Staging Treatment Intensity and Defining Resistant Depression: Historical Overview and Future Directions

Naji C. Salloum, MD,[a] and George I. Papakostas, MD[a,*]

ABSTRACT

Objective: To review existing staging models and definitions of treatment-resistant depression (TRD) and offer future directions within the context of up-to-date evidence.

Data Sources: A PubMed search was conducted on February 25, 2018, for articles in English on TRD staging or definition using the following keywords: *depressive disorder, treatment-resistant* OR *treatment resistant depression* cross-referenced with *staging* OR *degree* OR *level* OR *definition*. Relevant cross-references from identified articles were also included.

Study Selection: A total of 18 articles were identified that included a proposed TRD staging model, a proposed TRD definition, empirical work to support a model or definition, or any combination thereof.

Data Extraction: Included articles were summarized in chronological order in terms of the date the TRD staging model (and accompanying TRD definition if applicable) was first proposed. Findings from validation studies pertaining to staging or definition were then synthesized.

Results: Five staging models were identified. Strengths identified across staging models include rigorous assessment of adequacy of treatment, differentiation of resistance versus symptom return, assignment of equal weights to different pharmacotherapies, and accounting for augmentation. Future considerations should include differential weighting to specific augmentation agents based on available evidence, added weight to electroconvulsive therapy and ketamine treatments, and the addition of evidence-based psychotherapies. Dichotomous versus continuous approaches to TRD diagnosis were considered, with the latter (beginning with 1 failed trial) best explaining available data from large trials.

Conclusions: The most up-to-date evidence in the literature should guide future research in the definition and staging of TRD.

J Clin Psychiatry 2019;80(4):18r12250

To cite: Salloum NC, Papakostas GI. Staging treatment intensity and defining resistant depression: historical overview and future directions. *J Clin Psychiatry.* 2019;80(4):18r12250.

To share: https://doi.org/10.4088/JCP.18r12250

[a]Clinical Trials Network and Institute, Massachusetts General Hospital, Boston, Massachusetts

**Corresponding author:* George I. Papakostas, MD, One Bowdoin Sq, 9 Floor, Boston, MA 02114 (GPapakostas@partners.org).

Treatment-resistant depression (TRD) is associated with a reduced quality of life, high rates of medico-psychiatric comorbidities, increased health care expenditures, and social and occupational impairment leading to poor treatment outcomes.[1–4] The concept of TRD started emerging in the 1970s to describe a group of patients suffering from major depressive disorder (MDD) who fail to respond to treatment.[5] Despite the continued development of new antidepressants over the following decades, a large proportion of depressed patients fail to respond to available antidepressant therapies.[6,7]

It was recognized early on that to better understand how to optimize treatment choices for individuals with TRD, efforts should also be made to accurately and unambiguously define what TRD is. Several authors have postulated various definitions since the early 1990s, with the two most widely recognized and utilized definitions being (1) failure to respond to at least 1 antidepressant and (2) failure to respond to 2 or more antidepressants during the current episode.[8–11] Efforts were further extended to develop staging schemes for different subcategories of subjects based on ascending levels of treatment resistance.[9,12–15] The definition and staging of TRD often rely, however, on expert opinion not always substantiated by the existing empirical evidence. This discordance explains the relative heterogeneity of TRD definitions across modern research studies.[16]

Nonetheless, over time, informative data accumulate, offering opportunities to continually reassess and revise old viewpoints. Ultimately, a simple, rational, broad, and easily generalizable definition of TRD that is widely accepted can help in the design of future studies and development of treatment guidelines and algorithms, as well as provide a regulatory basis for approval of novel therapies in MDD.

Toward that end, in the present work, we systematically review, chronologically describe, and critically evaluate the different definitions and staging schemes of TRD within their historical, evidence-based context and in the context of contemporary clinical evidence. We also extrapolate a set of criteria that we think are most supported by current evidence and offer future directions to refine the way we define and stage TRD.

METHODS

Search Methods

We conducted a search in PubMed for articles published on or before February 25, 2018, in English, that were related to the staging or definition of TRD. The full PubMed search text was as follows: (*depressive disorder*, *treatment-resistant* [MeSH Terms] OR (*depressive* [All Fields] AND *disorder* [All Fields] AND *treatment-resistant* [All Fields]) OR *treatment-resistant depressive disorder*

Clinical Points

- Although there is no universally agreed upon operational definition of treatment-resistant depression (TRD), the vast majority of research studies define TRD as a dichotomous measure after 1 or 2 failed adequate trials.
- In clinical practice, TRD should be thought of on a continuous spectrum starting with 1 failed adequate antidepressant trial.
- Clinicians should adapt data from phase III TRD studies to best suit the needs of patients in real-world treatment settings.

[All Fields] OR (*treatment* [All Fields] AND *resistant* [All Fields] AND *depression* [All Fields]) OR *treatment resistant depression* [All Fields]) AND (*staging* [All Fields] OR *definition* [All Fields] OR *degree* [All Fields] OR *level* [All Fields]).

Study Selection

The full PubMed search identified 498 articles. We further searched bibliographies of relevant articles and identified 5 works that were not identified in our original search results. Subsequently, we reviewed the title and abstract of the 503 articles and excluded 473 articles based on the following exclusion criteria: non-English, nonhuman, editorial, no TRD content, or no staging or definition content. For the remaining 30 articles, we reviewed the full text and excluded an additional 12 articles based on the same exclusion criteria. Of note, 1 study[17] used a data-driven analysis to derive an empirical definition of TRD. However, the a priori TRD definition used was whether or not a subject received brain stimulation, which we deemed poorly reflective of true TRD and hence excluded the study from our review. Data were extracted from a total of 18 articles (Figure 1), synthesized, and presented in the Results section.

Presentation of Findings

Selected studies are grouped into the different staging models and the respective definitions of treatment resistance used, and they are summarized and presented chronologically. The findings are also presented within the context of the existent evidence-based literature at the time each model was developed. We then present all studies that empirically test the different staging models or support definitions of TRD.

