM-D: 29.67 (1.43)	1.4 (1.0)	–	–	–
Perugi et al. (2021) [40]	Various pharmacological treatments	53.2 (9.8)	124	66	MADRS: 31.1 (6.1)	16.0 (11.1)	37.3 (13.1)	1.9 (2.8)	–
Slupski et al. (2020) [29]	Ketamine	–	35	–	–	–	–	–	–

Table 3 continued

Author(s), Year	Summary of current treatment(s)	Age, mean (SD), years	No. patients of relevance	% female (patients of relevance)	Disease severity, mean (SD)	Disease duration, mean (SD), years	Age at diagnosis, mean (SD), years	Duration of current episode, mean (SD), years	Number of previously failed treatments, mean (SD)
Nguyen et al. (2022) [30]	Ketamine/ esketamine	44 (12.3) ^h , 50 (12.7) ⁱ	40 ^h , 70 ⁱ	–	PHQ-9: 18.6 (4.9) ^h , 18.9 (4.6) ⁱ	–	–	–	–
Samalin et al. (2022) [31]	Esketamine	52.5 (16.7)	66	62	MADRS: 31.2 (5.4)	16.5 (14.4)	–	3.5 (2.5)	–
Tokursu et al. (2013) [28]	ECT	61.3 (13.6)	42	62	HAM-D-17: 21.9 (5.9)	7.4 (8.9)	53.3 (16.3)	0.8 (0.9)	–
Ghabrash et al. (2016) [33]	Valproate	42.0 (12.9)	14	–	MADRS: 33.9 (1.5) ^g CGI-S: 5.1 (0.3) ^g	–	–	–	–
Chugh et al. (2018) [32]	Onabotulinumtoxin A (Allergan) + current treatment	44.6 (11.2)	42	45	HAM-D-17: 32.8 (5.6) BDI: 46.4 (5.5) MADRS: 47.7 (4.4)	5.5 (8.2)	–	–	–
Ferreira- Garcia et al. (2018) [35]	Tranylcypromine + amitriptyline	44.7 (10.2)	31	–	HAM-D: 32.57 (3.64) CGI-S: 6.29 (0.46)	15.0 (8.5)	–	–	–

Table 3 continued

Author(s), Year	Summary of current treatment(s)	Age, mean (SD), years	No. patients of relevance	% female (patients of relevance)	Disease severity, mean (SD)	Disease duration, mean (SD), years	Age at diagnosis, mean (SD), years	Duration of current episode, mean (SD), years	Number of previously failed treatments, mean (SD)
Tundo et al. (2022) [34]	Pramipexole + current treatment	64.5 (13.2)	79	58	HAM-D-21: 18 ^j	15.0 ^j	47.3 (18.6)	–	4 ^j

BDI Beck Depression Inventory, *CGI-S* Clinical Global Impression-Severity, *ECT* electroconvulsive therapy, *HAM-D* Hamilton Depression Rating Scale, *MADRS* Montgomery–Åsberg Depression Rating Scale, *PHQ-9* Patient Health Questionnaire 9-item, *SD* standard deviation, *THC* tetrahydrocannabinol, *TMS* transcranial magnetic stimulation, *TRD* treatment resistant depression

^aPatients who did not achieve response to TMS treatment
^bPatients who achieved response to TMS treatment ($\geq 50\%$ decrease in PHQ-9 score)
^cPatients who had sustained remission (MADRS total score ≤ 12) for 6 months
^dPatients who had not experienced remission after 12 months
^eAntidepressant therapy group
^fSecond-generation antipsychotics plus antidepressants group
^gStandard error of the mean reported
^hPatients with tetrahydrocannabinol (THC)-positive urine
ⁱPatients with THC-negative urine
^jNeither standard deviation or standard error of the mean reported

