The assessment of resistance to antidepressant treatment: Rationale for the Antidepressant Treatment History Form: Short Form (ATHF-SF)

Harold A. Sackeim, Scott T. Aaronson, Mark T. Bunker, Charles R. Conway, Mark A. Demitrack, Mark S. George, Joan Prudic, Michael E. Thase, A. John Rush

1. Introduction

In recent years there has been intense interest in treatment-resistant depression (TRD) (Berlim and Turecki, 2007b; Ruhe et al., 2012). Earlier estimates suggested that two-thirds of patients in a major depressive episode (MDE) have substantial improvement following their first antidepressant medication trial (Klein et al., 1980; Souery et al., 1999). The Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study (Rush et al., 2004) challenged this perspective. In a large sample without a history of adequate antidepressant treatment failure in the current episode, approximately 30% remitted following treatment with citalopram. Non-remitters received up to three additional trials of antidepressant treatment (Rush et al., 2006b). The likelihood of acute benefit decreased with each subsequent trial, and, if remission was obtained, the likelihood of relapse increased. For example, in STAR*D the probability of both remitting and sustaining that remission for a year was less than 5% in patients receiving their third antidepressant treatment (Conway et al., 2017). This prospective study and similar data (Fife et al., 2017; Mahlich et al., 2018; Rizvi et al., 2014; Saveanu et al., 2015; Thomas et al., 2013) have led to the estimate that approximately one-third of patients in a MDE are characterized by TRD (Berlim and Turecki, 2007b; Cepeda et al., 2018; Fava, 2003; Thase, 2011). TRD is associated with increased morbidity and mortality (Banankhah et al., 2015; Reutfors et al., 2018; Souery et al., 2007), increased medical and psychiatric health care costs (Amos et al., 2018; Kubitz et al., 2013; Lepine et al., 2012; Mahlich et al.,...
The assessment of resistance to antidepressant treatment impacts on clinical decision making, the quality of research, and public health policy (Gaynes et al., 2018; Ruhe et al., 2012). Regulatory agencies like the FDA have increasingly labelled medications and medical devices as indicated for different levels of TRD, typically defined in terms of the number of “failed adequate” trials (Center for Drug Evaluation and Research (CDER), 2018). Recommendation of new therapeutic regimens and prognostic information conveyed to patients by treatment providers are strongly influenced by their assessment of prior treatment history. In research, inclusion/exclusion criteria now commonly stipulate the range of treatment resistance that characterizes samples. The reliable assessment of TRD is also fundamental to investigation of its neurobiology, phenomenological correlates, treatment options, and prognostic significance.

Various approaches have been taken to define, categorize, and/or quantify degree of treatment resistance (see Table 1) (Berlim and Turecki, 2007a, b; Hazari et al., 2013; Ruhe et al., 2012; Trevino et al., 2014). Most approaches are categorical. For instance, in line with the staged model originally proposed by Thase and Rush (1995, 1997), the European Medicines Agency stated that “a patient has been considered suffering from TRD when consecutive treatment with two products of different pharmacological classes, used for a sufficient length of time at an adequate dose, fail to induce a clinically meaningful effect (inadequate response)” (European Medicines Agency, 2013). However, TRD definitions vary in whether they require “failed” trials from 1 or more classes of pharmacological treatment, the number of such failed trials, the definition of what constitutes a failed trial, and whether they include brain stimulation interventions, such as electroconvulsive therapy (ECT) or transcranial magnetic stimulation (TMS), or psychotherapies as potential adequate failed trials (Berlim and Turecki, 2007a; Demyttenaere and Van Duppen, 2019; Gaynes et al., 2018; Ruhe et al., 2012). Some staging models prioritize particular treatment strategies. For example, the Massachusetts General Hospital (MGH) (Fava, 2003) and Maudsley (Fekadu et al., 2009) staging models give extra weight if patients had not benefited sufficiently from adequate pharmacological augmentation trials, as well as ECT. Some approaches evaluate dimensions that extend beyond the patient’s treatment history. In addition to the number of failed adequate antidepressant trials, the Maudsley model weights baseline illness severity and duration of illness in producing an overall score of the degree of treatment resistance (Fekadu et al., 2009; van Belkum et al., 2018). A Dutch extension of the Maudsley staging model also includes degree of functional impairment, co-morbid anxiety, and psychosocial stressors in quantifying TRD (Peeters et al., 2016).

Undoubtedly, heterogeneity in the definitions and metrics used to identify and quantify TRD impedes progress (Berlim and Turecki, 2007a; Demyttenaere and Van Duppen, 2019; Gaynes et al., 2018; Hazari et al., 2013; Ruhe et al., 2012). However, at their core, each approach to defining and measuring TRD requires assessment of the whether individual antidepressant treatment trials were adequately administered and resulted in insufficient clinical benefit. The models differ principally in how many “failed adequate” trials they require for classification, their stipulations regarding the pharmacological classes of these trials, and the extra weight, if any, given particular treatment strategies (pharmacological augmentation, ECT).

### 2. Definitions and models of treatment-resistant depression

<table>
<thead>
<tr>
<th>Antidepressant Treatment History Form (ATHF) (Sackeim et al., 2001; Sackeim, 2001)</th>
<th>Yes</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
<th>Yes</th>
<th>No</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>European Staging Model (European Medicines Agency, 2013; Souery et al., 1990)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Massachusetts General Hospital Staging Model (MGH) (Fava, 2003) and the Antidepressant Treatment Response Questionnaire (ATRQ) (Desseilles et al., 2011)</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Maudsley Staging Model Fekadu et al., 2009; van Belkum et al. (2018)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Conway et al. Staging Model (Conway et al., 2017)</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Essential features of assessment instruments and models of treatment-resistant depression.