RESULTS

Early Efforts and the Development of the Antidepressant Treatment History Form

The Antidepressant Composite Score (ACS), part of the Collaborative Depression Study funded by the National Institute of Mental Health Clinical Research Branch,[18] was an effort to standardize the way to assess the intensity of all previous antidepressant treatments by assigning a composite score based on the type, dosage, and duration of treatment (see Table 1). The dosage equivalents were predefined by consensus following "discussions among experienced clinicians and psychopharmacology researchers in the Collaborative Depression Study"[18(p460)] and supplemented by reviews of the literature performed between 1969 and 1982 (Baldessarini,[19] Klein and David,[20] Cole and David,[21] Appleton[22]). In 1990, Sackeim and colleagues[12] adapted the ACS to create their own version of antidepressant treatment intensity measurement, which was shown to correlate with time to relapse after electroconvulsive therapy (ECT). This scoring system was further developed and updated into the Antidepressant Treatment History Form (ATHF) in a 2001 article by Sackeim.[11]

These early schemes, the ACS and the Sackeim rating scale (and the ATHF, its updated version), had several notable advantages. First, these works represented the earliest recorded efforts to impose a requirement of minimum threshold for the duration of treatment when defining treatment intensity. These efforts highlighted the growing appreciation that most antidepressants' onset of action is on the order of weeks.[23] Second, all antidepressants are considered to have equivalent efficacy, which is more in tune with the results of meta-analyses[24–30] conducted over the past 30 years comparing the efficacy of individual antidepressant agents or antidepressant classes. This consideration was reflected in randomized, placebo-controlled, head-to-head clinical trials of antidepressants for major depressive disorder preceding 1986, the date of publication of the ACS[18]; 3 such studies[31–33] of agents commercially available in the United States summarily did not demonstrate dramatic differences in efficacy between agents.

On the other hand, some weaknesses are also noted. In the Keller et al[18] staging method, a single score is assigned to a given patient's past trials based solely on the trial that is considered the most "intense." This weakness was remediated in the ATHF by allowing for scoring of each trial separately and, subsequently, summing all scores. The ATHF also had several noteworthy limitations. An antidepressant trial that leads to nonresponse on the one hand or response and later relapse on the other hand is considered a failed trial in either case, a point made clear by Sackeim in 2001.[11] Another limitation rests in that both the ACS and the ATHF assume a linear relationship between antidepressant dose, across a broad range, and efficacy or degree of treatment intensity. While a relationship between blood level and clinical response has been recognized for many tricyclic antidepressant (TCAs), this cannot be extended to other, newer antidepressants that appear to have a "flat" dose-response "curve."[34] Finally, a failed treatment trial in the ATHF is assigned a score of 1, even for patients receiving subtherapeutic doses of medication or treated for a very short duration (less than 4 weeks). To illustrate this weakness, someone who tried 4 antidepressant trials at subtherapeutic doses or with short durations would be scored the same as someone having had an adequate TCA trial (score of 4), representing a clinical paradox.

Definition of treatment resistance. The early work by Sackeim et al[12] appears to have influenced their choice of

Figure 1. Flowchart of Study Selection

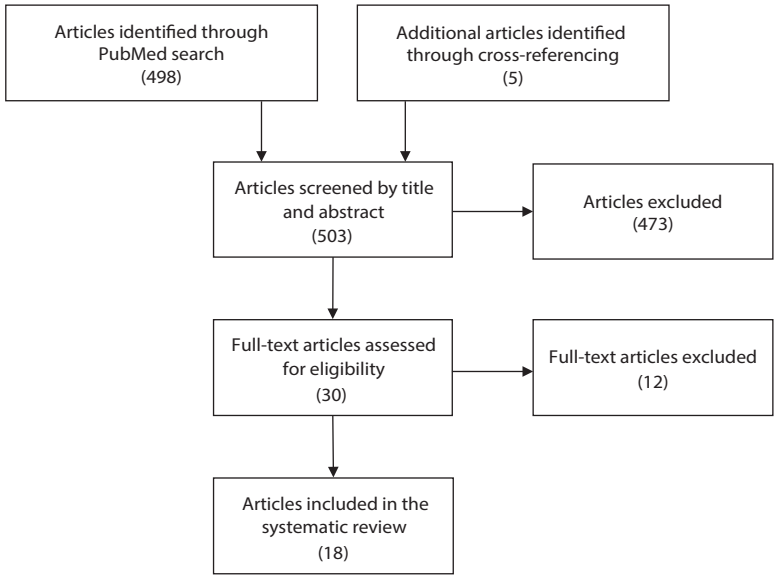

Table 1. TRD Staging Models and Definitions

Variable	ATHF	TR-S	ESM	MGH-S	MSM
Staging parameters	Sum of AD trials, scored individually on a scale of 1–4 depending on trial duration and dosage	Stage I: ≥ 1 AD trial failure Stage II: ≥ 2 AD trial failures from different classes Stage III: stage II + TCA failure Stage IV: Stage III + MAOI failure Stage V: Stage IV + bilateral ECT failure	(A) Nonresponder: 1 AD trial failure (B) TRD: 2 AD trial failures from different classes; subsections 1–5 according to trial duration (C) CRD: depressive episode ≥ 1 y despite multiple trials + augmentation	Sum of AD trials scored individually: Adequate monotherapy (1) Optimal monotherapy (1.5) Augmentation/combination (0.5) ECT (3)	Episode duration (score 1–3) Baseline symptom severity (1–5) AD trial failures (1–5) Augmentation (0–1) ECT (0–1) Maximum score of 15
Key advantages	All AD classes contribute equally to resistance All AD failed trials counted	Failed ECT contributes most to resistance score Accounts only for adequately delivered trials	All AD classes contribute equally to resistance	All AD classes contribute equally to resistance All AD failed trials counted Differentiation of adequate vs optimal trial based on dose Failed ECT contributes most to resistance score Augmentation/combination included	All AD classes contribute equally to resistance Failed ECT included Augmentation included
Key limitations	Tachyphylaxis considered as nonresponse Linear dose-response curve assumed for all ADs Subtherapeutic doses contribute to score	Across-class switch failure leads to higher resistance score MAOIs and TCAs lead to higher resistance scores Augmentation not included	Increased trial duration contributes to higher resistance Augmentation not included in stages 1–5 Failed ECT and pharmacotherapy contribute equally to degree of resistance	Failed monotherapy assumed to contribute more to resistance than failed augmentation Does not specify ECT number of sessions or modality	Illness and naturalistic factors as contributors to treatment resistance Binary score for ECT and augmentation
TRD definition	≥ 1 AD trial nonresponse	≥ 2 AD trial failures from different classes[a]	≥ 2 AD trial failures from different classes	≥ 1 AD trial nonremission	≥ 1 AD trial failure

