Massachusetts General Hospital SAFER Criteria for Clinical Trials and Research

Martin Desseilles, MD, PhD, Janet Witte, MD, MPH, Trina E. Chang, MD, MPH, Nadia Iovieno, MD, PhD, Christina Dording, MD, Heidi Ashih, MD, PhD, Maren Nyer, PhD, Marlene P. Freeman, MD, Maurizio Fava, MD, and David Mischoulon, MD, PhD

Major depressive disorder (MDD) is considered a major public health issue related to considerable morbidity and disability.[1,2] In addition, about half of patients with MDD will respond either partially or not at all to first-line antidepressant monotherapy, despite an adequate dosage and a sufficient duration of treatment.[3] For these patients, there is a clear need to study next-step antidepressant therapies, and reliable efficacy data are required in order to produce meaningful statistical analysis and increase the likelihood of finding effective, clinically applicable treatments.

We describe here a new instrument, the clinician-rated Massachusetts General Hospital *SAFER* interview. The acronym stands for interview's attention to the following criteria: State versus trait; Assessability; Face validity; Ecological validity; and Rule of three Ps (pervasive, persistent, and pathological). These criteria facilitate the assessment of important supplemental factors that go well beyond conventional diagnostic criteria and symptom severity to identify patients who would be valid for clinical trials. The use of SAFER might increase the quality of clinical trials by separating placebo responders from active drug responders and by restricting the sample to a more homogenously defined MDD population. That is, SAFER can help to ensure that patients are enrolled in studies that will allow assessment of the treatment modality's efficacy by minimizing the risk that factors unrelated to treatment will determine the patient's course of illness.

From the Cyclotron Research Centre, University of Liège, Belgium (Dr. Desseilles); Clinical Trials Network & Institute (Drs. Desseilles, Witte, Chang, Iovieno, Dording, Ashih, Nyer, Freeman, Fava, and Mischoulon) and Depression Clinical and Research Program (Drs. Desseilles, Witte, Chang, Iovieno, Dording, Ashih, Nyer, Fava, and Mischoulon), Department of Psychiatry, Massachusetts General Hospital, Boston, MA; Harvard Medical School (Drs. Witte, Chang, Iovieno, Dording, Ashih, Nyer, Freeman, Fava, and Mischoulon).

Correspondence: Martin Desseilles, MD, PhD, Cyclotron Research Centre, University of Liège B30, 8 Allée du 6 Août, B-4000 LIEGE, Belgium. Email: m.desseilles@ulg.ac.be

THE NEED FOR A NEW INSTRUMENT

Reliance on a categorical system of psychiatric nosology, such as the currently employed rigorous classification system of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM),[4] may not be sufficient for contemporary antidepressant drug development.[5] First, the current DSM-5 MDD criteria allow for the inclusion of patients with heterogeneous symptoms representing several depressive syndromes/subtypes. Second, a large number of symptoms commonly reported by depressed patients are not captured by the DSM-5 (e.g., somatic symptoms, such as headache, abdominal pain, musculoskeletal pain in the lower back, joints, and neck).[6,7] Third, situational factors (e.g., comorbid conditions, concurrent medications, or external circumstances) may contribute to episodes that appear similar to MDD but that fluctuate widely because of external psychosocial changes rather than because of treatment or spontaneous remission/exacerbation of the depressive illness. While DSM-5 now includes bereavement under the MDD diagnostic umbrella, other situational factors may be contributing to depressive symptoms to such an extent that including such a patient in a clinical trial for the treatment of depression may be inappropriate.*

Used in conjunction with established structured diagnostic instruments and symptom severity scales, SAFER overcomes many of these obstacles. It is not restricted to the DSM criteria, and it allows for other symptoms observed during the clinical interview, as well as situational factors, to be considered in evaluating eligibility for clinical trials. It also allows patients to be ruled out in cases where the symptomatology may not be strictly secondary or specific to MDD.

