Research Review Disposition of Comments Report

February 2018

Research Review Title: Definition of Treatment-Resistant Depression in the Medicare Population

Draft review available for public comment from 9:00 am September 1, 2017 to 5:00 pm on September 22, 2017.

Research Review Citation: Gaynes BN, Asher G, Gartlehner G, Hoffman V, Green J, Boland J, Lux L, Weber RP, Randolph C, Bann C, Coker-Schwimmer E, Viswanathan M, Lohr KN. Definition of Treatment-Resistant Depression in the Medicare Population. Technology Assessment Program. Project ID: PSYT0816. (Prepared by RTI–UNC Evidence-Based Practice Center under Contract No. HHSA290201500011I_HHSA29032006T). Rockville, MD: Agency for Healthcare Research and Quality. February 2018. http://www.ahrq.gov/clinic/epcix.htm.

Comments to Research Review

The Effective Health Care (EHC) Program encourages the public to participate in the development of its research projects. Each research review is posted to the EHC Program Web site or AHRQ Web site in draft form for public comment for a 3-4-week period. Comments can be submitted via the Web site, mail or email. At the conclusion of the public comment period, authors use the commentators' submissions and comments to revise the draft research review.

Comments on draft reviews and the authors' responses to the comments are posted for public viewing on the Web site approximately 3 months after the final research review is published. Comments are not edited for spelling, grammar, or other content errors. Each comment is listed with the name and affiliation of the commentator, if this information is provided. Commentators are not required to provide their names or affiliations in order to submit suggestions or comments.

The tables below include the responses by the authors of the review to each comment that was submitted for this draft review. The responses to comments in this disposition report are those of the authors, who are responsible for its contents, and do not necessarily represent the views of the Agency for Healthcare Research and Quality.

Commentator & Affiliation	Report Section	Comment	Response
Public Reviewer, Charles R. Conway, MD	General	Overall, the AHRQ report summarizing the literature on treatment-resistant depression (TRD) is well done and points out many of the current dilemmas facing psychiatry regarding this difficult to treat population (how you define TRD, how do you determine what constitutes a "failed trial", etc.). However, as detailed below, although there is debate about what defines TRD, there is no question that it exists and that it leads to considerable suffering and death (via suicide). The report makes several suggestions regarding how the field should progress in creating consensus; however, the report should place greater emphasis on the need to currently aggressively pursue studies that use adequate measures of mood, functional outcomes, and suicide reduction.	We appreciate the reviewer's positive comments. We agree on the importance of aggressively pursuing measures of mood, functional outcomes, and suicide reduction; we emphasize these types of measures when we discuss the package of outcome measures. We did not specifically mention a suicide measure, but we have now added in research recommendations consideration of a measure of suicidality as a key outcome. The consensus process certainly could recommend such a type of measure.
Public Reviewer, Nathaniel Z Counts, JD Mental Health America	General	Mental Health America (MHA) appreciates the Agency for Healthcare Research and Quality's (AHRQ's) thoughtful systematic review of definitions of treatment resistant depression (TRD) used in the literature. To build on this, MHA asks that AHRQ conduct a second analysis of the studies reviewed to determine, based on the evidence available, whether there are common factors that describe or predict TRD, other than the definitions used in the studies. For example – do individuals who are categorized as having TRD have especially high PHQ-9 scores, or share certain aspects of social complexity? It would greatly advance the field if AHRQ were able to identify factors that could be screened for earlier that predict three treatment failures, instead of having an individual undergo three treatment failures and then being classified as TRD. If AHRQ is unable to identify these factors, noting this gap in research would be valuable.	We appreciate the author's comments. While the questions posed are important, and important to answer, they are outside of the scope of this Technology Assessment. We encourage the reviewer nominate such a topic for AHRQ to consider at https://effectivehealthcare.ahrq.gov/get-involved/suggest-topic. Further, we note that, in the Future Research section, we recommend the creation of large, coordinated treatment registries, which could support these kinds of analyses.