[a]Authors also suggest "modest resistance may include an inadequate response to a single antidepressant trial."
Abbreviations: AD = antidepressant, ATHF = Antidepressant Treatment History Form, CRD = chronic resistant depression. ECT = electroconvulsive therapy, ESM = European Staging Model, MAOI = monoamine oxidase inhibitor, MGH-S = Massachusetts General Hospital Staging Model, MSM = Maudsley Staging Method, TCA = tricyclic antidepressant, TRD = treatment-resistant depression, TR-S = Thase and Rush Staging Model.

definition of TRD. Specifically, Sackeim and coauthors had reached the data-driven conclusion that "failure to respond to a single adequate antidepressant trial may convey as much information about likelihood of relapse as failure to respond to multiple trials."[12(p102)] After acknowledging that alternative definitions existed, as proposed in a publication by a European-based group (Souery et al, 1999[9]), Sackeim put forth a more decisive position on the subject in the 2001 publication, in which he stated that "treatment resistance for major depression may also more broadly be defined as the administration of an adequate dose of an antidepressant medication (or at minimal plasma levels) for sufficient

duration, with good treatment adherence and yet resulting in nonresponse or lack of remission."[11(pp11–12)] However, in a more recent article, Conway et al[35] defined TRD as failing 2 trials of antidepressants from different classes, in alignment with the definition put forth by Souery et al,[9] while also proposing a binary staging scheme with suggested treatment options for each category.

The Thase and Rush Staging Model

An influential article by Thase and Rush[13] in 1997, still heavily cited to this day, borrows the concept of illness staging used in oncology to lay out the foundations for a TRD staging model in an effort to offer guidance for the sequential treatment of patients who do not respond to a first antidepressant trial. The Thase and Rush Staging Model (TR-S) consists of 5 ascending stages of resistance to treatment, defined primarily by the type and number of antidepressant trials.

Many key points related to this scheme fall in line with current evidence. First, it accounts only for adequately delivered trials, albeit without specifying precise thresholds for dose and duration. The language was probably intentionally kept broad due to the absence of consensus in the field at the time on the optimal dose and duration of antidepressant trials, with studies showing continued increase in response and remission rates after 4, 6, and 8 weeks of treatment, therefore challenging the notion that only 4 weeks are needed for a trial to be considered adequate.[36–38] Another advantage to this model is that a failed trial of ECT carries the most weight to the degree of treatment resistance, mirroring the consistent findings, over the past several decades, of ECT's superiority over any other available treatment for depression.[39] The model, however, mentions only bilateral ECT, neglecting other modes of administration, including unilateral and bifrontal, which were shown more recently to have comparative efficacy.[40,41] Last, this scheme seems to address the problem of a ceiling effect, albeit partially: a patient will need to have failed an increasing number of trials to be considered more treatment resistant. However, the failed trials need to be of different classes, in that, for example, a patient who failed 1 selective serotonin reuptake inhibitor (SSRI) and a patient who failed 3 SSRIs would both be considered to be in stage I.

This hierarchical categorization further points to some weaknesses in the model. First, the requirement of 2 failed antidepressant trials with different classes in stage II makes the assumption that the latter condition translates into a higher degree of resistance than that of someone who fails 2 antidepressant trials from the same class. This notion no longer holds true, particularly in light of the evidence from the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial[42] demonstrating no significant difference in outcomes between a switch to a within-class or out-of-class agent after failing to remit with SSRI monotherapy. This evidence is substantiated by a meta-analysis[43] of studies comparing within versus across class switches; the number needed to treat (NNT) 1 patient if switched to a non-SSRI rather than another SSRI was 22, significantly higher than an NNT of 10 set by the United Kingdom's National Institute of Clinical Excellence. Second, the fact that one has to fail a TCA or monoamine oxidase inhibitor (MAOI) to be considered stage III or IV treatment-resistant, respectively, assumes the supremacy of these 2 medication classes over all others. Thase and Rush must have developed their hypothesis drawing on some evidence preceding their 1997 article, showing superiority of TCAs over SSRIs for melancholic depression and MAOIs over TCAs in nonresponders or those with atypical depression.[44–48] However, STAR*D results challenge those earlier findings; there was no advantage, in terms of efficacy, in switching to nortriptyline versus mirtazapine after failure of 2 consecutive antidepressant trials.[49] Likewise, no significant difference in efficacy was observed between tranylcypromine and venlafaxine plus mirtazapine following 3 failed antidepressant trials.[50] Moreover, there is no consideration for augmentation strategies in the staging model, even though the authors acknowledge the usefulness of augmenting with agents such as lithium, T_3, pindolol, buspirone, antidepressants, and neuroleptics, especially for stage III–resistant patients.[13] Finally, although an adequate course of psychotherapy, especially cognitive-behavioral therapy, has been shown to have similar efficacy to pharmacotherapy in MDD[51,52] and was recently demonstrated to be an efficacious augmentation therapy in TRD,[53] it was not considered in the Thase and Rush model or any TRD staging models to date (except that of Conway et al.[35]).

Definition of treatment resistance. Rush and colleagues[10] put forth 2 definitions without definitively arguing for one over the other. On one hand, they suggest that "modest resistance may include an inadequate response to a single antidepressant trial" while reporting on the other hand that "a general sense is that if depression has not adequately benefited from at least two adequate trials of medications from different classes in the current episode, clinically significant treatment resistance is present,"[10(p744)] citing the 2001 article by Sackeim.[11] Interestingly, as we mentioned in the previous section, Sackeim himself seemed to endorse the 1 medication cutoff definition while only acknowledging the existence of an alternative definition (≥2 trials) proposed by the European group (Souery et al, 1999[9]). In a STAR*D review article, Rush and colleagues[54] argue that a categorical definition of 2 or more trials would be the most reasonable choice based on the empirical data generated by STAR*D in which significantly lower remission rates were observed after failure of 2 treatment strategies.