THE OPERATIONALIZATION OF SAFER CRITERIA

We have sought to use SAFER to refine the diagnosis of depression by assessing the *persistence*, *pervasiveness*, and *pathological* nature of the current depressive episode.[5] The SAFER interview aims to confirm the diagnosis and exclude

---

* The reader should note that, although this manuscript was originally written based on our experience with DSM-IV MDD, the issues of concern remain more or less the same as we transition to DSM-5.
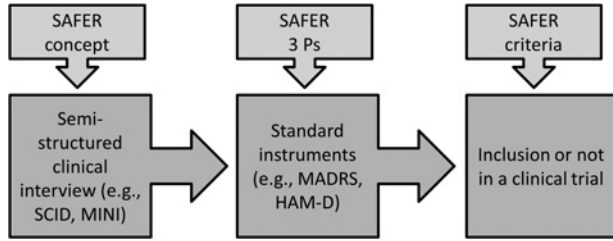
Figure 1. The SAFER is used at three steps of the clinical interview for a clinical trial. First, the SAFER (Figure 2) is used in addition to structured or semi-structured clinical interviews (e.g., the Structured Clinical Interview for DSM-IV [SCID] or Mini–International Neuropsychiatric Interview [MINI]). Second, the rule of the three Ps (i.e., Pervasiveness, Persistence, and Pathological, see Figure 2) is used for each depressive symptom revealed by the clinical interview, with or without the use of standard instruments for assessing depression severity (e.g., the Montgomery–Åsberg Depression Rating Scale (MADRS) or Hamilton Rating Scale for Depression (HAM-D)). Third, the SAFER criteria (Text Box 1) are filled out, and the patient is included or excluded from the clinical trial.

any depression that appears to be a trait, not a state, and any symptoms that are nonspecific or not readily assessable.[5] The overall concept is depicted in Figure 1.

The principles of the SAFER criteria and a description of the 3Ps –persistence, pervasiveness, and pathology– are presented in Figure 2.

Assessability means that the patient should be able to answer the clinician's questions—both from the depression scale and the general interview—in a relatively clear and unambiguous manner. Patients who, for example, provide many qualifiers to each answer, who provide data that are contradictory or inconsistent, or who seem to have difficulty understanding the questions would not likely yield useful data in a clinical trial, as it is difficult to determine whether they are actually in a major depressive episode (MDE), how severe their symptoms are, and whether those symptoms change over the course of the treatment period.

Face validity refers to the patients' symptoms clearly mapping to those of DSM-5 MDD—namely, daily depressed mood, anhedonia, sleep disturbance, and so on.

Ecological validity means that the symptoms should present the way that we know them to present in classic MDD. For example, a patient with MDD is expected to be depressed most of the day nearly every day. A patient who is depressed about half of all days, and who may have frequent periods of several weeks of euthymia alternating with periods of depression, does not seem like the usual MDD patient, and also would not be a good candidate for a study, as his symptoms would be too volatile to allow assessment of the efficacy of a new treatment.

Operationalization of the SAFER involves the use of focused questions around the 3Ps: pervasiveness, persistence, and pathology. Historically, we operationalized the SAFER criteria by inquiring about the 3P's for every symptom of the depression scale used in the study—for example, the Hamilton Depression scale. More recently, we have found that focusing the SAFER questions on the DSM symptoms of depression that the patient endorses, either individually or in a global manner, is more efficient and less cumbersome for the clinician and the patient. Each of the instrument's 9 items are applied globally: does the overall presentation

Figure 2. The SAFER criteria. The Rule of the Three Ps (for Pervasiveness, Persistence and Pathological) is used for depressive symptoms (e.g., depressed mood, fatigue, guilt, and sleep disturbances). The clinician asks about these characteristics for EACH item on the basic depression questionnaire and for any symptoms that may appear during the clinical interview.

Text Box 1
Operationalization of the SAFER Criteria
SAFER Criteria Inventory
(© Massachusetts General Hospital)

1. Persistent Symptoms:
   □ Definitely (All of the patient's symptoms are present most of the day, nearly every day)
   □ Possibly (Some, but not all, symptoms are present most of the day, nearly every day)
   □ Unlikely (The majority of the symptoms are not present most of the day, nearly every day)

2. Pervasive Symptoms:
   □ Definitely (Symptoms impact all domains and/or contexts)
   □ Possibly (Symptoms impact many domains and/or contexts, but not all)
   □ Unlikely (Symptoms impact a minority of domains and/or contexts)

3. Pathological Symptoms:
   □ Definitely (Symptoms are disruptive and have affected behavior or function, and are distinguishable from the patient's normal functioning)
   □ Possibly (Symptoms are sometimes disruptive of behavior or function, or are not reliably distinguishable from the patient's normal functioning)
   □ Unlikely (Symptoms are not disruptive and have not affected the patient's behavior or function, and are not distinguishable from the patient's normal functioning)

4. State (not Trait) Symptoms:
   □ Definitely (Patient can remember a time when he or she felt well.)
   □ Possibly (Patient may remember a time when he or she felt better, though the depression appears to be an exacerbation of previous dysthymia or chronic depression)
   □ Unlikely (The patient is chronically dysthymic or depressed with no identifiable reference point for a well period.)