Commentator & Affiliation	Report Section	Comment	Response
		With well over one million PHQ-9 results in its screening database from a self-selected, largely help-seeking population, MHA may be able to help in identifying these factors. Identification of factors that predict or describe TRD outside of three treatment failures is critical for health care payment and delivery. Without a TRD designation, payers or re-insurers will have less available granularity for risk-adjustment through mechanisms like the Hierarchical Condition Categories – depression will be depression and pay as such. If TRD could be identified and risk stratified at a higher tier, it would promote more aggressive treatment of higher needs earlier, and decrease the likelihood of adverse selection – ultimately rewarding providers' efforts to manage more complex needs. MHA appreciates that this request is time and labor intensive, and as noted above stands ready to help or support in any way appropriate, and thanks the AHRQ for its thoughtful analysis of a challenging issue. Please do not hesitate to reach out to Nathaniel Z Counts, JD, Senior Policy Director, at ncounts@mentalhealthamerica.net for questions or support.	
Public Reviewer, Nathaniel Z Counts, JD Mental Health America	General	A very thorough and accurate report!. A few opportunities for improvements: add the Aaronson et al AM J Psych report on 5 year registry of outcomes for TRD Published on line before the cut off date and should have been in the report. Will diverse opinions on threshold for number of failed trials, it is worth emphasing that EMA, Canada, FDA and Australia, concur at least two failed trials is minimum. The risk	Thank you. We have added the Aaronson citations. We have identified the EMA as defining TRD as a minimum of two trials in Table 6. We could not find data indicating that FDA, Canada, or Australia had officially confirmed this number of 2 failed trials as a minimum threshold.

Commentator & Affiliation	Report Section	Comment	Response
		factors associated with TRD are clear and widely published. No need to examine further. See studies from STARD, Perlis 2011, and multiple epidemiology studies.	
Public Reviewer, Phyllis Foxworth, Depression and Bipolar Support Alliance	General	*Efforts to improve definition and measurement of success include the perspectives of those who live with TRD.* For people who live with TRD, the past 25 years have seen anemic progress in the development of meaningful new treatments. Innovation has been incremental. People are consequently frustrated by, and losing hope for a solution. Modest improvement in clinical outcomes is simply no longer enough. Of course the first priority for treatment is ensuring that a person living with TRD is provided a pathway out of crisis and onto stability. However, all too often, this baseline stability is also the end goal established for successful long-term care. "Stable" or "better" are not always synonymous with "well." The goals for clinical success must focus on providing a pathway to a life well-lived as defined by the individual. Because this is often not the defined objectives for clinicians and researchers, the potential exists for the patient's definition of success to be obscured.	We agree. We suggest a multi-stakeholder consensus process, which includes patients and patient advocates in the research development process, and importantly also involves patient-centered outcomes (e.g., ability to work, improved sleep, improved pleasure from daily activities).
Public Reviewer, Phyllis Foxworth, Depression and Bipolar Support Alliance	General	*Measures of treatments' efficacy need to evolve.* DBSA firmly believes that evaluative measurements for TRD treatments should not only include symptom relief reduction, but must also include—and put a significant emphasis on well-being and impact on quality of life. To do this effectively, these measurements must be informed by the lived experience of individuals living with TRD. While symptom relief and side-effect management are already typically captured for clinical trials, a data gap remains in measuring well-being. In order for patients to make informed treatment decisions regarding their	We agree. We suggest a multi-stakeholder consensus process with patients and patient advocates in the research development that importantly involves patient-centered outcomes (e.g., ability to work, improved sleep, improved pleasure from daily activities). Of note, our charge was not to identify a preferred tool, but to clarify the need to use tools that measure quality-of-life outcomes.