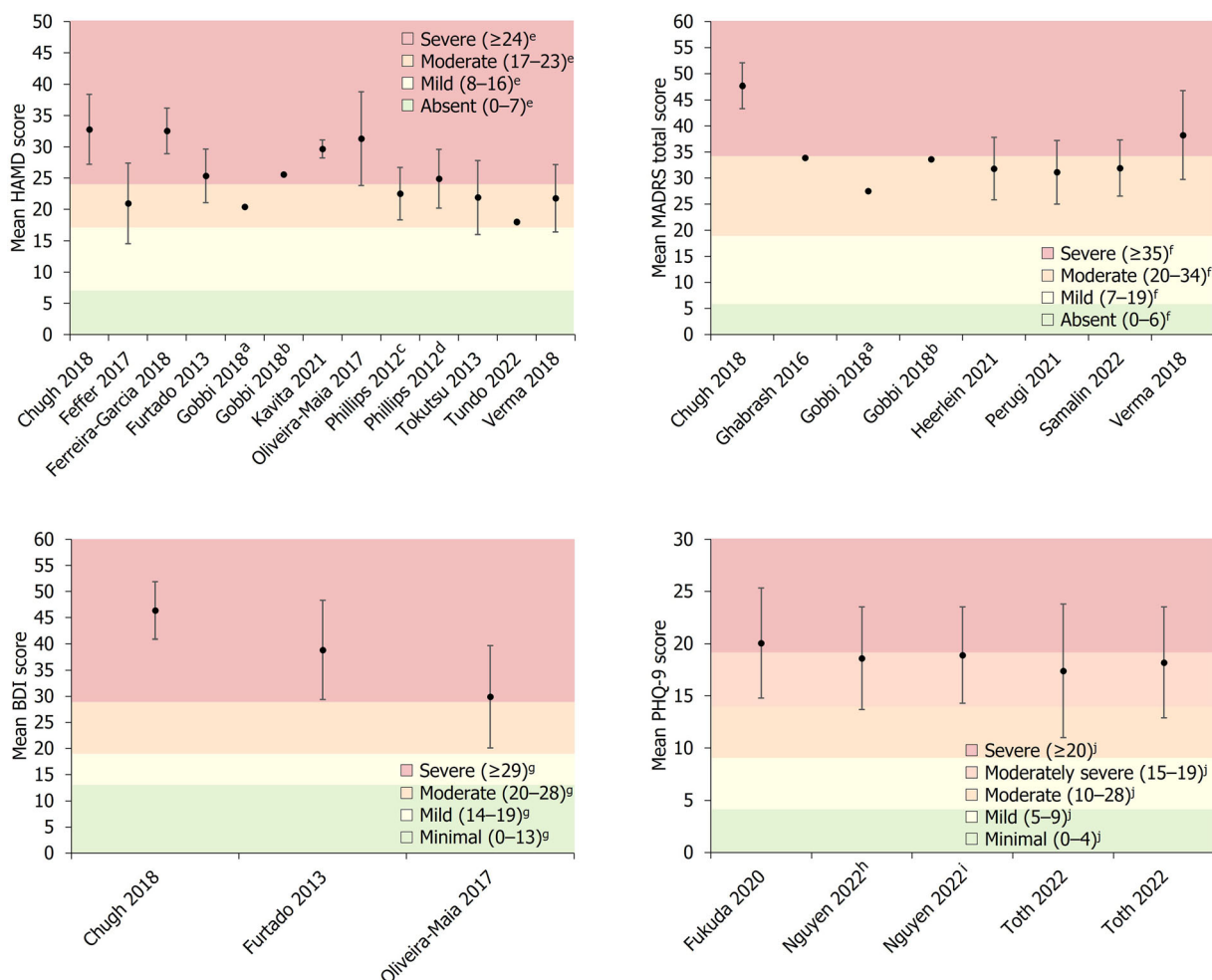


Fig. 2 Baseline depression severity scores. Data are presented as mean $\pm$ SD (where available). ^aGroup receiving antidepressant treatment. ^bGroup receiving second-generation antipsychotic plus antidepressant treatment. ^cPatients who subsequently experienced remission. ^dPatients who did not subsequently achieve remission. ^eSeverity classification for HAM-D score defined by Zimmerman et al. [41]. ^fSeverity classification for MADRS score defined by Muller et al. [42]. ^gSeverity classification for

BDI score defined by Beck et al. [43]. ^hPatients who were THC-positive. ⁱPatients who were THC-negative. ^jSeverity classification for PHQ-9 score defined by Kroenke et al. [44]. *BDI* Beck Depression Inventory, *HAM-D* Hamilton Depression Rating Scale, *MADRS* Montgomery–Åsberg Depression Rating Scale, *PHQ-9* 9-question Patient Health Questionnaire, *SD* standard deviation, *THC* Tetrahydrocannabinol

using HAM-D data) ranged from 0 to 57.1%. Of studies explicitly reporting medium-term remission rates (> 8 weeks to ≤ 6 months) using MADRS (2/20) or HAM-D (1/20) scores, rates of remission and response were generally higher than those reported in the acute setting, ranging from 16.7 to 70.9% in the overall study populations. In these studies, rates of response

without remission ranged from 9.8 to 80.6%. Long-term (≥ 12 months) rates of remission and/or response were reported by 2/20 studies using MADRS and 1/20 studies using HAM-D. Long-term remission rates ranged from 19.2 to 54.8% in the overall study populations, while rates of response without remission ranged from 11.6 to 15.9% (only reported using MADRS data).

Table 4 Change in depression severity score from baseline to final prespecified timepoint

	Summary of current treatment(s)	<i>n</i> (with baseline score)	Baseline score (mean)	Final prespecified timepoint	<i>n</i> (final prespecified timepoint)	Final prespecified timepoint score (mean)	Change from baseline (%)	Change from baseline (absolute)	Remission at final prespecified timepoint (<i>n</i> [%])	Response at final prespecified timepoint without remission at final prespecified timepoint (<i>n</i> [%])	Partial response at final prespecified timepoint (<i>n</i> [%])
HAM-D											
Furtado (2013) [21]	TMS	29	25.34	6 weeks	29	13.8	–	–	–	–	–
Feffer (2017) [22]	TMS	32	20.97	4–6 weeks	32	14.44	–	–	6 (18)	9 (28)	–
Oliveira-Maia (2017) [23]	TMS	43	31.3	2 weeks	41	–	–22.7	–	–	0 (0)	–
Verma (2018) [25]	TMS	14	21.79	3 weeks	14	10.64	–49.74	–	–	NR (57.1)	–
Gobbi (2018 ^a) [37]	Various pharmacological treatments	36	20.4	3 months	36	12.9	–	7.5	–	–	–
Gobbi (2018 ^b) [37]	Various pharmacological treatments	50	25.6	3 months	50	15.1	–	10.5	–	–	–
Kavita (2021) [39]	Various pharmacological treatments	52	29.67	16 weeks	52	16.26	–	–	–	–	–
Tokutsu (2013) [28]	ECT	42	21.9	1 year	–	–	–	–	23 (54.8)	13 (NR)	–