The European Staging Model

In 1999, a European-based group suggested alternative criteria for the definition and staging of TRD[9] consisting of 3 distinct categories: non-responder (NR—nonresponse to 1 adequate weeks antidepressant trial of 6–8 weeks' duration), TRD (failure to respond to 2 adequate trials of different classes of antidepressants), and chronic resistant depression

(CRD—a resistant depressive episode lasting more than 1 year despite multiple adequate interventions, including augmentation strategies). The TRD category was further divided into 5 stages depending on the duration of the trials. To note, treatment response was defined as less than 50% reduction from baseline in Clinical Global Impressions scale (CGI), Hamilton Depression Rating Scale, or Montgomery-Asberg Depression Rating Scale scores.

There are many strengths to this model. The model acknowledges that failure to 1 adequate antidepressant trial falls within the treatment resistance spectrum. Furthermore, it does not differentiate between antidepressant classes with respect to their contribution to treatment resistance, which represents an improvement on the TR-S system. However, some disadvantages stand out: (1) The use of different thresholds of trial duration as indicators for TRD staging is not supported by empirical evidence and results in a paradox whereby a patient who fails 2 antidepressant trials of 1-year duration is considered more treatment-resistant than someone who has failed 5 antidepressant trials of 6–8 weeks each. (2) The model does not account for augmentation trials in TRD stages 1 to 5. (3) ECT is assumed to contribute the same level of resistance as an antidepressant (eg, SSRI) would. (4) The CRD status does not differentiate patients in terms of treatment intensity in the past year (the term used, several treatments, is vague regarding number of trials and dosage levels), thereby creating a "ceiling" effect. (5) There is no mention of psychotherapeutic treatment modalities in the model.

Definition of treatment resistance. In this article,[9] cited by many[10,11] as the source for defining TRD as 2 failed trials, the authors acknowledge that at the time, it was challenging to operationalize the definition of TRD due to the paucity of empirical data. Although they then proceed to define TRD as "failure to respond to 2 adequate trials of different classes of antidepressants," they are quick to point out that "the proposed criteria should not be considered as absolute definition of TRD but as a standardized instrument to use within a collaborative research project….The criteria of two trials allows for controlled studies, using cross-over design with different drugs."[9(p90)] This type of collaborative work was later undertaken by the European multicenter project "Patterns of Treatment Resistance and Switching Strategies in Unipolar Affective Disorder,"[55,56] the results of which we examine in a later section of this review.

The Massachusetts General Hospital Staging Method

To rigorously and systematically identify previous adequately failed trials, Fava and Davidson[8] proposed an initial version of the Antidepressant Treatment Response Questionnaire (ATRQ), the elaborated form of which has been widely used since.[8,57] The initial version of the ATRQ exclusively focused on antidepressant monotherapy and employed several key concepts, including (1) quantification of the level of improvement with treatment (less than 25%, 25%–49%, 50%–75%, and greater than 75% symptom improvement), (2) minimum duration criterion (≥ 6 weeks), and (3) minimum dose criterion tailored for each antidepressant. Later on, Fava[14] proposed an elaborated staging version of the ATRQ to address methodological challenges seen in previous models. The new model—the Massachusetts General Hospital Staging Model (MGH-S)—also accounted for (1) a distinction between an adequate and an optimal trial based on doses tailored for each antidepressant, (2) the inclusion of augmentation/combination strategies and ECT. In addition, an explicit point-based system was introduced (1 point for each adequate monotherapy, 1.5 for each optimal monotherapy, 0.5 for each augmentation/combination, and 3 for ECT), which yielded a composite score for treatment resistance.

Several advantages to this staging system should be mentioned: (1) Any failure to an adequate antidepressant trial, regardless of class, accounts for 1 point. (2) The model recognizes the importance of the antidepressant dose and duration intensity by allocating an extra 0.5 point to each trial optimization. (3) Augmentation strategies, which are supported for their efficacy in TRD by an abundance of evidence,[58,59] are also accounted for. (4) The generated composite score, which is a continuous variable, is not susceptible to the limitation of the "ceiling effect" seen in the TR-S model. (5) ECT carries the most weight on the overall resistance score.

Despite the many improvements made over previous staging methods, a few limitations should be noted. First, each augmentation strategy contributes only 0.5 point to the overall score, indirectly assuming that an antidepressant monotherapy trial is more efficacious and therefore, if failed, contributes to a higher degree of resistance than does an augmentation trial. This assumption is challenged particularly with evidence from placebo-controlled randomized controlled trials demonstrating the efficacy of augmentation with atypical antipsychotics in MDD patients who failed antidepressant monotherapy.[58,60–62] Second, 3 points are assigned to a failed ECT treatment course, without further clarification on the number of treatments and mode of administration. Further, the lack of upper limit to the total score may lead to some individuals' receiving a very large score, therefore skewing the data. Last, this model, like others, does not include psychotherapy. In fact, Fava, in his original 2003 article,[14] acknowledges that the model may be incomplete without psychotherapy being accounted for, however mentions that "since no studies have tested such assumption, future investigations need to address this issue."[14(p655)]

Definition of treatment resistance. Fava and Davidson[8] define TRD as the "failure to respond to at least one antidepressant trial of standard doses lasting 6 weeks or more." In a 2003 article, Fava[14] further rejects the definition that posits the need for 2 antidepressant trials from different classes, highlighting 2 methodological flaws: the assumption that (1) nonresponse to 2 agents of different classes is more difficult to treat than nonresponse to 2 agents of the same class and (2) switching within a class is less effective than switching to a different class.

The Maudsley Staging Method

In 2009, Fekadu et al[15] suggested the Maudsley Staging Method (MSM), an alternative staging method for TRD. The most noteworthy deviation from previously developed staging methods was the addition of illness characteristics (ie, current major depressive episode duration and severity). A score is generated to indicate the degree of resistance, with a maximum score of 15. Of note, another group[63] later added several items (functional impairment, comorbid anxiety, personality disorders, and psychosocial stressors) to the MSM to create the Dutch measure for quantification of treatment resistance in depression.