The standardization of psychiatric diagnosis through the use of structured interviews and objective criteria was a major advance. Far less attention, however, has been paid to standardizing the assessment

2018; Olsson et al., 2018; Russell et al., 2004), and markedly reduced quality of life (Johnston et al., 2018; Mrazek et al., 2014).
of antidepressant treatment history despite its importance in treatment selection and prognosis, and its public health implications (Philip et al., 2010; Ruhe et al., 2012; Thase and Rush, 1995; Wijeratne and Sachdev, 2008). The ATHF (Sackeim, 2001) has been perhaps the most widely used instrument to systematically assess antidepressant treatment trials and characterize treatment resistance. This instrument provides explicit criteria for evaluating the adequacy of treatment with pharmacological and brain stimulation interventions deemed to be effective in MDE. Of instruments with a similar purpose (e.g., the MGH Antidepressant Treatment Response Questionnaire [ATRQ]) (Desseilles et al., 2011), the ATHF is unique in providing separate criteria for unipolar and bipolar MDE, as well as psychotic and non-psychotic MDE (Hazarzi et al., 2013).

The ATHF uses a 5-point scale to rate each antidepressant trial (see Table 2). Scores of 1 or 2 apply to “inadequate” trials, in which the treatment was given in insufficient dose or duration. Scores of 3 and above indicate different degrees of treatment resistance. For pharmacological agents, trials with scores of 4 or 5 denote either use of a medication above the established minimally effective dose (score of 3) or the use of specific augmentation strategies (e.g., an antidepressant medication plus lithium). The ATHF also requires judgment regarding adherence and clinical outcome. Patients are not considered as resistant to a treatment if they did not receive it due to non-adherence. Likewise, the concept of resistance or a “failed adequate trial” requires that patients manifest insufficient clinical improvement on the specific regimen adhered to at sufficient dose and duration. To receive a score of 3 or higher, the ATHF requires that the clinical improvement in the trial be below the level associated with response. Response is usually defined as a 50% reduction in MDE symptom severity, and corresponds to a judgment that substantial clinical improvement was obtained (Frank et al., 1991; Rush et al., 2006a).

The ATHF produces summary scores for the total number of antidepressant trials attempted, the number of “failed adequate trials” (ratings of 3 or higher), the rating of the most potent trial (1–5), and the total potency across all trials. Since the ATHF scores each trial independently, and also groups pharmacological treatments into specific classes (e.g., TCA, MAOI, etc.), the ATHF can be used to derive any of the various TRD definitions and metrics that have been proposed.

The ATHF was first developed in studies of ECT, where retrospective assessment of the degree of medication resistance was found in some, but not all, studies to prospectively predict both short-term ECT outcomes (Dombrovski et al., 2005; Heijnen et al., 2011; Prudic et al., 1996; Prudic et al., 1996; Rasmussen et al., 2007; Sackeim et al., 2009; Sachem et al., 2000; Sachem et al., 2008; van den Broek et al., 2004) and likelihood of postECT relapse (Prudic et al., 2013; Rasmussen et al., 2009; Sackeim et al., 1990, 2001a). The ATHF was then applied in a variety of studies with other brain stimulation interventions, both to specify minimal and maximal levels of treatment resistance in samples and as a predictor of outcome (Aaronson et al., 2017; George et al., 2010; Lisanby et al., 2009; O'Reardon et al., 2007; Rush et al., 2005a; Sackeim et al., 2001c). With the same purposes of sample definition and outcome prediction, the ATHF was also applied in psychopharmacological studies (Blumberger et al., 2011; Hsu et al., 2016; Joel et al., 2014; Joo et al., 2005; Kocsis et al., 2008), and in relation to the adequacy of treatment surrounding sentinel events, such as suicide (Oquendo et al., 1999, 2002), or the adequacy of care in particular subgroups (Baca-Garcia et al., 2009; Dew et al., 2005), such as psychotic depression (Andreescu et al., 2007; Mulsant et al., 1997). The first version of the ATHF showed strong inter-rater reliability (Prudic et al., 1990; Sackeim et al., 1990), as did a subsequent version with computerized scoring (Oquendo et al., 2003). Summary scores on the ATHF have shown predictive validity in multiple prospective studies of treatment outcome and long-term course following a variety of types of pharmacological and brain stimulation interventions.

4. Antidepressant Treatment History Form: Short Form (ATHF-SF)

The last published revision of the ATHF criteria was in 2001 (Sackeim, 2001). In the interim there have been substantial changes in the psychopharmacological, brain stimulation, and psychotherapy interventions considered effective in the treatment of MDE. New medications and strategies are now in common practice, such as the use of atypical antipsychotic medication in bipolar MDE or combination treatment with an antidepressant and an atypical antipsychotic medication for unipolar MDE. During this period, two neuromodulation treatments, TMS and VNS, were approved by the FDA with specific labelling for use in TRD. Many investigators using the ATHF made local modifications to account for this information.

The need for updating this instrument led to the creation of a workgroup tasked with producing a complete revision. The authors of this article constituted this workgroup, and are the authors of the new instrument, the Antidepressant Treatment History Form: Short Form (ATHF-SF). The group was chaired by Dr. Harold A. Sackeim, and the work of the group was partly supported by LivaNova, Inc, who intends to use the ATHF-SF in studies of VNS in TRD (see Disclosures). The original ATHF is now considered the long-form of the instrument (Sackeim, 2001).

In addition to this article, this workgroup produced four other documents: the ATHF-SF Instruction Manual, ATHF-SF Data Collection Forms for the Current Episode and as a Composite of Prior Episodes, and the ATHF-SF Scoring Checklist. The manual provides detailed instructions for collecting relevant data from a variety of sources and

<table>
<thead>
<tr>
<th>Table 2</th>
<th>ATHF rating criteria for antidepressant treatments.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating</td>
<td>Criteria</td>
</tr>
<tr>
<td>0</td>
<td>No treatment or medication with known psychotropic action</td>
</tr>
<tr>
<td>1</td>
<td>Any medication &lt; 4 weeks or &lt; minimum adequate daily dose (or blood level)⁹ ECT: 1–3 treatments</td>
</tr>
<tr>
<td>2</td>
<td>Any medication ≥ 4 weeks at less than the minimum adequate daily dose or blood level⁸ ECT: 4–6 treatments</td>
</tr>
<tr>
<td>3</td>
<td>Any medication ≥ 4 weeks and higher than minimum adequate daily dose or blood level⁹ For psychotic MDE: combination with an antipsychotic (≥ 400 CPZ equivalents) for ≥ 3 weeks ECT: 7–9 unilateral treatments</td>
</tr>
<tr>
<td>4</td>
<td>Any medication ≥ 4 weeks at higher dose or blood level or any medication at level 3 augmented with lithium ≥ 2 weeks⁸ For psychotic MDE: combination with an antipsychotic (≥ 400 CPZ equivalents) for ≥ 3 weeks ECT: 10–12 unilateral treatments; 7–9 bilateral treatments</td>
</tr>
<tr>
<td>5</td>
<td>Any medication at level 4 augmented with lithium ≥ 2 weeks⁸ For psychotic MDE combination with an antipsychotic (≥ 400 CPZ equivalents) for ≥ 3 weeks ECT: ≥ 13 unilateral treatments; ≥ 10 bilateral treatments</td>
</tr>
</tbody>
</table>

ATHF = Antidepressant Treatment History Form; Minimum adequate daily dosages are provided for all medications that can be rated “2” or higher. A higher dose range is also provided for ratings of “4” or “5.”