Table 4 continued

	Summary of current treatment(s)	<i>n</i> (with baseline score)	Baseline score (mean)	Final prespecified timepoint	<i>n</i> (final prespecified timepoint)	Final prespecified timepoint score (mean)	Change from baseline (%)	Change from baseline (absolute)	Remission at final prespecified timepoint (<i>n</i> [%])	Response at final prespecified timepoint (<i>n</i> [%])	Response without remission at final prespecified timepoint (<i>n</i> [%])	Partial response at final prespecified timepoint (<i>n</i> [%])
Chugh (2018) [32]	Onabotulinumtoxin A (Allergan) + current treatment	42	32.8	3 weeks	42	23.9	− 27	−9.0	0 (0)	−	−	−
Ferreira-Garcia (2018) [35]	Tranylcypromine + amitriptyline	31	32.57	12 weeks	31	11.8	− 63.7	−	−	−	25 (80.6)	−
Tundo (2022) [34]	Pramipexole + current treatment	79	18	24 weeks	79	2	−	−	56 (70.9)	60 (75.9)	−	−
MADRS												
Verma (2018) [25]	TMS	8	38.25	3 weeks	8	25.5	− 33.6	−	−	NR (37.5)	−	−
Gobbi (2018 ^a) [37]	Various pharmacological treatments	36	27.5	3 months	36	16.7	−	10.8	−	−	−	−
Gobbi (2018 ^b) [37]	Various pharmacological treatments	50	33.6	3 months	50	19.0	−	14.6	−	−	−	−
Heerlein (2021) [16]	Various pharmacological, psychotherapeutic and neuromodulation treatments	411	31.8	12 months	146	20	−	−	28 (19.2)	−	17 (11.6)	−

Table 4 continued

	Summary of current treatment(s)	<i>n</i> (with baseline score)	Baseline score (mean)	Final prespecified timepoint	<i>n</i> (final prespecified timepoint)	Final prespecified timepoint score (mean)	Change from baseline (%)	Change from baseline (absolute)	Remission at final prespecified timepoint (<i>n</i> [%])	Response at final prespecified timepoint (<i>n</i> [%])	Partial response at final prespecified timepoint (<i>n</i> [%])
Perugi (2021) [40]	Various pharmacological treatments	124	31.1	6 months	89	21.5	–	–	16 (18)	–	9 (10.1)
Slupski (2020) [29]	Ketamine	(NR)	–	4 weeks	35	–	–	–	5 (NR)	8 (NR)	–
Samalin (2022) [31]	Esketamine	50	31.2	4 weeks	25	19.5	–	–	–	22 (47.8)	–
Ghabrash (2016) [33]	Valproate	14	33.9	7 months	14	13.6	–	–	–	–	–
Chugh (2018) [32]	Onabotulinumtoxin A (Allergan) + current treatment	42	47.7	3 weeks	42	35	– 27	–12.7	0 (0)	–	–
CGIS											
Gobbi (2018 ^a) [37]	Various pharmacological treatments	36	4.5	3 months	36	3	–	1.5	–	–	–
Gobbi (2018 ^b) [37]	Various pharmacological treatments	50	5.4	3 months	50	3.4	–	2.0	–	–	–
Ghabrash (2016) [33]	Valproate	14	5.1	7 months	14	2.6	–	–	–	–	–

Table 4 continued

	Summary of current treatment(s)	<i>n</i> (with baseline score)	Baseline score (mean)	Final prespecified timepoint	Final prespecified timepoint	Final prespecified score (mean)	Change from baseline (%)	Change from baseline (absolute)	Remission at final prespecified timepoint (<i>n</i> [%])	Response at final prespecified timepoint (<i>n</i> [%])	Response without remission at final prespecified timepoint (<i>n</i> [%])	Partial response at final prespecified timepoint (<i>n</i> [%])
Ferreira-Garcia (2018) [35]	Tranylcypromine + amitriptyline	31	6.29	12 weeks	31	2.39	–	–	–	–	–	–
Perugi (2021) [40]	Various pharmacological treatments	NR	–	6 months	89	–	–	–	–	–	–	–
BDI												
Furtado (2013) [21]	TMS	29	38.86	6 weeks	29	23.41	–	–	–	–	–	–
Oliveira-Maia (2017) [23]	TMS	46	29.9	2 weeks	43	–	– 18.2	–	–	–	–	–
Chugh (2018) [32]	Onabotulinumtoxin A (Allergan) + current treatment	42	46.4	3 weeks	42	33.9	– 27	– 12.5	–	–	–	–
PHQ-9												
Fukuda (2020) [26]	TMS	15	20.07	9 weeks	15	–	– 65.7	–	–	–	–	–

Table 4 continued

Summary of current treatment(s)	<i>n</i> (with baseline score)	Baseline score (mean)	Final prespecified timepoint	<i>n</i> (final prespecified timepoint)	Final prespecified timepoint score (mean)	Change from baseline (%)	Change from baseline (absolute)	Remission at final prespecified timepoint (<i>n</i> [%])	Response at final prespecified timepoint (<i>n</i> [%])	Response without remission at final prespecified timepoint (<i>n</i> [%])	Partial response at final prespecified timepoint (<i>n</i> [%])
Totd (2022 ^c) [27]	11	17.4	6 weeks	11	–	– 13.7	–	–	–	–	–
Totd (2022 ^d) [27]	10	18.2	6 weeks	10	–	– 73.4	–	–	–	–	–

BDI Beck Depression Inventory, *CGI-I-S* Clinical Global Impression-Severity, *ECT* electroconvulsive therapy, *HAM-D* Hamilton Depression Rating Scale, *MADRS* Montgomery–Åsberg Depression Rating Scale, *NR* not reported, *PHQ-9* Patient Health Questionnaire 9-item, *TMS* transcranial magnetic stimulation

^aAntidepressant therapy group
^bSecond-generation antipsychotics plus antidepressants group
^cPatients who did not achieve response to TMS treatment
^dPatients who achieved response to TMS treatment ($\geq 50\%$ decrease in PHQ-9 score)