Commentator & Affiliation	Report Section	Comment	Response
		unique and complex mental health conditions, limiting research to clinical effectiveness using measurements such as the HAM-D or MADR scales that focus on symptom relief is inadequate. Patients are seeking interventions that help them function in life rather than alleviate their symptoms. Examples include: • Relief from bad decision-making • Ability to work and earn an income • Getting better sleep Encouraging studies to include measurement tools that include wellness outcomes as defined by people with TRD would improve the potential to identify treatments with outcomes that are meaningful to patients. For example, AHRQ should recommend elevating the importance of existing measurement tools that address function, such as the Sheehan Disability Scale, and/or that address wellness, such as theWHO-5 Scale. The Sheehan Scale is a comparative tool that asks people to evaluate disruptions in various areas of life such as work/school, family, and social life. Meanwhile, the WHO-5 evaluation tool asks a person to report on the active presence of certain positive aspects of overall well-being, such as feeling “calm and relaxed,” or “active and vigorous.” Both are useful in allowing not only for the mood-related improvements necessary to achieving complete wellness, but also the interpersonal and relational aspects of individuals’ experiences of TRD. In addition resiliency scales that measure the ability to thrive in the face of adversity such as the Connor-Davidson Resiliency Scale would support identifying treatments that improve the quality of life irrespective of the clinical definition of success.	We have added mention of the Sheehan Disability Scale and the WHO-5 scale now in Research Recommendations, which are more commonly used. We did not mention the Connor-Davidson scale because it is less commonly know

Commentator & Affiliation	Report Section	Comment	Response
Public Reviewer, Phyllis Foxworth, Depression and Bipolar Support Alliance	General	*Implications of chronic, versus episodic, experiences of TRD.* Of note in the literature review, 63% of the studies terminated in six months or less. Success should not be defined by controlling this week's, month's or year's episode, but by reducing the severity and eliminating the reoccurrence of symptoms over the entire lifetime. This is not often the defined objective for clinicians or researchers, but it is of vital importance to people experiencing TRD. DBSA envisions exploration of chronic versus episodic experiences and how treatments may need to differ for the chronic recurrence of TRD symptoms.	We agree, and the report emphasizes the need to identify a meaningful standard length of treatment given the chronicity of TRD. We note the importance that study durations "provide enough time for patients to receive an adequate dose and duration of the intervention." We have also added a sentence addressing how risk of relapse increases with greater level of treatment resistance, suggesting a need for even longer trials for the more severely resistant and an acknowledgement that this group might require different kinds of interventions than then with less chronic and more episodic depression (e.g., see Executive Summary, Research Recommendations section (and in full report).
Public Reviewer, Phyllis Foxworth, Depression and Bipolar Support Alliance	General	*DBSA notes that payers, including the Centers for Medicare and Medicaid Services (CMS), hesitate to include novel treatments for depression.* Because payers tend to resist coverage for new treatments, an inadvertent disincentive for research and innovation exists. This vacuum could be filled by encouraging research that identifies how TRD treatment options that target whole health and wellness are cost-effective and reduce the overall burden on the healthcare system.	This is an important point, but the suggestion goes beyond what we reviewed for this Technology Assessment.
Public Reviewer, Phyllis Foxworth, Depression and Bipolar Support Alliance	General	DBSA supports AHRQ's initiative around TRD. We sincerely hope that the committee's work will promote an environment that supports the development of better treatment options and encourages exploration of the steps that need to be taken in order to break	Excellent point—we agree, and have emphasized the importance of a patient-centered approach in our research recommendations in the