---

⁸ ≥ 14000 mg of MCPP daily or 550 mg of MBDP daily for 2 weeks⁹ ≥ 400 CPZ equivalents for 2 weeks
applying the ATHF-SF trial-by-trial rating criteria. Many applications of the ATHF have only evaluated antidepressant trials in the current episode, with the view that information from earlier episodes is often unreliable. However, some applications of the ATHF pertained to lifetime exposure to antidepressants or address exposure to a particular treatment, such as ECT [e.g. (Aaronson et al., 2017)]. Accordingly, the ATHF-SF has data collection forms dedicated to the current episode only or across prior episodes. The ATHF-SF Scoring Checklist provides a comprehensive list of all antidepressant treatments deemed efficacious in the treatment of MDEs based on review of the empirical evidence, treatment guidelines, regulatory approvals, and expert opinion, across the domains of pharmacotherapy, brain stimulation, and psychotherapy. Using a checklist format, notations are made as to whether or not the treatment was administered during the relevant time period and, if administered, whether the treatment exposure should be considered a “failed adequate trial”. The ATHF-SF presents explicit criteria to make this determination for every antidepressant treatment. These four documents can be freely downloaded at the end of this article as supplemental files. There are no restrictions on their use or dissemination.

The workgroup began by discussing fundamental issues in evaluating the adequacy of antidepressant treatments. These included identifying the domains of treatment that should be considered, the patient subgroups that required separate criteria (e.g., bipolar vs. unipolar MDE), how trial-by-trial ratings are made (e.g., dichotomous evaluation of adequacy vs. multi-level rating of trial potency), how antidepressant treatments should be grouped into classes, whether separate criteria should be offered for augmented or combination psychopharmacology trials (e.g., antidepressant plus lithium), and conventions for addressing trial dosage, duration, clinical benefit and adherence. The workgroup achieved consensus on these issues. Indeed, this article and the accompanying ATHF-SF documents were unanimously endorsed by this workgroup. After a draft ATHF-SF was created, the documents were shared with European experts in TRD, Drs. Koen Demyttenaere, Allan H. Young, and Thomas E. Schläpfer. Their comments prompted inclusion of some pharmaceutical agents not available in the United States, as well as modification of the psychotherapy criteria. The four ATHF-SF documents were finalized on December 5, 2018 as ATHF-SF Version 2018.1. The workgroup plans to update the ATHF-SF every 3–4 years.

This article reviews the key questions that must be addressed when evaluating the adequacy of antidepressant treatments and rating treatment resistance. In doing so, we detail the rationale behind the ATHF-SF.

5. Domains of treatment

Definitions of treatment resistance have often been restricted to psychopharmacological trials [e.g. (Thase and Rush, 1997)]. Some definitions include only ECT as a potentially adequate non-pharmacological treatment [e.g. (Fava, 2003; Fekadu et al., 2009)]. In recent years, brain stimulation interventions other than ECT have been found effective in the treatment of MDE. Indeed, TMS and VNS were each approved by the FDA specifically for different degrees of TRD. The ATHF-SF documents the administration of all brain stimulation interventions, and provides explicit criteria to assess the adequacy of three interventions with established efficacy, ECT, TMS, and VNS. Deep Brain Stimulation (DBS) (Dougherty et al., 2015; Holtzheimer et al., 2017), various forms of transcranial Electrical Stimulation (tES) (Brunoni et al., 2017; Inukai et al., 2016; Loo et al., 2018), and phototherapy (Mårtensson et al., 2015; Tuunainen et al., 2004) were determined at this time to have inconclusive evidence regarding efficacy in MDE, and, so while their use is documented, it is not rated for adequacy.

Only the staging model recently offered by Conway et al. (2017) includes insufficient benefit from an adequate trial of psychotherapy as contributing to the assessment of TRD. Neither the original ATHF nor the other instruments used to rate the adequacy of antidepressant trials included psychotherapy as a potential source of treatment resistance. However, particular forms of psychotherapy have an evidence base that supports their efficacy in the treatment of MDE. These include: Behavior Therapy (BT) (Harley et al., 2008; Lynch et al., 2015), Cognitive-Behavioral Therapy (CBT) (Cristea et al., 2015; Feng et al., 2012; Gould et al., 2012; Gregory, 2010; Tolin, 2017), Interpersonal Therapy (IPT) (Cuijpers et al., 2016; Markowitz and Weissman, 2012; Weissman et al., 2014), Problem-Solving Therapy (PST) (Bell and D’Zurilla, 2009; Cuijpers et al., 2018; Townsend et al., 2001), and Short-term Psycho-dynamic Psychotherapy (STPP) (de Roten et al., 2017; Driessen et al., 2015, 2018). Furthermore, in recent years, there has been a burgeoning literature establishing the efficacy of evidence-based psychotherapies specifically in the treatment of TRD (Fonagy et al., 2015; Souza et al., 2016; Town et al., 2017; Trivedi et al., 2011; van Bronswijk et al., 2018). Thus, a significant change in the ATHF is provision of explicit criteria to rate the adequacy of specific forms of psychotherapy.