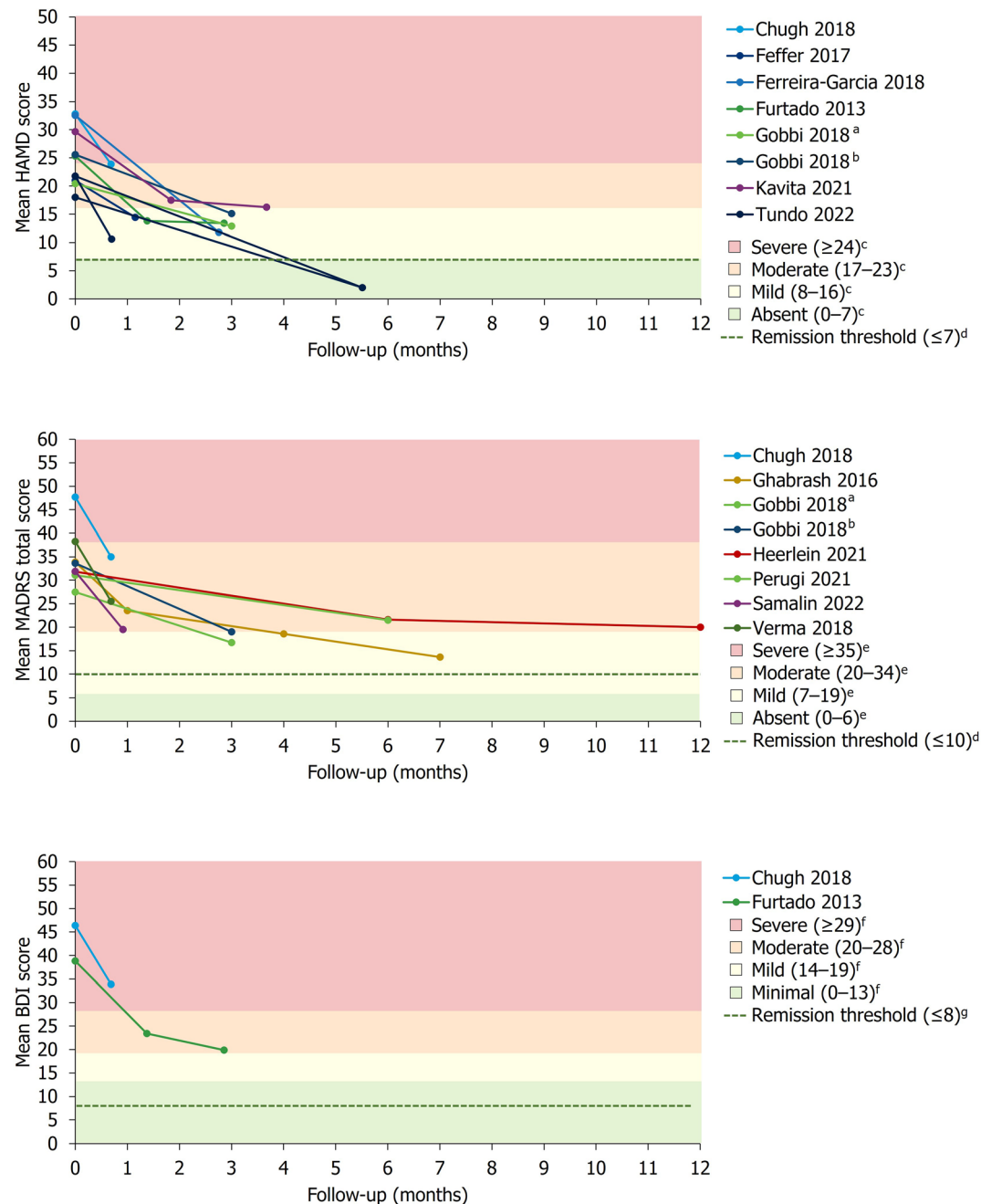


Fig. 3 Change in mean absolute depression severity scores over time. ^aGroup receiving antidepressant treatment. ^bGroup receiving second-generation antipsychotic plus antidepressant treatment. ^cSeverity classification for HAM-D score defined by Zimmerman et al. [41]. ^dRemission thresholds for HAM-D and MADRS

previously defined by Zimmerman et al. [45]. ^eSeverity classification for MADRS score defined by Muller et al. [42]. ^fSeverity classification for BDI score defined by Beck et al. [43]. ^gRemission threshold for BDI previously defined by Riedel et al. [46]

Health-Related Quality of Life Outcomes

Two studies, Heerlein et al. and Perugi et al., reported HRQoL data, the latter reporting on an Italian subset of patients in the study of the former [16, 38, 40]. These studies assessed HRQoL at baseline and after 6 months using the European Quality of Life Group, 5-Dimension 5-Level Scale (EQ-5D-5L), whereby an index score of 1 represents perfect health, 0 represents a health state equivalent to death and <0 represents a state worse than death [38]. Heerlein et al. reported a mean baseline EQ-5D-5L index of 0.41 in 397 patients [38]. In a separate publication reporting on the same study, after 6 months of receiving various treatments, the EQ-5D-5L score had increased by 0.11 in patients who did not respond to treatment, by 0.26 in those who experienced response without remission and by 0.34 in those who experienced remission [16]. After 12 months, the improvements from baseline were 0.11, 0.31 and 0.35, respectively [16]. Perugi et al. similarly reported a mean baseline EQ-5D-5L index of 0.4 in 121 patients [40]. After 6-months of receiving various treatments, the EQ-5D-5L index improved to 0.6 ($n = 85$) in patients remaining in the study, but was lower in those who did not respond to treatment (0.2; $n = 61$) versus those who responded (0.7; $n = 8$) or reached remission (0.9; $n = 16$).

Heerlein et al. also reported the impact of TRD on functioning [Sheehan Disability Scale (SDS)] and work productivity (WPAI). At baseline, according to the SDS, 61.6% of these patients experienced marked or extreme work impairment (mean SDS total score: 22.4), with WPAI scores revealing overall mean impairment of work and activity to be 60.5% and 73.3%, respectively. After 6 and 12 months of treatment, mean change from baseline in total SDS score was -2.67 and -2.91 , respectively, in those who did not respond to treatment, -7.58 and -7.00 in those with response without remission and -12.53 and -14.44 in those who experienced remission. Change in WPAI scores were not reported.