The MSM succeeds in applying equal weight to different antidepressant classes and, in congruence with contemporary empirical evidence, does not differentiate between within- and across-class switches. However, its major pitfall lies in the insertion of naturalistic illness factors for the assessment of treatment resistance. Characteristics such as episode duration and severity may be somewhat helpful in assessing disease progression and prognosis. However, they convey no information regarding the degree of treatment resistance in the current episode. To illustrate this point, a pharmacotherapy-naive individual with a moderate major depressive episode (MDE) lasting more than a year would qualify as treatment resistant according to this staging method despite not having tried any treatment yet. Moreover, an individual with a mild MDE of less than a year's duration who fails 2 antidepressant trials would score less (and therefore be considered less treatment resistant) than an individual who has a psychotic MDE but with good response to a first antidepressant trial. This again creates a clinical paradox.

Definition of treatment resistance. In accordance with the MGH-S definition, Fekadu et al[15] propose a treatment resistance definition of at least 1 antidepressant trial failure by rejecting the notion that treatment resistance is an "all-or-nothing phenomenon" and instead postulating that "it exists as a continuum," further elaborating that "failure of the first treatment is influential in treatment resistance and may be a useful starting point in any measure of this conceptual continuum."[15(p179)]

EMPIRICAL EVIDENCE

Empirical Evidence Supporting the Staging Models

There remains a paucity of empirical evidence to support the models described in the Results section of this article. However, a few studies evaluated the validity and predictive utility of some of these schemes. Petersen et al[64] tested the predictive power of the MGH-S versus the TR-S with respect to remission with subsequent treatments. After reviewing 115 charts of outpatients with MDD, the authors found a high correlation between the scores of the 2 models; however, the MGH-S had a significantly greater predictive value with respect to nonremission (CGI-Improvement score > 1). Further, retrospective evidence from charts of 88 discharged subjects with TRD showed that both the MSM and the TR-S, albeit the latter to a lesser extent, had a positive linear association with the likelihood of future nonremission (at discharge), with the MSM correctly predicting treatment resistance in 85.5% of cases.[15] Charts from 62 TRD subjects from the same cohort were examined for different clinical variables after their discharge for an average period of 29.5 months. The MSM score, but not the TR-S score, was found to be positively associated with a persistent depressive episode throughout the follow-up period, depressive episodes lasting 50% or longer of the follow-up period, months spent in a depressive episode, and functional impairment.[65] In another study[66] examining the long-term outcomes and predictors of TRD in 118 subjects with unipolar and bipolar depression in tertiary care, the MSM score was only marginally associated with achieving remission in follow-up (adjusted hazard ratio = 0.82; 95% CI, 0.68–0.99; $P = .04$). In a recent study,[67] MSM score was found to be correlated with percentage of time spent in a depressive episode in 643 subjects with MDD followed for 2 years. However, a poor retrospective account of treatment trial adequacy coupled with a naturalistic description of the course of illness irrespective of treatment limited the authors' ability to draw conclusions regarding treatment resistance.

Empirical Data Informing the Definition of TRD

Reviewing the main articles that have focused on this work, we find that the greatest point of discrepancy is whether a dichotomous or continuous model better fits the empirical data from these studies. In particular, the discrepancy appears to focus on whether or not a single antidepressant trial would qualify as resistance, or whether TRD defined as 2 or more treatment failures represents a more homogeneous clinical entity. One way to answer this question is to examine the outcome of (*a*) patients who are treatment-naive, (*b*) patients who fail exactly 1 antidepressant trial, and (*c*) patients who fail more than 1 trial. To date, data from 2 large trials that examined more than 1 sequential treatment approach can be utilized to answer this question.

We first examine data from STAR*D, a large multicenter US based trial to evaluate, in a sequential manner, the effectiveness of depression treatments in patients with MDD after failing their first antidepressant trial.[68] We can note that treatment-naive patients treated with citalopram monotherapy exhibited a 48.6% response rate. However, those who failed to respond to citalopram and were subsequently treated with venlafaxine, sertraline, bupropion, or cognitive-behavioral therapy had much lower response rates, namely between 26% and 30% (average = 28.5%). Regarding long-term outcomes, remitters following a single antidepressant trial demonstrated a significantly lower relapse rate (33.5%) than remitters following 2 or more antidepressant trials ($P < .0001$), while no statistically significant difference was observed between the relapse rates of any of the other 3 groups (relapse rates of 47.4%, 42.9%, and 50% for remitters after 2, 3, or 4 trials, respectively). Taken together, these results indicate that treatment-naive patients and patients who fail a single treatment are very dissimilar in short- and

long-term outcomes. Instead, there appears to be a more gradual transition to poorer outcomes as the number of treatment failures increases, with the greatest difference in outcome between treatment-naive subjects and subjects who failed exactly 1 treatment.

A similar picture emerges when examining the results of the European multicenter project designed to help answer important questions regarding TRD, including those concerning predictors of resistance and optimal treatment strategies.[56] Subjects who had failed an adequate antidepressant trial were recruited and randomized to a 4-week trial of citalopram versus desipramine. Subsequently, nonresponders to this first phase of the study were further randomized to continue the same treatment or switch to the alternate treatment for another 4 weeks, resulting in 4 arms (citalopram-citalopram, desipramine-desipramine, citalopram-desipramine, and desipramine-citalopram). Response rates to phase 1 were 54% in the citalopram group and 55% in the desipramine group. Much lower response rates to subsequent treatments would lend support to the notion that TRD is more homogeneously defined as 2 (rather than 1) or more trial failures. However, response rates from the 4 arms ranged between 38% and 67%, hence overlapping with response rates from the first phase. Therefore, outcomes of patients who failed 1 versus 2 treatments seem more similar than not, reinforcing the notion that a cutoff at 2 failed antidepressant trials is arbitrary and not supported by empirical evidence.