The workgroup considered and rejected inclusion of complementary, dietary, or naturopathic supplements as efficacious antidepressant treatments (e.g., S-adenosyl-l-methionine [SAM-e], St. John’s Wort, omega-3 fatty acids, folate, B-vitamins). Their exclusion was based on insufficient or inconsistent evidence of efficacy, especially as monotherapies, and concerns about consistency of dosing and bioavailability (American Psychiatric Association, 2010; Sarris et al., 2016). For example, the NIH National Center for Complementary and Integrative Health provides this bottom-line regarding use of St. John’s wort in depression: “St. John’s wort isn’t consistently effective for depression. Do not use it to replace conventional care or to postpone seeing your health care provider.”

6. Patient subgroups requiring distinct criteria

Other than the ATHF, instruments that assess the adequacy of antidepressant treatment have been restricted to nonpsychotic, unipolar MDE (Dessilels et al., 2011). This restriction was based on the evidence that differing pharmacological regimens are effective in the acute treatment of bipolar and unipolar MDE or nonpsychotic and psychotic MDE. For example, some medications have established efficacy in the acute treatment of bipolar MDE without similar information in the case of unipolar MDE (Bhagwagar and Goodwin, 2005; Bowden, 2002, 2003; Deshauer et al., 2005; Fountoulakis et al., 2017; Goodwin et al., 2016; Grunze et al., 2013; Muzina and Calabrese, 2003; Pacchiarotti et al., 2013). Consequently, in the ATHF-SF adequate trials can be achieved with lithium, specific anticonvulsants, or specific atypical antipsychotic medications in the acute treatment of bipolar, but not unipolar, MDE. In contrast, clozapine is treated differently, as criteria are given for adequate trials in either unipolar or bipolar MDE (Suppes and Rush, 1996; Suppes et al., 1999).

6.1. Unipolar vs. bipolar MDE

Another difference between bipolar and unipolar depression in the ATHF-SF is that augmentation strategies apply only to unipolar MDE. In bipolar, nonpsychotic MDE, all treatments are evaluated individually, regardless of their concurrent administration. Monotherapy trials with traditional antidepressant medications, lithium, specific anticonvulsants, and specific atypical antipsychotic medications can meet criteria for adequacy. In unipolar MDE, clozapine is the only antipsychotic that be adequate monotherapy. In contrast, in unipolar MDE, the administration of augmentation trials is specifically evaluated. An adequate augmentation trial in unipolar patients can involve administration of adequate dose and duration of lithium, triiodothyronine (T3), or 5 specific atypical antipsychotics concurrently with adequate dose and duration of an antidepressant medication (e.g., nortriptyline plus lithium). The reason for this difference is that placebo-controlled, randomized trials have supported the efficacy of a variety of monotherapies in bipolar MDE, without comparable evidence
in unipolar MDE (Geddes and Miklowitz, 2013; Goodwin et al., 2016; Thase and Sachs, 2000). On the other hand, the evidence supporting the efficacy of lithium (Bauer et al., 2000; Fawcett, 2003; Januel et al., 2003; Nelson et al., 2014; Schule et al., 2009), T3 (Cooper-Kazaz et al., 2007; Joffe and Sokolov, 2000), or atypical antipsychotics (Gobbi et al., 2018; Nelson and Papakostas, 2009; Papakostas et al., 2007; Simons et al., 2017; Wang et al., 2015) in unipolar MDE derives mainly from trials where they were used in combination with or to augment antidepressant medications.

The role of traditional antidepressant medications in the acute treatment of bipolar MDE is controversial (Geddes and Miklowitz, 2013; Goodwin et al., 2016; Malhi et al., 2015; Pacchiarotti et al., 2013). Overall, the evidence for efficacy is relatively sparse and, at times, inconsistent (Sachs et al., 2007; Sidor and Macqueen, 2011; Vazquez et al., 2011, 2013). There are also the concerns that in some patients these medications may provoke mixed states or switches into mania (Baldessarini et al., 2013; Licht et al., 2008; Vazquez et al., 2011), or may destabilize the long-term course by promoting relapse or rapid-cycling (Altshuler et al., 1995; El-Mallakh et al., 2015). If a traditional antidepressant medication is used in bipolar MDE, most treatment guidelines recommend that a mood stabilizer be co-administered (Geddes and Miklowitz, 2013; Goodwin et al., 2016; Malhi et al., 2015; Pacchiarotti et al., 2013).

In the ATHF-SF, like the ATHF, no distinction is made between bipolar and unipolar MDE in evaluating mono- and combination treatments with antidepressant medications. This approach was taken since meta-analyses have indicated that overall traditional antidepressants are more effective than placebo in bipolar MDE, and that the size of this effect is comparable to what is typically observed with unipolar MDE (Sidor and Macqueen, 2011; Vazquez et al., 2011, 2013). Furthermore, MDE with reversed vegetative features (i.e., atypical depression) is especially common in individuals with bipolar II MDE (Akiskal, 2005; Baldessarini et al., 2010; Blanco et al., 2012). In this subgroup, MAOIs, a specific class of traditional antidepressants, may be especially effective (Liebowitz et al., 1988; Quitkin et al., 1988).

Much of the concern about traditional antidepressants in bipolar MDE pertains their potential to produce mood instability, leading to the recommendations for combined treatment with a mood stabilizer. This destabilization is not universally observed or may only characterize a subgroup (Amsterdam et al., 2015). Regardless, the side effect potential of interventions does not impinge on ATHF-SF ratings of trial adequacy. Rather, interventions presumed to be potentially efficacious are evaluated in terms of dose, duration, adherence, and clinical outcome in order to determine resistance to a minimally adequate treatment regimen. Thus, ATHF-SF conventions for evaluating treatment resistance are not prescriptive regarding optimal care.

6.2. Psychotic vs. nonpsychotic MDE

Like the ATHF, the ATHF-SF has distinct criteria for pharmacological treatment of nonpsychotic and psychotic MDE. This approach derives from evidence that in psychotic MDE monotherapy with traditional antidepressant or antipsychotic medication is substantially less effective than combination treatment (Blumberger et al., 2011; Farahani and Correll, 2012; Glassman and Roose, 1981; Nelson et al., 1986; Spiker et al., 1985; Wijckstra et al., 2010, 2015). Thus, in the ATHF-SF, no trial of an antidepressant medication is considered adequate in the treatment of psychotic MDE unless that trial is combined with adequate dosage and duration of an antipsychotic medication. The minimal adequate dosage for the antipsychotic medication is 400 mg/d chlorpromazine (CPZ) equivalents based on research with first generation antipsychotics suggesting this cutoff (Nelson et al., 1986). The ATHF-SF Instruction Manual contains an appendix providing CPZ equivalents for first and second-generation antipsychotic medications.