Quality Assessment of Included Studies

The quality of the included studies, as indicated by the AHFMR quality assessment checklist, was moderate to good (Supplementary Fig. 1). Of the included studies, the description of subjects and settings was generally appropriate, with just two studies providing only a partially adequate description. While only three studies included an appropriate sample size for the study design and target population, all studies provided an adequate description of the statistical analysis methods employed. There was a consistent lack of adjustment for confounding, which was either not done or not reported in almost all (19/20) studies.

DISCUSSION

This systematic review has identified 20 real-world studies, comprising a variety of pharmacological and non-pharmacological treatments, reporting baseline characteristics and clinical outcomes in patients with TRD. There was substantial heterogeneity in the definition of TRD and the means of assessment, and manner of reporting, on the burden of illness and treatment outcomes, preventing the quantitative synthesis of results. Nevertheless, patients with TRD consistently presented with moderate to severe depression, long durations of illness and poor HRQoL. Only two studies assessed the latter, suggesting that greater emphasis is placed on clinical outcomes than patient-centred outcomes in real-world studies. Treatment outcomes varied greatly. While many patients typically experienced a level of response by the end of the included studies' follow-up period, rates of remission were generally low. Studies predominantly involved relatively small sample sizes, followed-up over relatively short durations, highlighting the need for larger-scale, longer-term studies.

In their criteria for TRD, studies did not consistently define what constitutes response failure, nor adequate improvement, with the latter ranging from remission to response. Despite the definition of TRD being centred around the number of prior treatment failures

and the well-established negative relationship between the number of prior treatment failures and the probability of relapse from acute response over time, very few studies reported the absolute number of prior treatments received [5]. The tools used to assess disease severity were varied, with most studies only reporting outcomes using a single tool. As the clinical tools developed to assess TRD focus on different elements of the disease and use a range of assessment methods, a more complete picture of the disease and its burden could be developed by consistently using multiple tools within individual studies. Similarly, the sample size and follow-up duration of the included studies was wide ranging, but studies typically featured relatively small sample sizes monitored over a period of several weeks to a few months, potentially reducing the robustness of the findings. Of the included studies, the most frequently utilised treatment was TMS, while several studies also reported on patients receiving multiple treatments, often comprising two or more different treatment types. Collectively, the heterogeneity of the included studies suggests a lack of clinical consensus and standardisation in the severity classification and monitoring of TRD in real-world practice [11].

Applying previously defined cut-offs, the severity of depression at baseline, which has been identified as the most important prognostic factor for TRD [47], ranged from moderate to severe. Disease duration at baseline was typically greater than 10 years, with the current MDE frequently spanning several years. These findings are consistent with an earlier review of the burden of TRD, which reported that, on average, patients with TRD had MDD durations of 4.4 years and had completed 4.7 unsuccessful drug treatments [11]. Taken together, the common concurrence of MDD that spans many years, with severe and prolonged MDEs, is indicative of the substantial and often unremitting burden imposed by TRD, which exceeds that of MDD alone [3].

Definitions of remission and response varied between studies. In those studies reporting rates of responsiveness to treatment in the acute setting, remission rates were generally low. Given the established propensity among

patients with TRD for severe and long-lasting MDD and MDEs, it follows that longer-term interventions and follow-up periods are likely to be required for many patients to achieve remission. Indeed, studies of medium- and long-term follow-up durations typically reported higher rates of treatment responsiveness than those of acute interventions, but were nevertheless highly variable, with many patients not reaching remission after 12 months of treatment. Heerlein et al. reported that, after 6 months of initiating a new antidepressant treatment, only 16.7% of patients with TRD achieved remission, rising only to 19.2% at 12 months [16]. Similarly, Perugi et al., reporting on a subset of the previous study, demonstrated that, after 6 months of receiving various treatments, only 18% of patients had experienced remission, rising only marginally to 22.7% at 12 months [40]. This is in agreement with an earlier review of studies of patients with TRD, which captured data from nearly 60,000 patients and reported wide-ranging rates of remission and response, averaging 20% and 36%, respectively [11]. Indicative of the severe burden and unmet need for effective treatments experienced by patients with TRD, the aforementioned study also reported a 17% prevalence of prior suicide attempts in this population. Collectively, these studies demonstrate that existing treatment strategies are often insufficiently effective to enable patients with TRD to experience remission, particularly over short time durations of treatment.

Health-related quality of life was rarely assessed within the included studies, suggesting that, in real-world practice, greater emphasis is still placed on clinical outcomes than patient-centred outcomes. However, it should also be considered that there are few disease-specific tools for the assessment of HRQoL in patients with MDD, which may contribute to the infrequency of HRQoL evaluation. Nevertheless, it has been previously demonstrated that, when compared with MDD patients without treatment resistance, patients with TRD experience significantly lower HRQoL and a greater impairment of work activity and productivity [3]. Both studies that reported HRQoL data demonstrated low HRQoL scores in the TRD

population [16, 38, 40]. In these studies, patients with TRD exhibited mean baseline EQ-5D-5L index scores (0.4 and 0.41) that were substantially lower than the general adult population, which typically ranges between ~ 0.70 and 0.95 globally [38, 40, 48]. Patients with TRD were also likely to be experiencing significant work and activity impairment, culminating in high levels of absenteeism and presenteeism [10]. In these two studies, patients who did not achieve remission during the study period experienced further declines or minimal improvement in HRQoL. Importantly, although remission rates were low in these studies, those patients with TRD who did experience remission also experienced substantial increases in HRQoL, emphasising the merit of remission as the primary treatment goal [16, 40]. Future studies would benefit from increased assessment and continued monitoring of HRQoL to better capture the personal burden of TRD and to assess the efficacy of treatment options on functioning and productivity. This, alongside consistent use of a range of clinical outcome assessment tools, could build a more comprehensive picture of the TRD population and specific unmet needs that patients experience in their day-to-day lives.