DISCUSSION AND FUTURE DIRECTIONS

A number of methods are currently in use for staging the treatment of MDD. Each of these has its strengths and limitations and helps inform critical ingredients necessary for any staging schema in TRD. We summarize them as follows:

1. Most antidepressant trials require a minimum duration and dosage to work for the average patient. These should be clearly defined from the onset.
2. A staging mechanism should exclusively measure the strength of treatment during the current episode and thus differentiate true resistance from transient improvement ("poop-out"), illness relapse/recurrence, or lifetime treatment intensity (ie, previous episodes). However, physicians should not overlook past history of treatment failures, as these may inform future treatment choice, even though they do not strictly count toward treatment resistance during the current episode.
3. There should be a differentiation between minimally effective treatment and optimal treatment (with respect to dosage) since the majority of traditional antidepressants appear to provide additional benefits at higher doses.
4. Augmentation therapies, commonly used in clinical practice, should be reflected when calculating treatment strength, as should psychotherapy and other evidence-based brain stimulation therapies.
5. Due to their greater acute efficacy, treatment with ECT or ketamine should be ascribed higher scores.
6. Staging should consist of a method for measuring treatment strength, rather than incorporating non–treatment-related variables believed to serve as predictors of treatment outcome.

Of the staging methods reviewed, the MGH-S captures most of these criteria. Future modifications should focus on incorporating novel treatments such as ketamine and other rapidly acting drugs, with attention given to defining minimal qualifying dose and duration,[69] and an appropriate strength score for these. There should also be a distinction between adjunctive therapies with empirical evidence similar to that of antidepressant monotherapies (ie, atypical antipsychotics) and those with lesser evidence (and hence a lower attributed score). In addition, combination treatments (2 antidepressants), when adequate, should also receive scores reflecting the use of 2 antidepressant therapies.[70]

Finally, one major issue in ascertaining TRD status is recall bias. Independent interviews performed by trained clinicians with experience in the conduct of clinical trials, and validated scales measuring history of treatment intensity during the current episode, can help mitigate the risk of enrollment of inappropriate study subjects due to recall bias. In parallel, the provider's and electronic health records, prescription records, and pharmacy records can help partly mitigate this problem, particularly in the era of mandated electronic medical records and measurement-based care in clinical practice, which includes patient-rated scales such as the Quick Inventory of Depressive Symptomatology, the Symptoms of Depression Questionnaire, or the 9-item Patient Health Questionnaire.[7]

With respect to the definition of TRD, Souery and colleagues[9] argued for excluding patients with 1 treatment failure from this definition. However, the data from the 2 largest trials, as shown earlier in this article, do not fit this hypothesis. In our point of view, based on these data, failure of a single adequate treatment should qualify as TRD. An issue equally important as how scientists define TRD involves how the regulatory process views TRD. To date, phase III studies designed by sponsors to evaluate the efficacy, safety, and tolerability of potential treatments for TRD define the latter as 2 or more unsuccessful trials of antidepressant monotherapy during the current episode. In addition, TRD phase III programs to date (olanzapine-fluoxetine combination[71,72] and, more recently, intranasal esketamine[73–75]) randomized patients who were nonresponders to a new antidepressant plus the investigational agent versus a new antidepressant plus placebo. Data generated by such trials are limited in 2 ways with respect to common clinical practice. First, only antidepressant monotherapy counts toward meeting criteria for TRD, resulting in patients who have failed 1 monotherapy plus numerous adjunctive treatments (eg, pharmacotherapy, psychotherapy, ECT) not formally meeting criteria

(although the esketamine phase III program did make an exception for adjunctive aripiprazole or quetiapine[73–75]). Such patients would be excluded from TRD trials using the existing regulatory definition. In addition, in the face of partial response or even nonresponse, many clinicians resort to adding a new therapy or switching to another treatment. Substituting a treatment with 2 new ones, as in the phase III programs mentioned above, does not represent augmentation or switching and is extremely rare in clinical care. To complicate matters further, the new draft guidance by the US Food and Drug Administration on the matter, published in June 2018,[76(p6)] requires all programs seeking indications for TRD labeling to randomize patients who have failed 2 or more antidepressant trials to the investigational product (as monotherapy) or continue the antidepressant to which they had failed to respond (switch). Therefore, this new language removes the requirement of switching all TRD patients to a new antidepressant when testing the efficacy of an investigational drug, which is an improvement. However, it also removes the option of seeking an adjunctive therapy indication for TRD labeling. Given that (1) we do not know whether augmentation versus switching is preferable in TRD (ASCERTAINTRD[77] is examining this question) and (2) adjunctive treatments are currently more popular in TRD than monotherapies,[78] it is unclear how this new guidance helps clinicians, researchers, or patients. Our recommendation would be to simply designate adjunctive or monotherapy indications for TRD. Additional indications could be given for patients who specifically fail to respond to therapies proven more efficacious (intravenous ketamine, ECT, intravenous scopolamine). Meanwhile, clinicians should adapt data from phase III TRD studies to best suit the needs of patients in real-world treatment settings. Such adaptation includes considering these treatment options for monotherapy or adjunctive treatment for MDD patients who may not meet strict regulatory definitions of TRD or typical entry criteria for phase III studies.

LIMITATIONS

This review has several limitations that should be noted. First, we found only a few validation studies for some but not all staging models presented, limiting our ability to decisively assess the utility of different staging systems and calls for more emphasis on empirical research in this area. Second, we acknowledge that other experts in the field may interpret the data extracted from our search and presented in this review differently. A recent article,[79] for instance, also provides an independent and unbiased review of staging methods in TRD accompanied by the authors' impression of their relative strengths and merits. Unlike ours, which focuses on a detailed examination of the methods measuring treatment resistance, that review also encompasses a discussion on the prevalence and possible clinical and demographic risk factors for TRD. We therefore encourage the reader interested in furthering their knowledge on this topic to seek additional resources in the literature, learn about the different viewpoints, and drive conclusions accordingly. Finally, although we did a thorough and systematized search of the literature, important works might have been inadvertently omitted from inclusion in the present review, particularly new or unpublished works.

Submitted: March 16, 2018; accepted February 4, 2019.

Published online: June 4, 2019.

Potential conflicts of interest: **Dr Salloum** reports no financial relationship relevant to the subject of this article. As part of his current fellowship training, he also works in the digital medicine group at Pfizer Inc; however, that work has no relevance or conflict of interest with the subject of this review article.