ECT is the only non-pharmacological intervention with substantial evidence regarding potential differential efficacy in bipolar and unipolar MDE and nonpsychotic and psychotic MDE. Response and remission rates to ECT do not differ as a function of polarity, although bipolar patients as a group improve more quickly and require fewer treatments (Daly et al., 2001; Sackeim and Prudic, 2005; Sienaert et al., 2009). Both non-psychotic and psychotic MDE remit at high rates with ECT, and there is evidence that psychotic MDE may be especially ECT responsive (Flint and Rifat, 1998; Janicak et al., 1989; Petrides et al., 2001; Sobin et al., 1996; Solan et al., 1988). Thus, the ATHF-SF criteria for adequate trials of ECT do not differ for these subgroups. Due to the absence of evidence of differential efficacy, ATHF-SF criteria also do not differentiate among these subgroups in rating the adequacy of other brain stimulation interventions or psychotherapies.

7. Evaluation of treatment adequacy: dichotomy or continuum?

As indicated in Table 2, the original ATHF scored each antidepressant trial on a 5-point scale, with scores of “3” corresponding to an adequate trial and higher scores reflecting greater potency due to either increased dosage or blood level of a medication or use of specific augmentation strategies. This approach to scoring potency made use of the ATHF complex. Data collection involved tracking changes in dosages of medication and their specific duration to identify the highest dosage for the minimally acceptable period. The concurrent administration of medications needed to be tracked to identify trials requiring higher scores due to adequate augmentation regimens. The approach taken in the ATHF confounded higher potency scores due to high dosage with higher scores due to use of an adequate augmentation strategy.

Since the publication of the ATHF, alternative instruments have been proposed, such as the Antidepressant Treatment Record (ATR) (Carpenter et al., 2012; Dunner et al., 2014) and the ATRQ (Desseilles et al., 2011). Both instruments use a checklist format to identify minimally adequate treatment trials. Both instruments provide minimum dosages for antidepressant medications and minimum durations at that dose to identify adequate trials. Both instruments also collect information on clinical response to these trials, in order to identify “failed, adequate trials.” Both instruments make separate ratings of trials involving pharmacological augmentation or ECT.

Like the ATR and ATRQ, the ATHF-SF differs from the ATHF in that the scoring of each trial requires only the dichotomous determination of whether the trial was adequate, i.e., corresponding to a score of “3” on the ATHF. This change markedly simplifies scoring. All antidepressant interventions, pharmacological or otherwise, are rated independently, with notable exceptions made for (a) psychotic depression and (b) specific augmentation trials in unipolar depression, as discussed below. Other than these two exceptions, there is less need with the ATHF-SF to keep track whether medications were administered simultaneously or to track the dose changes of every medication. This greatly simplifies data collection and scoring. To be adequate, all pharmacological agents must be given for a minimum of 4 weeks at or above the stipulated minimum threshold dose or blood level. Once this is ascertained for a medication (or comparable criteria for brain stimulation and psychotherapy interventions), no further inquiry is necessary.

These changes were also prompted by the fact that few studies using the ATHF have reported potency scores [e.g. (Prudic et al., 1990; Sackeim et al., 1990)]. Almost all reports have focused on the total number of trials administered and the number of failed adequate trials (either in the current episode or lifetime). These summary scores are available with the ATHF-SF.

The ATHF-SF provides criteria to rate the adequacy of all interventions deemed effective treatments of MDE. These interventions are assessed in terms of whether they were administered and whether the intervention is considered a “failed adequate” trial. Two other types of intervention are also documented. Some interventions are borderline in terms of their evidence of efficacy. Promising treatments are
documented in the ATHF-SF that have not yet met the threshold for established efficacy, or where knowledge is lacking of minimal effective dose and duration. Examples include ketamine (Fava et al., 2018; Sanacora et al., 2017; Singh et al., 2016; Wilkinson et al., 2018) and pramipexole (Dell'OssO and Ketter, 2013; Fawcett et al., 2016; Gauthier et al., 2017; Lattanzi et al., 2002). In contrast, reboxetine is included in this grouping since information subsequent to its regulatory approval in Europe has created doubt regarding its efficacy. For brain stimulation interventions, this grouping includes various forms of tES, DBS, and light therapy. The administration of any of these specific interventions is noted (at any dose or duration), but they cannot constitute adequate trials. The third grouping involves classes of medication often used adjunctively in MDE treatment but thought to have minimal efficacy on core MDE symptoms. These classes include benzodiazepines, stimulants, specific anticonvulsants and antipsychotics in bipolar MDE, and specific antipsychotics in unipolar MDE. Notation is made if any member of the class was administered during the rating period, regardless of whether multiple members of the class were prescribed.

The structure of the ATHF-SF distinguishes among treatment domains (pharmacotherapy, brain stimulation, psychotherapy), classes of treatment, and specific interventions within a class. For pharmacological treatment, classes are defined following traditional distinctions regarding presumed pharmacological mechanisms (e.g., SSRI, SNRI, TCA, anticonvulsants, etc.). Each brain stimulation or psychotherapy intervention is considered a separate class. Thus, the ATHF-SF can provide summary scores for the number of antidepressant trials attempted and the number of “failed adequate” trials within the three domains, within specific treatment classes, and across all interventions.

8. Combination and augmentation strategies

As noted, due to the evidence that the combination of antipsychotic and antidepressant medication is more effective than either alone in the treatment of psychotic MDE (Blumberger et al., 2011; Farahani and Correll, 2012; Glassman and Roose, 1981; Nelson et al., 1986; Spiker et al., 1985; Wijkstra et al., 2010, 2015), the ATHF-SF requires such combination treatment for an adequate pharmacological trial in this subgroup.