This SLR was conducted in accordance with best practice guidelines, such as the use of two independent systematic reviewers to review abstracts and full-text articles against the inclusion and eligibility criteria [49]. Nevertheless, conclusions drawn from this systematic review are naturally limited by the information available in, and the methodological quality of, the included published literature. While the present study's evaluation of real-world evidence enables greater confidence in the relevance and applicability of findings to the patient population, the lack of standardisation prevented quantitative synthesis of the findings. Owing to the inclusion of only real-world evidence studies, and the heterogeneity of the interventions and study designs employed, safety outcomes were not captured. This decision was made since, outside randomised controlled trials, safety outcomes are inconsistently monitored and reported, thus limiting comparability. Nevertheless, it must be considered

that treatment side effects are likely to influence subsequent treatment decisions, clinical outcomes and patient-centred outcomes, such as HRQoL and functioning. As such, it may be of value to investigate real-world safety patterns of therapies used for TRD in future studies.

CONCLUSIONS

This review demonstrates that there is a paucity of studies investigating real-world treatment of patients with TRD. Those that do are heterogeneous in their definition, assessment and monitoring of TRD and feature a wide range of treatment types and durations. The lack of evidence, together with the heterogeneity of studies that are available, make drawing specific conclusions about this patient population in the real world challenging. However, more broadly, studies found in this review show that patients with TRD had typically experienced long-lasting MDD, with moderate or severe MDEs spanning multiple years. Rates of response to treatment varied greatly between studies, but remission rates were typically low. Few studies had investigated long-term treatment outcomes in this patient population, for whom response to treatment is notoriously elusive. Longer durations of study intervention and follow-up were associated with marginally greater gains in favourable treatment outcomes, but remission rates typically remained low even after a year of treatment. Health-related quality of life was seldom measured, suggesting that greater emphasis is typically placed on the reporting of clinical outcomes over patient outcomes. When it was measured, HRQoL was reported to be particularly low in patients with TRD. Thus, while there is a lack of clinical consensus on the definition of TRD, the condition doubtless carries a high burden of illness and there exists an unmet need for more effective treatment options. Furthermore, future real-world studies would benefit from the application of standardised modalities of assessment, monitoring and reporting of treatment effectiveness, including greater consideration of HRQoL outcomes, to better understand

the burden on patients affected by this condition.

Medical Writing/Editorial Assistance The authors also acknowledge Andrew Wilhelmsen, PhD, and Jessica Patel, PhD, from Costello Medical, UK, for medical writing and editorial assistance based on the authors' input and direction. Third-party writing assistance was funded by Janssen EMEA.

Authors' Contributions. Substantial contributions to study conception and design: Albino J. Oliveira-Maia, Ania Bobrowska, Eric Constant, Tetsuro Ito, Yerkebulan Kambarov, Hannah Luedke, Siobhán Mulhern-Haughey, Christian von Holt; literature review and data extraction: Ania Bobrowska, Hannah Luedke; substantial contributions to analysis and interpretation of the data: Albino J. Oliveira-Maia, Ania Bobrowska, Eric Constant, Tetsuro Ito, Yerkebulan Kambarov, Hannah Luedke, Siobhán Mulhern-Haughey, Christian von Holt; drafting the article or revising it critically for important intellectual content: Albino J Oliveira-Maia, Ania Bobrowska, Eric Constant, Tetsuro Ito, Yerkebulan Kambarov, Hannah Luedke, Siobhán Mulhern-Haughey, Christian von Holt; final approval of the version of the article to be published: Albino J Oliveira-Maia, Ania Bobrowska, Eric Constant, Tetsuro Ito, Yerkebulan Kambarov, Hannah Luedke, Siobhán Mulhern-Haughey, Christian von Holt.

Funding. This study was sponsored Janssen EMEA. Support for third-party writing assistance for this article, provided by Andrew Wilhelmsen, PhD, and Jessica Patel, PhD, from Costello Medical, UK, was funded Janssen EMEA in accordance with Good Publication Practice (GPP3) guidelines (<http://www.ismpps.org/gpp3>). Rapid Service and Open Access Fees were also funded by Janssen EMEA.

Data Availability. Data available from the corresponding author upon reasonable request.

Declarations

Conflicts of Interest. Albino J. Oliveira-Maia: Received grants from Schuhfried GmbH,

Janssen and Compass Pathways, Ltd; received payment, honoraria, travel support or advisory board fees from MSD, Janssen, Angelini, Neurolite AG, and the European Monitoring Centre for Drugs and Drug Addiction; investigator-driven research funded by Fundação para Ciência e Tecnologia (PTDC/SAU-NUT/3507/2021; PTDC/MED-NEU/1552/2021; PTDC/MED-NEU/31331/2017), Fundação para Ciência e Tecnologia and FEDER (PTDC/MED-NEU/30845/2017_LISBOA-01-0145-FEDER-030845; FCT-PTDC/MEC-PSQ/30302/2017-IC&DTLISBOA-01-0145-FEDER), the European Commission Horizon 2020 program (H2020-SC1-2017-CNECT-2-777167-BOUNCE; H2020-SC1-DTH-2019-875358-FAITH), the European Research Council (grant ERC-2020-STG-Grant agreement 950357) and the European Joint Programme in Rare Diseases (Joint Translational Call 2019) through Fundação para Ciência e Tecnologia (EJPRD/0001/2020); Vice-President of the Portuguese Society for Psychiatry and Mental Health; Head of the Psychiatry Working Group for the National Board of Medical Examination (GPNA) at the Portuguese Medical Association and Portuguese Ministry of Health. Eric Constant: Received honoraria, advisory board fees and travel support from AstraZeneca, BMS, Lilly, Wyeth, Servier, Janssen Cilag, Pfizer, Lundbeck, Viatrix and Teva. Tetsuro Ito, Yerkebulan Kambarov, Siobhán Mulhern-Haughey, Christian von Holt: Employees of Janssen; hold Johnson & Johnson company stocks/stock options. Ania Bobrowska, Hannah Luedke: Employees of Costello Medical, UK.