5. Acute Symptoms:
   □ Definitely (Patient has symptoms of depression that began or worsened during the current episode)
   □ Possibly (Symptoms of depression do not appear to have begun with the current episode)
   □ Unlikely (Patient does not have symptoms of depression)

6. Specificity of Symptoms:

   NOTE: A situational depression may improve or worsen in response to external circumstances. Thus, symptomatic change (either improvement or worsening) may not be reliably attributed to drug treatment.

   □ Definitely (Patient has MDD does not have any other condition as the primary cause of these symptoms)

   □ Possibly (There is significant doubt as to whether MDD is the primary diagnosis, because patient appears to have a symptomatic condition, e.g., PTSD or GAD, that may be responsible for many or most of the symptoms)
   □ Unlikely (Patient's symptoms are likely caused by something other than MDD: another medical or psychiatric diagnosis, concomitant medications, alcohol or drug abuse, external circumstances, or, the depressive episode seems highly situational and the patient may spontaneously improve with changed circumstances)

7. Valid Symptoms (Ecological and Face Validity):
   Using the responses obtained from the severity scale and SAFER questions 1–6, ascertain whether:
   a) Symptoms clearly map to the primary nosological entity and occur with the frequency, intensity, duration, course, and impact consistent with our knowledge of DSM-diagnosed MDD in a real-world setting. Symptoms do not map more closely to any other condition which might be considered a primary diagnosis, e.g. PTSD.
   b) Symptoms are not exaggerated, they represent a change from baseline, and they have had real impact on behavior or level of function over at least the past 4 weeks.
   c) Symptomatic change is likely to matter to the patient's quality of life.
   d) If recurrent, the characteristics of the current episode are similar to previous episodes. (Ask the patient this question if necessary.)
   □ Definitely
   □ Possibly
   □ Unlikely

8. Assessable Symptoms:
   □ Definitely (Patient answers questions with reasonable certainty, there is good internal consistency between answers to items, and the patient rarely stumbles, seems unsure, or contradicts himself/herself)
   □ Unlikely (Patient stumbles on answers, answers in non sequiturs or digresses a lot, doesn't seem sure of answers, and/or contradicts herself/himself in answering questions)

   The patient's symptoms must be able to be reliably and sensitively measured with the employed rating instruments.

9. Valid Patient (meets ALL SAFER criteria for this clinical trial):
   □ Yes (SAFER questions 1–4 must be scored as "Definitely" or "Possibly," and SAFER questions 5–8 must be scored as "Definitely")
   □ No (The above conditions for "Yes" are not met)

   A response of "Unlikely" to any of questions 1–8 means that the patient will be a SAFER fail.

suggest MDD or something else? As for the time frame, SAFER is designed to focus on the entire period of depression, as described by the patient. However, if the depression is long-standing or has had a fluctuating course, we instruct clinicians to focus on the past month, to facilitate the data-gathering process.

The operationalization of the SAFER criteria is depicted in Text Box 1. A score of 1 means that the answer to the question posed in the item is "definitely," meaning, for example, that the symptoms have been present during the acute episode (item 5). A score of 2 means the answer is "possibly," meaning that the answer is less clear but is, overall, thought to be affirmative; for example, a case of a depression that began shortly after a job loss might cast doubt on item 6 (specificity of symptoms) and would earn a score of 2. A score of 3 means that the answer to the question of interest is "unlikely"; for example, a patient who cannot give clear answers to the clinician would get a rating of 3 on the item 8 (assessability).

For items 1 to 8 in the text box (Operationalization of the SAFER Criteria), the rater will select the score 1 (Definitely), 2 (Possibly) or 3 (Unlikely). Then, at item 9, the rater will assess the overall validity of the patient as a study candidate. This requires that the first four symptom criteria (the 3Ps, and state vs. trait) should be scored Definitely or Possibly (score 1 or 2), and items 5 (Acute Symptoms), 6 (Specificity of Symptoms), 7 (Valid Symptoms [Ecological and Face Validity]), and 8 (Assessable Symptoms) have to be scored Definitely (score 1). A yes on item 9 (Valid patient) would represent a SAFER "pass" if the patient has sufficient and valid target symptoms of the primary nosological entity, and for at least four weeks, to yield meaningful, measurable scores in a clinical trial. A SAFER "fail" will occur when the symptoms are not present (item 5 is Possibly or Unlikely), are nonspecific to depression (item 6 is Possibly or Unlikely), have no ecological or face validity (item 7 is Possibly or Unlikely), or the patient is not assessable (item 8 is Possibly or Unlikely)