In unipolar MDE, there is also a substantial literature on the use of particular augmentation strategies following lack of response with a traditional antidepressant medication. Randomized placebo-controlled trials have shown that adding lithium (Bauer et al., 2000; Fawcett, 2003; Januel et al., 2003; Nelson et al., 2014; Schule et al., 2009), T3 (Bauer et al., 2000; Fawcett, 2003; Januel et al., 2003; Nelson et al., 2014; Schule et al., 2009), or specific atypical antipsychotic medications (Gobbi et al., 2018; Nelson and Papakostas, 2009; Papakostas et al., 2007; Simons et al., 2017; Wang et al., 2015) are effective augmentation strategies (Thase, 2009; Thase and Rush, 1997; Zhou et al., 2015). The workgroup felt that the information conveyed from failure to benefit from an adequate augmentation strategy in unipolar MDE was different from failing adequate monotherapy trials of the constituents. In contrast, augmentation strategies are less established in bipolar MDE. Instead, in this subgroup there is evidence that specific agents often used in augmentation strategies, such as lithium or particular anticonvulsants or atypical antipsychotics are effective as monotherapies (Geddes and Miklowitz, 2013; Goodwin et al., 2016; Thase and Sachs, 2000). Consequently, in nonpsychotic bipolar MDE all treatments are evaluated individually regardless of their concurrent use, as if they were administered as monotherapies. This is also the case in nonpsychotic, unipolar MDE, except for the concurrent administration of augmentation strategies with lithium, T3, or specific atypical antipsychotics. While augmentation trials imply prior lack of benefit with the antidepressant being “augmented”, in the ATHF-SF any period of failed adequate concurrent treatment is considered when evaluating “augmentation” trials.

9. Dosage, duration, adherence, and clinical outcome

It is generally agreed that any definition of a “failed adequate” trial must operationalize (a) the minimum dose of an intervention considered effective; (b) the minimum duration of treatment at or above that dosage for the trial to be considered adequate; (c) the extent to which nonadherence with the regimen disqualifies a trial; and (d) the extent of clinical improvement that justifies characterizing the trial as unsuccessful or “failed” (Berlim and Turecki, 2007a; Sackeim, 2001). In reviewing 47 RCTs in TRD, Berlim and Turecki (2007b) found marked variability in the conventions used to define each component.

9.1. Dosage and duration

The ATHF, ATR, and ATRQ provide criteria for the minimum effective dose and duration for pharmacological treatments to be considered adequate. These criteria largely reflect the minimum dose for which there is evidence from RCTs demonstrating antidepressant efficacy. The ATHF-SF follows this principle. In addition, for some medications blood level information takes precedence over oral dosage (e.g., nortriptyline, lithium), and minimal values are given for adequate blood levels.

There is also considerable variability in the definitions used for adequate duration of pharmacotherapy. A minimum of four weeks has been the most commonly used cutoff, but TRD studies have also required six or eight weeks for trials to be considered adequate (Berlim and Turecki, 2007b). The ATHF (Sackeim, 2001) and ATR (Carpenter et al., 2012; Dunner et al., 2014) use a four-week minimum cutoff, while the ATRQ (Desseilles et al., 2011) requires 6 weeks. Given uncertainty about the optimal duration cutoff, the ATRQ also documents whether the minimal dosage was given for at least 10 weeks.

Other than ketamine (Sanacora et al., 2017), there is little evidence that antidepressant regimens differ in their speed of improvement (Gelenberg and Cheson, 2000; Machado-Vieira et al., 2010). Consequently, the workgroup determined that use of the same treatment duration criterion across all pharmacological interventions was justified and would greatly simplify data collection and scoring. While the duration of acute phase treatment in antidepressant RCTs has increased in recent decades (Sackeim et al., 2006), this does not bear directly on the question of what minimally adequate trial duration is informative of treatment resistance. There is substantial evidence that early clinical improvement with antidepressant pharmacotherapy has positive predictive value regarding ultimate response or remission, and that early lack of symptomatic improvement has negative predictive value (Leuchter et al., 2009; Nierenberg et al., 1995; Papakostas et al., 2006; Sackeim et al., 2006; Szegedi et al., 2009).

The ATHF-SF maintains the convention adopted in the ATHF and ATR requiring a four-week minimum duration to declare a trial as adequate. While this choice was arbitrary, it had the advantage of recognizing that total duration of exposure to a medication is often longer due to upward titration of dose at treatment outset. More critically, it was thought that failure to show substantial improvement after a minimum four-week trial was informative of likely resistance to that medication. Use of this cutoff also had the advantage of continuity with earlier studies using the ATHF and ATR.

Even though the original ATHF only required a minimum four-week duration of medication exposure, in many studies the total number of medication trials administered was two-to-four times more common than the total number of failed adequate medication trials (George et al., 2010; Lisanby et al., 2009; Sackeim et al., 2001b, 2009). Low rates of adequate antidepressant treatment (i.e., below minimal thresholds for dose and/or duration) have been repeatedly documented, since the earliest assessment of treatment adequacy in MDE (Keller et al., 1982, 1986) and until the present (Andreescu et al., 2007; Fife et al., 2018; Kocsis et al., 2008). For example, using the ATHF, Kocsis et al. (2008) found that only 33% of 80 patients with chronic...
forms of MDE had ever received an adequate antidepressant medication trial. The minimum treatment duration for pharmacotherapy is perhaps the ATHF-SF criterion most likely to be modified for particular research needs. It should be recognized that lengthening this criterion (or increasing medication dose) may purify TRD samples by reducing the false positive rate, but will also increase false negatives and reduce the incidence of TRD.

The adequacy of pharmacological treatment can be characterized in terms of oral dosage (or blood levels) and treatment duration. In contrast, multiple dimensions potentially impact on the adequacy of brain stimulation and psychotherapy interventions. For example, the efficacy and side effects of ECT are contingent on electrode placement and electrical dosage relative to seizure threshold, electrical waveform, and other technical factors (Sackeim et al., 1993, 2000, 2008; Semkovska et al., 2016). TMS can be characterized in terms of site of stimulation and method of ascertainment, dosing relative to motor threshold, stimulation parameters and pulse number, coil geometry, frequency of treatments, treatment number, etc. (Blumberger et al., 2018; George et al., 2010; Levkovitz et al., 2015; O’Reardon et al., 2007). Completely different dimensions apply to psychotherapies, including session duration, frequency, and number, individual, couple, family, or group format, manual-guided, therapist experience, theoretical orientation, etc. (Garland et al., 2010). The workgroup recognized that patients rarely can report the technical details regarding previous brain stimulation or psychotherapy interventions, and this information is also often missing from medical records.