Ethical Approval. This article is based on previously conducted studies and does not contain any new studies with human participants or animals performed by any of the authors.

Open Access. This article is licensed under a Creative Commons Attribution-Non-Commercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative

Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

1. Salahudeen MS, Wright CM, Peterson GM. Esketamine: new hope for the treatment of treatment-resistant depression? A narrative review. *Ther Adv Drug Saf*. 2020;11:2042098620937899.
2. Fava M, Kendler KS. Major Depressive Disorder. *Neuron*. 2000;28(2):335–41.
3. Jaffe DH, Rive B, Denoe TR. The humanistic and economic burden of treatment-resistant depression in Europe: a cross-sectional study. *BMC Psychiatry*. 2019;19(1):247.
4. National Institute for Health and Care Excellence. Depression in adults: treatment and management (NG222) 2022. <https://www.nice.org.uk/guidance/ng222/resources/depression-in-adults-treatment-and-management-pdf-66143832307909>. Accessed 7 Dec 2022.
5. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry*. 2006;163(11):1905–17.
6. Souery D, Amsterdam J, de Montigny C, Lecrubier Y, Montgomery S, Lipp O, et al. Treatment resistant depression: methodological overview and operational criteria. *Eur Neuropsychopharmacol*. 1999;9(1–2):83–91.
7. European Medicines Agency. Guideline on clinical investigation of medicinal products in the treatment of depression. EMA/CHMP/185423/2010 Rev 2 2013. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-investigation-medicinal-products-treatment-depression_en.pdf. Accessed 18 Aug 2022.
8. Al-Harbi KS. Treatment-resistant depression: therapeutic trends, challenges, and future directions. *Patient Prefer Adherence*. 2012;6:369–88.
9. Thase ME. Treatment-resistant depression: prevalence, risk factors, and treatment strategies. *J Clin Psychiatry*. 2011;72(5): e18.
10. Heerlein K, Young AH, Otte C, Frodl T, Degraeve G, Hagedoorn W, Oliveira-Maia AJ, Perez Sola V, Rathod S, Rosso G. Real-world evidence from a European cohort study of patients with treatment resistant depression: Baseline patient characteristics. *J Affect Disord*. 2021;283:115–22.
11. Mrazek DA, Hornberger JC, Altar CA, Degtiar I. A review of the clinical, economic, and societal burden of treatment-resistant depression: 1996–2013. *Psychiatr Serv (Washington, DC)*. 2014;65(8): 977–87.
12. Halaris A, Sohl E, Whitham EA. Treatment-resistant depression revisited: a glimmer of hope. *J Pers Med*. 2021;11(2).
13. Crown WH, Finkelstein S, Berndt ER, Ling D, Poret AW, Rush AJ, et al. The impact of treatment-resistant depression on health care utilization and costs. *J Clin Psychiatry*. 2002;63(11):963–71.
14. Mendlewicz J. Towards achieving remission in the treatment of depression. *Dialogues Clin Neurosci*. 2008;10(4):371–5.
15. Gaynes BN, Warden D, Trivedi MH, Wisniewski SR, Fava M, Rush JA. What did STAR*D teach us? Results from a large-scale, practical, clinical trial for patients with depression. *Psychiatr Services*. 2009;60(11):1439–45.
16. Heerlein K, Perugi G, Otte C, Frodl T, Degraeve G, Hagedoorn W, et al. Real-world evidence from a European cohort study of patients with treatment resistant depression: treatment patterns and clinical outcomes. *J Affect Disord*. 2021;290:334–44.
17. Ionescu DF, Rosenbaum JF, Alpert JE. Pharmacological approaches to the challenge of treatment-resistant depression. *Dialogues Clin Neurosci*. 2015;17(2):111–26.
18. European Medicines Agency. Esketamine Nasal Spray Summary of Product Characteristics 2021. <https://www.ema.europa.eu/en/medicines/human/EPAR/spravato>. Accessed 18 Aug 2022.
19. Food and Drug Administration. Spravato Prescribing Information 2019. https://www.accessdata.fda.gov/drugsatfda_docs/label/2019/211243lbl.pdf. Accessed 18 Aug 2022.