THE POTENTIAL ADVANTAGES OF SAFER

Nuanced Approach to Depression and Pathophysiological Processes

SAFER is not limited to a symptom list; instead, it takes into account a variety of symptoms in depressed patients that can be psychological or somatic[8] and that may be based on underlying pathophysiological processes. This more nuanced approach is important because,

> since the publication of DSM-III in 1980, there has been a steady decline in the teaching of careful clinical evaluation that is targeted to the individual person's problems and social context and that is enriched by a good general knowledge of

psychopathology. Students are taught to memorize DSM rather than to learn complexities from the great psychopathologists of the past.[9(p108)]

Consideration of the complexity of a person's symptom constellation (i.e., a phenomenological approach) might be reinstituted by the systematic use of SAFER in clinical trials. SAFER inquires about possible causes of depression that are not attributable to MDD per se—for example, depressive symptoms secondary to medical illness (e.g., hypothyroidism), concurrent medication side effects (e.g., from interferon), and comorbid psychopathology (e.g., posttraumatic stress disorder). By properly attributing symptoms, SAFER could therefore provide a clearer diagnosis of the patient.

Potential for Reanalyzing Data from Large Studies

It might be worthwhile to reanalyze data from large studies by applying post hoc SAFER criteria to dichotomize patients as "SAFER passes" or "SAFER fails" and then reinterpret the results. This approach may give additional insight into the actual efficacy of antidepressant treatment in patients who are more rigorously selected for depression—whose symptoms might be more tractable with antidepressants. Thus, by reanalyzing data from older studies that have available information about these factors, we could, for example, remove patients who are "SAFER fails" and analyze only patients who are "SAFER passes." We might, as a result, obtain different efficacy findings for the experimental therapy of that particular study. This same approach could be applied to individual chart reviews by clinicians in practice, particularly in cases of complex patients who have not responded well to treatment. The implementation of SAFER in clinical settings may therefore enrich the psychiatric diagnostic process, allowing for a more precise understanding of the patient's situation and assisting in the selection of effective treatments.

Development of More Targeted Treatments

The use of SAFER might lead to the development of treatments targeting specific symptoms and subpopulations of patients, which may result in fewer trials of inappropriate therapies and thus more rapid alleviation of depressive symptoms. This approach may also encourage pharmaceutical companies to develop drugs for more narrowly defined depressive syndromes and subtypes, as opposed to drugs targeting the largest number of individuals. Narrower treatment indications may pose a marketing/economic challenge, but targeting specific treatment populations may increase signal detection in two ways: by decreasing variability due to diagnostic heterogeneity and by decreasing the time needed to identify an appropriate treatment for a specific patient, thus saving precious time for the patient and clinician.

CASE ILLUSTRATIONS

Below are four case illustrations of patients that the SAFER identifies as not appropriate for a clinical trial.

– A subject became depressed one year ago. He attributes his depression largely to many years of psychosocial stressors or external factors and recognizes that a change in life circumstances would do much to improve his depression. In particular, he attributes his depression to being laid off from his job. He believes that a new job will alleviate his depression and that a drug could not change anything because he has been through a similar situation before, with depressive symptoms that started when he was laid off five years ago and that improved when he got a job. Such a patient would fail the SAFER assessment due to lack of specificity arising from external circumstances (the loss of his job) (Figure 2). During the relatively short period of the clinical trial, it would be hard to ascertain the origin of a symptom's improvement (either the antidepressant or the new job).

– A subject has been depressed all her life and cannot recall a time in her life when she felt happier for more than a few hours, and even those occasions were rare. This patient would fail the SAFER on the ground of "state versus trait" symptoms, meaning that the symptoms represent a chronic, long-standing trait characteristic and are unlikely to change during a brief pharmacological intervention. The situation would potentially be different, though in cases of "double depression" (MDD superimposed on dysthymia), where an exacerbation of preexisting symptoms may be measurable (Figure 2).