In this context, the workgroup adopted a minimalist approach to evaluating trial adequacy, similar in concept to the approach taken with medications. Brain stimulation and psychotherapy interventions all require a minimum number of “sessions” or treatments within a prescribed timeframe. Specifically, an adequate trial of ECT corresponds to a course of at least eight treatments given within a 5-week period (Kellner et al., 2016; Sackeim et al., 2009). An adequate trial of TMS involves at least 20 sessions within a 6-week period of either fast (≥ 5 Hz) left frontal or slow (≤ 1 Hz) right frontal stimulation (Carpenter et al., 2012; Chen et al., 2013; Gaynes et al., 2014; Liu et al., 2014). Given the evidence for an especially slow onset of benefit, an adequate trial of VNS corresponds to implantation and stimulation for a year or more, with the current set at 0.75 mA or higher (Aaronson et al., 2017; Rush et al., 2005a, 2005b; Sackeim et al., 2001b, 2007).

Five forms of evidence-based psychotherapy are identified in the ATHF-SF as potentially adequate treatments: Behavior Therapy (BT), Cognitive-Behavioral Therapy (CBT), Interpersonal Therapy (IPT), Problem-Solving Therapy (PST), and Short-term Psychodynamic Psychotherapy (STPP). To be considered adequate, patients must participate in at least 12 sessions within a 15-week period. Thus, the same criteria for minimum intensity of treatment apply to all psychotherapy interventions. Treatment may be given in an individual, couple, family, or group format as long as the primary orientation reflects one of the evidence-based psychotherapies and the primary target of treatment is the patient’s depression.

9.2. Adherence

Studies of TRD rarely describe compliance or adherence criteria when evaluating antidepressant resistance (Berlim and Turecki, 2007b). However, failure to receive a therapy due to nonadherence (for whatever reason) contraindicates evaluating the treatment as adequate. The ATHF-SF explicitly states that to rate a trial as adequate, “there should not be evidence that the patient was substantially non-adherent with the treatment regimen.” For pharmacotherapy, substantial non-adherence would involve evidence (including self-report) that 25% or more of the prescribed medication doses were not administered during the relevant period. For psychotherapy, evidence of non-adherence might include refusal to engage in sessions, lack of completion of assignments, etc. Non-adherence is rarely an issue for brain stimulation interventions as their administration does not typically require active patient participation. Explicit evidence of non-adherence is needed to exclude a trial from being rated as adequate.

9.3. Treatment outcome

When used solely to evaluate the adequacy of care, treatment therapeutic outcome is not a relevant consideration. The ATHF has been used to characterize the intensity or adequacy of pharmacological treatment in a variety of subgroups (Andreasen et al., 2007; Kocsis et al., 2008; Mulsant et al., 1997; Oquendo et al., 1999, 2002). However, the concept of treatment resistance requires there be insufficient benefit from an adequate trial. Studies of TRD have varied markedly in the criteria used to document that previous trials were unsuccessful. Criteria have ranged from failing to achieve complete remission to failing to manifest a 30% reduction in symptom severity (Berlim and Turecki, 2007b). This diversity reflects the tension between the consensus that the goal of antidepressant treatment should be symptomatic remission, especially since the extent of residual symptoms may predict relapse, and the clinical reality that for some patients partial improvement may be particularly meaningful in the context of a chronic and treatment-resistant course.

In line with the majority of studies in this area (Berlim and Turecki, 2007b), the ATHF-SF takes a middle course between these positions. To be rated as a “failed, adequate” trial, patients should not have shown marked clinical improvement or remission at the end of the treatment period. The ATHF-SF uses a modified Clinical Global Impression — Improvement Scale (CGI-I) (Guy, 1976), where scores of “1” and “2” correspond to “very much improved” and “much improved.” Any trial receiving a score of 1 or 2 is disqualified as a “failed, adequate” trial. Scores of 3 through 7 correspond to the traditional CGI-I anchors ranging from “minimally improved” to “very much worse.” Any score in this range would constitute an unsuccessful trial. This demarcation also corresponds to the distinction between “response” and “non-response” (Frank et al., 1991; Rush et al., 2006a), such that any patient showing a 50% or greater improvement in symptom severity should not be considered as having a failed trial. The ATHF-SF adds another anchor in scoring the CGI-I for each trial. Scores of “8” correspond to a trial in which the patient was much or very much improved but then relapsed on the same regimen. Such instances of antidepressant tachyphylaxis often indicate the onset on a new MDE (Targum, 2014). The ATHF-SF Instruction Manual makes specific recommendations on how to evaluate instances of loss of benefit for previously effective antidepressant treatment.

10. Sources of data

The reliability and validity of the information submitted to ATHF-SF scoring are enhanced by comprehensive attempts to obtain information from diverse sources. Pharmacy records may be especially useful in documenting prescribed oral dosage and the duration of treatment. Patient and family member interviews may be key to evaluating adherence and clinical outcome. Consultation with past providers may be needed to verify clinical outcome and the reason for trial discontinuation. Reviewing medical records may be necessary to document blood levels. While obtaining comprehensive information on antidepressant trial details can be onerous and time consuming, such information is necessary for accurate evaluation of treatment adequacy and resistance.

11. Conclusions

There is considerable diversity in the definitions and metrics used to categorize and quantify antidepressant treatment resistance. However, all approaches to this problem require that patients experience insufficient benefit during one or more adequate antidepressant treatment trials, and the identification of “failed, adequate trials” is at the
core of methods for assessing TRD. The evaluation of antidepressant treatment trials in terms of their adequacy, adherence, and clinical outcome raises a host of complex issues, and the empirical evidence often offers only limited guidance. The field has not achieved consensus on many of these topics. Nonetheless, the reliable and valid identification of TRD is fundamental to treatment planning, research, and public policy. We have summarized the rationale behind the ATHF-SF and described how it addresses each of the key issues.