20. Food and Drug Administration. Symbyax Prescribing Information 2009. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2009/021520s022lbl.pdf. Accessed 7 Dec 2022.
21. Furtado CP, Hoy KE, Maller JJ, Savage G, Daskalakis ZJ, Fitzgerald PB. An investigation of medial temporal lobe changes and cognition following antidepressant response: a prospective rTMS study. *Brain Stimul.* 2013;6(3):346–54.
22. Feffer K, Lapidus KAB, Braw Y, Bloch Y, Kron S, Netzer R, et al. Factors associated with response after deep transcranial magnetic stimulation in a real-world clinical setting: results from the first 40 cases of treatment-resistant depression. *Eur Psychiatry.* 2017;44:61–7.
23. Oliveira-Maia AJ, Press D, Pascual-Leone A. Modulation of motor cortex excitability predicts antidepressant response to prefrontal cortex repetitive transcranial magnetic stimulation. *Brain Stimul.* 2017;10(4):787–94.
24. Feffer K, Fettes P, Giacobbe P, Daskalakis ZJ, Blumberger DM, Downar J. 1 Hz rTMS of the right orbitofrontal cortex for major depression: safety, tolerability and clinical outcomes. *Eur Neuropsychopharmacol.* 2018;28(1):109–17.
25. Verma R, Kumar N, Kumar S. Effectiveness of adjunctive repetitive transcranial magnetic stimulation in management of treatment-resistant depression: a retrospective analysis. *Indian J Psychiatry.* 2018;60(3):329–33.
26. Fukuda AM, Hindley LE, Kang JWD, Tirrell E, Tyrka AR, Ayala A, et al. Peripheral vascular endothelial growth factor changes after transcranial magnetic stimulation in treatment-resistant depression. *NeuroReport.* 2020;31(16):1121–7.
27. Toth C, King Johnson ML, Heinzerling A, Trapp N. Response to TMS treatment for depression associated with higher levels of psychological well-being. *J Psychiatr Res.* 2022;150:142–6.
28. Tokutsu Y, Umene-Nakano W, Shinkai T, Yoshimura R, Okamoto T, Katsuki A, et al. Follow-up study on electroconvulsive therapy in treatment-resistant depressed patients after remission: a chart review. *Clin Psychopharmacol Neurosci.* 2013;11(1):34–8.
29. Slupski J, Cubala WJ, Gorska N, Slupska A, Galuszko-Wegielnik M. Copper concentrations in ketamine therapy for treatment-resistant depression. *Brain Sci.* 2020;10(12):1–9.
30. Nguyen J, Boerth J, Lacro J. Influence of cannabis on effectiveness of esketamine and ketamine for treatment resistant depression. In: College of Psychiatric and Neurologic Pharmacists (CPNP) Annual Meeting. 2022.
31. Samalin L, Rotharmel M, Mekaoui L, Gaudre-Wattinne E, Codet MA, Bouju S, et al. Esketamine nasal spray in patients with treatment-resistant depression: the real-world experience in the French cohort early-access programme. *Int J Psychiatry Clin Pract.* 2022;26(4):352–62.
32. Chugh S, Chhabria A, Jung S, Kruger THC, Wollmer MA. Botulinum toxin as a treatment for depression in a real-world setting. *J Psychiatr Pract.* 2018;24(1):15–20.
33. Ghabrash MF, Comai S, Tabaka J, Saint-Laurent M, Booi L, Gobbi G. Valproate augmentation in a subgroup of patients with treatment-resistant unipolar depression. *World J Biol Psychiatry.* 2016;17(2):165–70.
34. Tundo A, Betro S, Iommi M, de Filippis R. Efficacy and safety of 24-week pramipexole augmentation in patients with treatment resistant depression. A retrospective cohort study. *Prog Neuropsychopharmacol Biol Psychiatry.* 2022;112:110425.
35. Ferreira-Garcia R, Da Rocha Freire RC, Appolinario JC, Levitan MN, Halkjaer-Lassen RD, Bueno JR, et al. Tranylcypromine plus amitriptyline for electroconvulsive therapy-resistant depression. *J Clin Psychopharmacol.* 2018;38(5):502–4.
36. Phillips JL, Batten LA, Aldosary F, Tremblay P, Blier P. Brain-volume increase with sustained remission in patients with treatment-resistant unipolar depression. *J Clin Psychopharmacol.* 2012;73(5):625–31.
37. Gobbi G, Ghabrash MF, Nunez N, Tabaka J, Di Sante J, Saint-Laurent M, et al. Antidepressant combination versus antidepressants plus second-generation antipsychotic augmentation in treatment-resistant unipolar depression. *Int Clin Psychopharmacol.* 2018;33(1):34–43.
38. Heerlein K, Young AH, Otte C, Frodl T, Degraeve G, Hagedoorn W, et al. Real-world evidence from a European cohort study of patients with treatment resistant depression: baseline patient characteristics. *J Affect Disord.* 2021;283:115–22.
39. Kavita, Sahu MK, Dhone PG, Tiwari RK, Hishikar R. A prospective observational study on the effects of antidepressant treatment on hypothalamic-pituitary-adrenal axis regulation in treatment-resistant depression. *J Clin Diagn Res.* 2021;15(4):FC4–7.
40. Perugi G, Calo P, De Filippis S, Rosso G, Vita A, Adami M, et al. Clinical features and outcomes of 124 Italian patients with treatment resistant

- depression: a real-world, prospective study. *Front Psychiatry*. 2021;12: 769693.
41. Zimmerman M, Martinez JH, Young D, Chelminski I, Dalrymple K. Severity classification on the Hamilton Depression Rating Scale. *J Affect Disord*. 2013;150(2):384–8.
 42. Müller MJ, Szegedi A, Wetzel H, Benkert O. Moderate and severe depression. Gradations for the Montgomery-Asberg Depression Rating Scale. *J Affect Disord*. 2000;60(2):137–40.
 43. Beck AT, Steer RA, Brown GK. BDI-II, Beck Depression Inventory: manual: Psychological Corp; 1996.
 44. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606–13.
 45. Zimmerman M, Posternak MA, Chelminski I. Derivation of a definition of remission on the Montgomery-Asberg Depression Rating Scale corresponding to the definition of remission on the Hamilton Rating Scale for depression. *J Psychiatr Res*. 2004;38(6):577–82.
 46. Riedel M, Möller HJ, Obermeier M, Schennach-Wolff R, Bauer M, Adli M, et al. Response and remission criteria in major depression—a validation of current practice. *J Psychiatr Res*. 2010;44(15): 1063–8.
 47. Lundberg J, Cars T, Lööv S-Å, Söderling J, Sundström J, Tiihonen J, et al. Association of treatment-resistant depression with patient outcomes and health care resource utilization in a population-wide study. *JAMA Psychiat*. 2022;80(2):167–75.
 48. Janssen B, Szende A. Population norms for the EQ-5D. In: Szende A, Janssen B, Cabases J, editors. Self-reported population health: an international perspective based on EQ-5D. Dordrecht: Springer; 2014. p. 19–30.
 49. Lefebvre C GJ, Briscoe S, Featherstone R, Littlewood A, Marshall C, Metzendorf M-I, Noel-Storr A, Paynter R, Rader T, Thomas J, Wieland LS. Chapter 4: Searching for and selecting studies. 2022. In: *Cochrane handbook for systematic reviews of interventions* [Internet]. Cochrane. Available from: www.training.cochrane.org/handbook. Accessed 2 June 2023.