Commentator & Affiliation	Report Section	Comment	Response
		out from the current dynamic of incremental, slow improvement, to one of exciting breakthroughs. Part of this depends upon a transformation of the way we currently measure success and to tools used to measure it. We urge the committee to provide guidance from those living with TRD to bend the focus of scientific discovery towards the things that matter most to us.	Evidence Summary (p.) and the Discussion of the main report (p
Public Reviewer, **Kenneth Kramer, PhD** **Allergan**	General	Major depressive disorder has a significant influence on the way that American adults live, work, and interact with the world around them. Over 16% of American adults experienced a depressive episode in 2015.[4] While available prescription drug treatments and psychotherapy effectively help many of these patients, there are still some (upwards of 20% or more) whose depression does not respond to treatment. It is critical that the mental health community continue to work together towards developing innovative treatments for all patients living with depression. A consensus definition of TRD is one step towards that goal. [4] National Institute of Mental Health. Major Depression Among Adults (https://www.nimh.nih.gov/health/statistics/prevalence/major-depression-among-adults.shtml).	Thank you for this observation.
Peer Reviewer #1	General	I think the topic of this report is very important and will hopefully have some excellent impact on future research. I think the key questions are appropriate and very explicitly stated...but a few inconsistencies in places. This is an excellent summary and I would use it.	We appreciate the reviewer's comments
Peer Reviewer #2	General	The report asks the right foundational questions: What is TRD? How best can we learn more about its origins, course, and most beneficial treatment strategies?	We agree with the reviewer's suggestions. Accordingly, we have made the suggested change in key messages and modified references to TRD treatment failures throughout the

Commentator & Affiliation	Report Section	Comment	Response
		The population and audience are clear. I appreciate that the report points to certain challenges (especially around exclusion criteria like co-occurring conditions, suicidality, etc.) in comparing the people who participate in research to real world presentations with no one excluded. The key questions are clear and make sense. I suggest addressing some language issues that occur sporadically throughout the report. I'll provide some specific pages/line numbers below, but here are general, overarching points: > In many if not most instances, to refer to patients as "people", "individuals", "people with TRD", "the TRD population", etc. would not obscure meaning or make the syntax unduly awkward. > Some prose within the report suggests individuals receiving treatment are failing to respond, rather than treatments being inadequate. For example, right away, within the "Purpose and Key Messages" on the second page after the cover, consider line 14: "failure of patients to respond or go into remission". I think some minor rewording could eliminate the implication that people are personally failing and/or somehow willfully resisting the efficacy of treatments. An alternative (along the lines of page iv, lines 36-37) might be "a treatment's failure to eliminate or at least significantly improve". > The TRD definition synthesized within lines 10-15 on page ES-4 is excellent. I'd offer this as a model for descriptions of TRD within the whole report. It makes the depression itself the entity that's resistant/problematic, not the person with the depression.	report per his suggestions. Thanks again for the comment.

Commentator & Affiliation	Report Section	Comment	Response
		The foregoing are not essential changes, but if we can improve the report language such that it's less stigmatizing, that's a good thing!	
Peer Reviewer #3	General	This report nicely captures in appropriate detail the issues at hand. The report is definitely clinically meaningful, the target population and audience are explicitly defined (with a comprehensive, if frustrating, survey of the many competing definitions of TRD!), and the key questions are appropriate and explicitly stated. That many questions remain open to discussion and debate is, unfortunately, a reflection of the continuing difficulty in executing a unified approach to research aimed at putting to rest lingering uncertainties. Through no fault of the report's authors, we are left with much work yet to do in the area of TRD diagnosis and treatment. The report's suggestions for adopting common data elements and agreement across the field on fundamental issues involving tools for evaluation and treatment decision-making are sound. What remains unclear is whether the understanding of the underlying pathophysiology of depression and mechanisms of action of antidepressant interventions will be required before such research can truly move forward.	We appreciate the reviewer's comments and agree with his observations on the gaps in this area.
Public Reviewer, Kenneth Kramer, PhD Allergan	Abstract	The lack of a consensus definition among researchers, medical professionals, and patients leads to inconsistent drug development protocols or worse, poor real-world patient outcomes. Unfortunately, no definition of TRD is included in the latest edition of the Diagnostic and Statistical Manual of Mental Disorders ("DSM-5"), maintained by the American Psychiatric Association. This creates confusion because it is not clear whether TRD is a characteristic of the disease or a manifestation of failed attempts at treatment. To substantially improve	We agree with the reviewer, and the call to develop the package of outcomes (e.g., Executive Summary, p.) involves forging the kind of consensus definition the reader suggests.