Dr Papakostas has served as a consultant for Abbott, Acadia,* Alkermes, AstraZeneca, Avanir, Axsome Therapeutics,* Boston Pharmaceuticals, Brainsway, Bristol-Myers Squibb, Cephalon, Dey, Eli Lilly, Genentech,* Genomind,* GlaxoSmithKline, Evotec, H. Lundbeck, Inflabloc, Janssen Global Services,* Jazz, Johnson & Johnson,* Methylation Sciences, Mylan,* Novartis Pharma, One Carbon Therapeutics,* Osmotica,* Otsuka, PAMLAB, Pfizer, Pierre Fabre, Ridge Diagnostics (formerly known as Precision Human Biolaboratories), Shire, Sunovion, Taisho, Takeda, Theracos, and Wyeth. He has received honoraria (for lectures or consultancy) from Abbott, Acadia, Alkermes, Asofarma America Centroamérica y Caribe, AstraZeneca, Avanir, Bristol-Myers Squibb, Brainsway, Cephalon, Dey, Eli Lilly, Evotec, Forest, GlaxoSmithKline, Inflabloc, Grunbiotics, Jazz, H. Lundbeck, Medichem, Meiji Seika, Novartis, Otsuka, PAMLAB, Pfizer, Pharma Trade SAS, Pierre Fabre, Ridge Diagnostics (formerly known as Precision Human Biolaboratories, Shire, Sunovion, Takeda, Theracos, Titan, and Wyeth. He has received research support (paid to hospital) from AstraZeneca, Bristol-Myers Squibb, Forest, the National Institute of Mental Health, Neuralstem*, PAMLAB, Pfizer, Ridge Diagnostics (formerly known as Precision Human Biolaboratories), Sunovion, Tal Medical, and Theracos. He has also served (not currently) on the speaker's bureau for Bristol-Myers Squibb and Pfizer.

*Asterisk denotes activity undertaken on behalf of Massachusetts General Hospital.

Funding/support: None.

REFERENCES

1. Greden JF. *J Clin Psychiatry*. 2001;62(suppl 16):26–31.
2. Gaspersz R, Lamers F, Kent JM, et al. *J Clin Psychiatry*. 2017;78(2):207–213.
3. Russell JM, Hawkins K, Ozminkowski RJ, et al. *J Clin Psychiatry*. 2004;65(3):341–347.
4. Kautzky A, Dold M, Bartova L, et al. *J Clin Psychiatry*. 2018;79(1):16m11385.
5. Kelly D. *Pharmakopsychiatr Neuropsychopharmakol*. 1974;7(4):199–204.
6. Nierenberg AA, Katz J, Fava M. *Psychiatr Clin North Am*. 2007;30(1):13–29.
7. Trivedi MH, Rush AJ, Wisniewski SR, et al; STAR*D Study Team. *Am J Psychiatry*. 2006;163(1):28–40.
8. Fava M, Davidson KG. *Psychiatr Clin North Am*. 1996;19(2):179–200.
9. Souery D, Amsterdam J, de Montigny C, et al. *Eur Neuropsychopharmacol*. 1999;9(1–2):83–91.
10. Rush AJ, Thase ME, Dubé S. *Biol Psychiatry*. 2003;53(8):743–753.
11. Sackeim HA. *J Clin Psychiatry*. 2001;62(suppl 16):10–17.
12. Sackeim HA, Prudic J, Devanand DP, et al. *J Clin Psychopharmacol*. 1990;10(2):96–104.
13. Thase ME, Rush AJ. *J Clin Psychiatry*. 1997;58(suppl 13):23–29.
14. Fava M. *Biol Psychiatry*. 2003;53(8):649–659.
15. Fekadu A, Wooderson S, Donaldson C, et al. *J Clin Psychiatry*. 2009;70(2):177–184.
16. Berlim MT, Turecki G. *Eur Neuropsychopharmacol*. 2007;17(11):696–707.
17. Cepeda MS, Reps J, Fife D, et al. *Depress Anxiety*. 2018;35(3):220–228.
18. Keller MB, Lavori PW, Klerman GL, et al. *Arch Gen Psychiatry*. 1986;43(5):458–466.
19. Baldessarini RJ. *Chemotherapy in Psychiatry*. Cambridge, MA: Harvard University Press; 1977:87–98.
20. Klein DR, David JM. *Diagnosis and Drug Treatment of Psychiatric Disorders*. Baltimore, MD: Williams & Wilkins Co; 1969:189–192.
21. Cole JD, David JM. Antidepressant drugs. In: Freedman AM, Kaplan HI, Sadock BJ, eds. *The Comprehensive Textbook of Psychiatry*. Baltimore, MD: Williams & Wilkins Co; 1980:1940–1948.
22. Appleton WS. *J Clin Psychiatry*. 1982;43(1):12–27.
23. Quitkin FM, Rabkin JG, Ross D, et al. *Arch Gen Psychiatry*. 1984;41(8):782–786.
24. Undurraga J, Baldessarini RJ. *Neuropsychopharmacology*. 2012;37(4):851–864.
25. Papakostas GI, Homberger CH, Fava M. *J Psychopharmacol*. 2008;22(8):843–848.
26. Papakostas GI, Nelson JC, Kasper S, et al. *Eur Neuropsychopharmacol*. 2008;18(2):122–127.