– When describing her past psychiatric history, a subject recalls being very depressed for a year. She intermittently has had symptoms for one to two weeks at a time every few months. Although she reported significant symptoms in the two-week window evaluated in the Structured Clinical Interview for DSM-IV (SCID), she also reports a past history of suddenly feeling almost completely cured a few days after starting an antidepressant. Closer examination of the overall timeline raises the question of whether this case is actually one of MDD, because the overall course of the illness and this rapid recovery is not consistent with what is known about MDD in the "real world" (Figure 2). This case would constitute a SAFER fail on the ground that the symptoms lack ecological validity relating to the overall course of the illness, severity of symptoms, or presentation of symptoms. For example, symptoms that are very intermittent may pass a SCID if captured in the right window of two weeks, but they may fail a week later at the SAFER interview if they have improved or generally do not have a significant impact on the patient's condition. This case is one of a second-level check to correct an initial interview with a SCID that may have been a close pass.

– A subject has a hard time describing his mood and is unable to answer interview questions in a clear and unambiguous manner. His answers to questions about specific symptoms are filled with qualifiers, digressions, and irrelevant details, along with a general vagueness. He has poor recall regarding the time course of his illness and symptoms, and is not able to describe how his current state compares to his nondepressed baseline. This case is a SAFER fail for a lack of assessability (Figure 2).

LIMITATIONS
SAFER is not without limitations. First, this new instrument requires systematic investigation of its face, construct, and predictive validity or lack thereof. The instrument was designed to be of use to research groups that carry out antidepressant clinical trials—in particular, by ensuring that patients selected are more likely to have MDD as opposed to "depression" that is purely secondary to situational factors or to medication side effects, or that is simply not consistent with the "real world" presentation of MDD. We believe that SAFER could also be of use to clinicians in practice who are faced with the decision of whether to treat a patient with antidepressants, to recommend other interventions (e.g., psychotherapy), or to simply observe the patient as he or she negotiates a particular situation, such as searching for a new job or managing grief. Patients who do not respond to antidepressants in clinical trials, or who are placebo responders, will often respond (or not) similarly in standard clinical practice. Therefore, if clinicians use the SAFER instrument in their initial assessment of patients, it may allow them to select treatment more effectively, which would improve clinical care and presumably result in better outcomes.

Second, patients may retrospectively associate external events with the onset of their symptomatology even when objective evidence suggests no connection. Thus, in clinical practice, the separation between "endogenous" and "reactive" depression might be challenging or even impossible to ascertain. While major life events (e.g., death of a loved one) and the burden of minor stressors (e.g., daily hassles) have been shown to predispose patients to MDD, SAFER aims to tease apart depressive symptoms that might be caused by an external event or context that may obscure the potential clinical impact of the drug—or lack of it—during the trial.

Third, by using the narrow SAFER inclusion criteria in a clinical trial, we obtain a highly selected group of patients with clearly defined episodic MDD leading to assessable functional impairments. While this narrowing may contribute to improved signal detection in drug development, it may limit generalizability of findings to a more narrow population.

CONCLUSION
The SAFER criteria identify patients who, despite meeting DSM criteria, would be poor candidates for drug trials. By vetting patients for enrollment into clinical trials, we hope to increase the likelihood of separation between novel

antidepressant treatments and placebo. To be eligible for clinical trials, patients must have a clinical presentation with both face and ecological validity that goes beyond mere symptom identification. The assessment of diagnostic validity could be made more accurate by the use of the SAFER criteria. Finally, while we have used the SAFER primarily for studies of MDD, the instrument could potentially be adapted for use with other conditions, such as bipolar disorder, anxiety disorders, or psychotic disorders.

REFERENCES

1. Greenberg PE, Kessler RC, Birnbaum HG, et al. The economic burden of depression in the United States: how did it change between 1990 and 2000? J Clin Psychiatry 2003;64:1465–75.
2. Greenberg PE, Stiglin LE, Finkelstein SN, Berndt ER. The economic burden of depression in 1990. J Clin Psychiatry 1993;54:405–18.
3. Fava M. Diagnosis and definition of treatment-resistant depression. Biol Psychiatry 2003;53:649–59.
4. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. Washington, DC: American Psychiatric Press, 2013.
5. Targum SD, Pollack MH, Fava M. Redefining affective disorders: relevance for drug development. CNS Neurosci Ther 2008;14:2–9.
6. Cassano P, Fava M. Depression and public health: an overview. J Psychosom Res 2002;53:849–57.
7. Kirmayer LJ, Robbins JM, Dworkind M, Yaffe MJ. Somatization and the recognition of depression and anxiety in primary care. Am J Psychiatry 1993;150:734–41.
8. Fava M. Somatic symptoms, depression, and antidepressant treatment. J Clin Psychiatry 2002;63:305–7.
9. Andreasen NC. DSM and the death of phenomenology in America: an example of unintended consequences. Schizophr Bull 2007;33:108–12.