Commentator & Affiliation	Report Section	Comment	Response
		patient wellbeing and health outcomes, Allergan is committed to assisting - in any capacity we can - with the development of a consensus definition of TRD that supports the needs of all parties: Allergan agrees with the conclusion of the draft technical assessment which establishes that no consensus definition currently exists to properly define treatment-resistant depression. However, recognizing the regulatory importance of establishing a TRD definition, Allergan strongly urges that physicians, researchers, manufacturers and patients work together in order to establish a regulatory and clinical definition which the FDA may then act upon when operationalizing its role with manufacturers during the drug development process. Allergan acknowledges the importance of establishing a consensus definition of TRD which can be used by the FDA to guide future clinical trial design, and establish a regulatory pathway, that will result in effective and safe treatments for those suffering from treatment-resistant depression.	
Public Reviewer, A. John Rush MD	Executive Summary	EXECUTIVE SUMMARY: Would add that the high disability, cost mortality associated with TRD requires that research move forward and that at least two meds are a minimum, though for some treatments with higher risk might require more treatments when treatment risk is higher. Finally, the increasing likelihood of relapse with greater levels of resistance argues for need to	This point is key. We agree, and the report emphasizes the need to identify a meaningful standard length of treatment given the chronicity of TRD. We note the importance that study durations “provide enough time for patients to receive an adequate dose and duration of the intervention.” Also, we have now added a sentence addressing how risk of relapse increases with greater level of

Commentator & Affiliation	Report Section	Comment	Response
		demonstrate longer term efficacy n the more resistant patients	treatment resistance, suggesting a need for even longer trials for the more severely resistant and an acknowledgement that this group might require different kinds of interventions than then with less chronic and more episodic depression (e.g., see Executive Summary, p.
Peer Reviewer #1	Introduction	Introduction was good and the length was adequate. I did not find the rationale sufficiently detailed in terms of how some of the questions came about. I don't quite understand why a narrative review is what was used to describe the methods for KQ1 - 5. The flow diagram suggests that citations were screened by 2 reviewers with reproducible criteria to achieve 37 citations.	The nominator and sponsor of this Technology Assessment report (TA) originally proposed the structure of the TA (which does not have the usual "topic refinement" phase that many Evidence-based Practice Center systematic reviews have). As appropriate and necessary, we discussed the narrative questions and the "systematic" key questions with the sponsor, agency staff, and other experts. This approach is typical of most evidence reports done by our EPC. Regarding the second point about the flow diagram: We had agreed with the idea that we would handle KQs 1-5 generally as narrative questions (which often turn up for systematic reviews as well); nevertheless, we applied standard methods for the title/abstract and full-text review insofar as possible, to meet the typical global standards for this step.

Commentator & Affiliation	Report Section	Comment	Response
Peer Reviewer #2	Introduction	The introduction is clear and concise. On page ES-1, lines 21-22, consider adjusting "Such patients pose a common, challenging presentation to psychiatric and primary care clinicians." to "Such depression [or simply "TRD"] poses a common, challenging set of treatment questions and issues to psychiatric and primary care clinicians." or just changing "patients pose" to "depression poses" and leaving the rest of that sentence as-is, which may be necessary to honor the source that is cited. Also on page ES-1, line 29, how about changing "Patients with TRD incur" to "TRD represents". Again on page ES-1, lines 30-31, could we make "Treatment-resistant patients are twice as likely to be hospitalized; their cost of hospitalization is more than six times the mean total cost for depressed patients who are not treatment resistant." read "Individuals with TRD are twice as likely to be hospitalized; the cost of this hospitalization is more than six times the mean total cost for people with depression that is not treatment resistant." instead?	We agree and have made the recommendations suggested.
Peer Reviewer #3	Introduction	The Introduction sets the stage well. The description of the various approaches to defining TRD, as well as diagnostic and outcome measures is thorough and clear. The specific point that common depression rating scales, e.g. Hamilton, Beck, are best used as measures of severity and not for establishing a diagnosis of depression is an important one. One issue that could be more strongly highlighted is the role of patient preference in treatment selection, which may well play a role in clinical decision-making	We appreciate the reviewer's comments. The reviewer's point about the import of patient preference and the role of non-evidence-based factors in making