27. Papakostas GI, Fava M. *Eur Neuropsychopharmacol*. 2007;17(1):32–36.
28. Papakostas GI, Fava M. *Eur Psychiatry*. 2007;22(7):444–447.
29. Papakostas GI, Fava M. *Can J Psychiatry*. 2006;51(12):783–790.
30. Song F, Freemantle N, Sheldon TA, et al. *BMJ*. 1993;306(6879):683–687.
31. Itil TM, Shrivastava RK, Mukherjee S, et al. *Br J Clin Pharmacol*. 1983;15(suppl 3):433S–438S.
32. White K, Razani J, Cadow B, et al. *Psychopharmacology (Berl)*. 1984;82(3):258–262.
33. Georgotas A, McCue RE, Hapworth W, et al. *Biol Psychiatry*. 1986;21(12):1155–1166.
34. Bollini P, Pampallona S, Tibaldi G, et al. *Br J Psychiatry*. 1999;174(4):297–303.
35. Conway CR, George MS, Sackeim HA. *JAMA Psychiatry*. 2017;74(1):9–10.
36. Feiger AD, Bielski RJ, Bremner J, et al. *Int Clin Psychopharmacol*. 1999;14(1):19–28.
37. Trivedi MH, Rush AJ, Pan JY, et al. *J Clin Psychiatry*. 2001;62(3):158–163.
38. Rush AJ, Batey SR, Donahue RM, et al. *J Affect Disord*. 2001;64(1):81–87.
39. Group UER; UK ECT Review Group. *Lancet*. 2003;361(9360):799–808.
40. Semkovska M, Landau S, Dunne R, et al. *Am J Psychiatry*. 2016;173(4):408–417.
41. Dunne RA, McLoughlin DM. *World J Biol Psychiatry*. 2012;13(4):248–258.
42. Rush AJ, Trivedi MH, Wisniewski SR, et al; STAR*D Study Team. *N Engl J Med*. 2006;354(12):1231–1242.
43. Papakostas GI, Fava M, Thase ME. *Biol Psychiatry*. 2008;63(7):699–704.
44. Danish University Antidepressant Group. *Psychopharmacology (Berl)*. 1986;90(1):131–138.
45. Danish University Antidepressant Group. *J Affect Disord*. 1990;18(4):289–299.
46. McGrath PJ, Stewart JW, Nunes EV, et al. *Am J Psychiatry*. 1993;150(1):118–123.
47. Quitkin FM, Harrison W, Stewart JW, et al. *Arch Gen Psychiatry*. 1991;48(4):319–323.
48. Quitkin FM, McGrath PJ, Stewart JW, et al. *Arch Gen Psychiatry*. 1990;47(10):935–941.
49. Fava M, Rush AJ, Wisniewski SR, et al. *Am J Psychiatry*. 2006;163(7):1161–1172.
50. McGrath PJ, Stewart JW, Fava M, et al. *Am J Psychiatry*. 2006;163(9):1531–1541, quiz 1666.
51. Thase ME, Friedman ES, Biggs MM, et al. *Am J Psychiatry*. 2007;164(5):739–752.
52. DeRubeis RJ, Hollon SD, Amsterdam JD, et al. *Arch Gen Psychiatry*. 2005;62(4):409–416.
53. Nakagawa A, Mitsuda D, Sado M, et al. *J Clin Psychiatry*. 2017;78(8):1126–1135.
54. Rush AJ, Warden D, Wisniewski SR, et al. *CNS Drugs*. 2009;23(8):627–647.
55. Souery D, Serretti A, Calati R, et al. *J Clin Psychopharmacol*. 2011;31(4):512–516.
56. Souery D, Serretti A, Calati R, et al. *World J Biol Psychiatry*. 2011;12(5):364–375.
57. Chandler GM, Iosifescu DV, Pollack MH, et al. *CNS Neurosci Ther*. 2010;16(5):322–325.
58. Nelson JC, Papakostas GI. *Am J Psychiatry*. 2009;166(9):980–991.
59. Papakostas GI. *J Clin Psychiatry*. 2009;70(suppl 6):16–25.
60. Berman RM, Fava M, Thase ME, et al. *CNS Spectr*. 2009;14(4):197–206.
61. Marcus RN, McQuade RD, Carson WH, et al. *J Clin Psychopharmacol*. 2008;28(2):156–165.
62. Berman RM, Marcus RN, Swanink R, et al. *J Clin Psychiatry*. 2007;68(6):843–853.
63. Peeters FP, Ruhe HG, Wichers M, et al. *J Affect Disord*. 2016;205:365–371.
64. Petersen T, Papakostas GI, Posternak MA, et al. *J Clin Psychopharmacol*. 2005;25(4):336–341.
65. Fekadu A, Wooderson SC, Markopoulou K, et al. *J Clin Psychiatry*. 2009;70(7):952–957.
66. Fekadu A, Rane LJ, Wooderson SC, et al. *Br J Psychiatry*. 2012;201(5):369–375.
67. van Belkum SM, Geugies H, Lysen TS, et al. *J Clin Psychiatry*. 2018;79(1):17m11475.
68. Rush AJ, Trivedi MH, Wisniewski SR, et al. *Am J Psychiatry*. 2006;163(11):1905–1917.
69. Papakostas GI, Charles D, Fava M. *World J Biol Psychiatry*. 2010;11(2 pt 2):300–307.
70. Blier P, Ward HE, Tremblay P, et al. *Am J Psychiatry*. 2010;167(3):281–288.
71. Shelton RC, Williamson DJ, Corya SA, et al. *J Clin Psychiatry*. 2005;66(10):1289–1297.
72. Corya SA, Williamson D, Sanger TM, et al. *Depress Anxiety*. 2006;23(6):364–372.
73. A Study to Evaluate the Efficacy S, and Tolerability of Flexible Doses of Intranasal Esketamine Plus an Oral Antidepressant in Adult Participants With Treatment-resistant Depression. Identifier NCT02418585. ClinicalTrails.gov website. https://clinicaltrials.gov/ct2/show/NCT02418585?term=esketamine&draw=3&rank=30.
74. Janssen Research & Development. Randomized, double-blind study of flexibly-dosed intranasal esketamine plus oral antidepressant vs active control in treatment-resistant depression. Poster session presented at the 2018 Annual Meeting of American Society of Clinical Psychopharmacology. June 2018; Miami, Florida.
75. Janssen Research & Development. Efficacy and safety of esketamine nasal spray plus an oral antidepressant in elderly patients with treatment-resistant depression. Poster session presented at the 2018 Annual Meeting of American Society of Clinical Psychopharmacology. June 2018; Miami, Florida.
76. Food and Drug Administration. Major Depressive Disorder: Developing Drugs for Treatment, Guidance for Industry. FDA website. https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM611259.pdf. June 2018.
77. Comparative effectiveness research trial for antidepressant incomplete and non-responders with TRD (ASCERTAINTRD). Identifier NCT02977299. ClinicalTrials.gov website. https://clinicaltrials.gov/ct2/show/NCT02977299. In.
78. Goldberg JF, Freeman MP, Balon R, et al. *Depress Anxiety*. 2015;32(8):605–613.
79. Trevino K, McClintock SM, McDonald Fischer N, et al. *Ann Clin Psychiatry*. 2014;26(3):222–232.

Editor's Note: We encourage authors to submit papers for consideration as a part of our Early Career Psychiatrists section. Please contact Erika F. H. Saunders, MD, at esaunders@psychiatrist.com.