Declaration of interest

Dr. Harold A. Sackeim has served as a consultant and/or received research support from the brain stimulation companies:Brainsway Ltd., Cervel Neurotech Inc./NeoStim Inc., Cyberonics Inc., LivaNova PLC, Magstim Ltd., MECTA Corp., NeoSync Inc., Neuronetics Inc., and NeuroPace Inc. and from the pharmaceutical companies: Cambridge Neuroscience Inc., Eli Lilly & Co., Forest Laboratories, Hoffman-La Roche AG, Interneuron Pharmaceuticals Inc., Novartis International AG, Pfizer Inc., Warner-Lambert, Inc., and Wyeth-Ayerst, Inc. He holds non-remunerative patents with the MECTA Corporation for Focal Electrically-Administered Seizure Therapy (FEAST) and for Titration in the Current Domain. He is the originator of Magnetic Seizure Therapy. He has received royalties from Oxford University Press and Elsevier Inc. His effort in chairing this revision of the ATHF and drafting the ATHF-SF documents was partially supported by LivaNova PLC. LivaNova PLC partly supported the deliberations that resulted in the ATHF-SF with the aim of using this instrument in future research on VNS in TRD. However, neither LivaNova PLC, or any other commercial entity, had any substantive influence on the design or content of the ATHF-SF.

Dr. Scott T. Aaronson reports no conflicts of interest directly relating to this work. He does report a number of other relationships with commercial entities. He serves as a consultant to LivaNova PLC, Neuronetics Inc., Janssen Pharmaceuticals Inc., and Genomind. He has received research support from Neuronetics Inc. He has been paid speaker honoraria by Sunovion Pharmaceuticals Inc. and Neuronire Biosciences Inc.

Dr. Mark Bunker reports no conflicts of interest directly relating to this work. He is a consultant to LivaNova PLC.

Dr. Mark Demitrack reports no conflicts of interest directly relating to this work. He serves as a consultant and has stock ownership in Neuronetics, Inc. He is an employee of Axovant Sciences, Inc.

Dr. Charles R. Conway reports no conflicts of interest directly relating to this work. He has received research support from Bristol-Myers Squibb Co., Stanley Medical Research Institute, National Institute of Mental Health, NeoSync Inc., Cyberonics Inc., Taylor Family Institute for Innovative Psychiatric Research, August Busch IV Foundation, and Barnes-Jewish Hospital Foundation. He previously served as a speaker for Bristol-Myers Squibb Co. and Otsuka Pharmaceuticals Co. He currently serves as a paid consultant to LivaNova PLC in designing studies involving VNS.

Dr. Mark S. George reports no conflicts of interest directly relating to this work. He has no equity ownership in any device or pharmaceutical company. He does occasionally consult with industry, although he has not accepted consulting fees from anyone who manufactures a TMS device, because of his role in NIH and DOD/VA studies evaluating this technology. His total industry related compensation per year is less than 10% of his total university salary. In the past two years, his industry involvement has included Brainsonix Corp. (unpaid consultant), Brainway Ltd. (unpaid consultant, research grant, donated equipment), Cervel Neurotech Inc./NeoStim Inc. (unpaid consultant, research grant), LivaNova PLC (consultant), MECTA Corp. (unpaid consultant, research grant), Microtransponder Inc. (DSMB member), Neuronetics Inc. (unpaid consultant, research grant, donated equipment), NeoSync (unpaid consultant, research grant), Nerive (unpaid consultant), Pure Tech Health Ventures (consultant).

Dr. Joan Prudic reports no conflicts of interest.

Dr. Michael E. Thase reports no conflicts of interest directly relating to this work. He does report a number of other relationships with commercial entities. During the past 3 years, he has been an advisory consultant to Acadia, Alkermes, Allergan (Forest, Naurex), AstraZeneca, Cerocor, Eli Lilly, Johnson & Johnson (Janssen, Ortho-McNeil), Lundbeck, MedAvante, Merck, Modaksha, Nestlé (Pamlab), Neuronetics, Novartis, Otsuka, Pfizer, Shire, Sunovion, and Takeda. In addition to the National Institute of Mental Health, he received grant support from Acadia, the Agency for Healthcare Research and Quality, Alkermes, Assurex, Avanir, Forest Pharmaceuticals, Johnson & Johnson, Otsuka Pharmaceuticals, and Takeda. Dr. Thase received royalties from the American Psychiatric Press, Guilford Publications, Herald House and W.W. Norton & Company, Inc. Dr. Thase’s spouse, Dr. Diane Sloan, works for Peloton Advantage, which did business with Pfizer and AstraZeneca.

Dr. A. John Rush has served as a consultant to Akili Inc., American Psychiatric Association, Brain Resource Ltd., Compass Inc., Curbstone Consultant LLC., Cyberonics Inc., Eli Lilly, Emmes Corp., Holmusk, LivaNova PLC, National Institute of Drug Abuse, Santium Inc., Sunovion, Taj Medical, Takeda USA. He has received speaking fees from LivaNova and royalties from Guilford Publications and the University of Texas Southwestern Medical Center, Dallas, TX. (for the Inventory of Depressive Symptoms and its derivatives). He is a named co-inventor on U.S. Patent No. 7,795,033: Methods to Predict the Outcome of Treatment with Antidepressant Medication, Inventors: McMahon FJ, Laje G, Manji H, Rush AJ, Padlock S, Wilson AS and on U.S. Patent No. 7,906,283: Methods to Identify Patients at Risk of Developing Adverse Events During Treatment with Antidepressant Medication, Inventors: McMahon FJ, Laje G, Manji H, Rush AJ, Padlock S. His effort in assisting in the drafting of this revision of the ATHF and associated ATHF-SF documents was partially supported by LivaNova PLC.

Acknowledgements

We thank Professors Koen Demyttenaere, Allan H. Young and Thomas E. Schlüper for their comments on the ATHF-SF. Their comments prompted inclusion of some pharmaceutical agents not available in the United States, as well as modification of the psychotherapy criteria. We also thank Ms. Amy Keith and Ms. Renske De Zwaaf of LivaNova PLC for organizational assistance with the deliberations of the co-authors that resulted in the ATHF-SF 2018 Version 1.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jpsychires.2019.03.021.

References