Commentator & Affiliation	Report Section	Comment	Response
		in TRD….it must be acknowledged that ideal, equipoise studies sometimes cannot be achieved; clinical decision-making can be influenced by a host of factors and concerns that are independent of actual data.	treatment decisions is an important one. We highlight this in Introduction section at the end of paragraph 2 on page 1.
Peer Reviewer #1	Methods	The eligibility criteria were justifiable and the fact that TEP and key informants were surveyed to rank important outcomes is excellent. It adds credibility to the process. However I wish they would show these results....rather than describe the final outcomes. Don't have a sense if there was diverse representation of the TEP and if any patient representatives were included. Using the HDI to select countries also speaks to reproducibility and justification of the countries included in the narrative review. It is not clear why the CAM therapies were delimited as they were. It seems that many more such therapies are in wide use among people with depression. Perhaps more explanation as to how this short list was derived. The searches were well described and reproducible. The outcomes are common in the area of depression research. It was not clear how "primary" outcome was determined as the methods state that only the primary outcomes were extracted.	We appreciate the thoughtful consideration. The number of CAM therapies was limited because we required that eligible interventions be tested as a treatment to target TRD, and both CAM (and exercise) were infrequently tested as TRD interventions.
Peer Reviewer #2	Methods	The methods section is fine. Within Table A on page ES-2, could KQ 3 include something to represent patient-desired (as opposed to patient-reported or even patient-centered) outcomes?	We appreciate the reviewer's comments. These key questions were developed with the nominator/sponsor, and we

Commentator & Affiliation	Report Section	Comment	Response
		This might be "what if any consideration has there been of the outcomes most important and meaningful to people with TRD". Might the same item, KQ 3 in page ES-2's Table A, include something representing well-being and resiliency? I could envision "have considerations of not only symptom reduction, but also improvement on measures of well-being and resiliency, been included". A question to consider: do you think that co-occurring conditions, identified as a major issue within the TRD population, are represented adequately within the Key Questions in Table A? They seem implied within KQs 6, 9, 10, and 11, but they're not mentioned specifically. Is this a gap we should consider filling? On page ES-3, within the Outcomes section of Table B, I again wonder whether we might include outcome measures related to well-being, resiliency, and even co-occurring conditions.	need to report the specific questions we answered. We will leave as is. We appreciate the point, but these key questions were developed with the nominator/sponsor, and we need to report the specifics questions we answered. We will leave as is. . We appreciate the point, but these key questions were developed with the nominator/sponsor, and we need to report the specifics questions we answered. That said, we do believe that comorbidity, to the extent that it is described in the literature, is adequately addressed. We will leave as is. These are important considerations, but the outcomes listed in this PICOTS Table (Table B) reflect the inclusion criteria we used, so we must keep as is.
Peer Reviewer #3	Methods	The methodology is well-described and appropriate… [however] there is an important diagnostic point that is worth mentioning. One contributing factor to some cases of treatment-resistant depression has been identified as a failure to properly diagnose a patient presenting with depression as in fact suffering from bipolar disorder.	This point is an important clinical one, and we have added this point to the third paragraph of the Introduction section.

Commentator & Affiliation	Report Section	Comment	Response
Public Reviewer, Charles R. Conway, MD	Results	The report strongly recommends that registries should be done to study TRD; however, the largest and longest duration published registry of TRD (assessing vagus nerve stimulation versus treatment as usual) is not detailed in the paper); Aaronson et al., American Journal of Psychiatry, 2017 (electronically published March 31, 2017).	Thank you for the suggestion. We have now added this citation, which is an update of the one we had already cited.
Peer Reviewer #1	Results	The results are very succinctly laid out and I appreciate the KEY POINTS section in KQ1 to 5. The tables were clear and had good information. To my knowledge I have never seen consensus defined by the frequency of use within the literature. Consensus can vary in understanding but it implies a level of agreement or disagreement. There is no reference provided to show that this has been used previously in the literature. I would strongly suggest that a different term other than consensus is used. Frequency of publication does not in my mind imply consensus (majority agreement). In fact I can't quite make sense of the tables that follow and then the definition of consensus. Perhaps if it is better explained. Doesn't make sense to me as it is presented...although the information is useful.	This is an important observation. We agree that "consensus" may mean "majority agreement" that can be arrived at in numerous ways. For example, such a "majority agreement" can be determined by guidelines or best practices, or by consensus statements, or by systematic reviews that represent state of the art consideration of the relevant topic, or other ways. Our charge was to identify a "consensus" in the TRD literature. Since there is no clearly identified group to make such an agreement, we believed that what is presented in consensus statements, clinical practice guidelines, government reports, and the peer-reviewed literature is a legitimate marker for agreement among clinical, research, and policy experts, especially if a majority of those citations appear to agree with each other over time. We have now added this explanation to our Methods section in the main report, in the Data Synthesis subsection on page 10.

Commentator & Affiliation	Report Section	Comment	Response
Peer Reviewer #2	Results	The results section is fine—clear, generally adequate, descriptive, and mostly inclusive.	We appreciate the reviewer's comments.
		Once again, the TRD definition synthesized within lines 10-15 on page ES-4 is an ideal model for writing about depression itself as the problem, not the person with the depression.	Thank you.
		Within Key Question 3 on page ES-7, could we include comment about outcome measures related to well-being, resiliency, and co-occurring conditions? (The comment may be that these were not mentioned, but I think calling that out is important.)	This observation is important, but these measures were not part of our PICOTS selection criteria for studies, so we cannot add it.
		What if any mention of diversity and social/cultural considerations (race, ethnicity, country of origin, faith, experience of trauma, gender identification, sexual orientation, justice involvement, armed forces or Veteran status) in study design were present in the literature? In Key Question 4 on page ES-7, I think pointing to this—and I suspect the lack of such mention—would be useful.	Again, this point in TRD is important, but it is not one that addresses what the assigned TRD questions were.
		In Key Question 5 on page ES-7, I really think touching upon co-occurring conditions—not just mental health and substance use issues, but also "physical" health issues—makes sense, even if just to point out that exclusion criteria may have affected the prevalence of these within the studies, and/or that these were simply not mentioned.	We appreciate the question. However, we assessed the literature to see whether medical comorbidity had an important role, and it did not predict TRD, so we will leave the text as is.
		On page ES-8, in Key Question 6, lines 22-24, I again wonder whether any mention could be made of co-occurring conditions' inclusion/non-inclusion and impact.	We considered this suggestion, but comorbidities (along with a variety of other variables) were frequently considered as exclusion criteria, and

Commentator & Affiliation	Report Section	Comment	Response
		Also on page ES-8, in Key Question 6, lines 22-24, it may be desirable once again to bring up cultural and social considerations' inclusion/non-inclusion and impact.	this finding did not raise to the level of a key point. This point is interesting, but it is not one that we were asking to review for this Technology Assessment, so we have no information about it.
Peer Reviewer #3	Results	The studies are well-described, with an appropriate level of detail. Tables are especially clear and helpful. The literature is still heavily laden with a collection of inconclusive - and occasionally contradictory findings - that are handled as well as possible in the report. A couple of lines of ongoing research could usefully be included (e.g. KQ #5). Although far from definitive, additional factors predicting response (or nonresponse) to antidepressant treatments include psychosocial (history of childhood trauma) and biological (fMRI imaging measures) parameters. Nemeroff et al. (2003), studying patients with "chronic depression" found a history of childhood trauma to predict preferential response to psychotherapy and – confirmed more recently (Williams et al. 2016), nonresponse to antidepressant pharmacotherapy. An open-label pilot study of brain imaging found two types of connectivity patterns on resting state fMRI to predict positive antidepressant response to behavioral activation psychotherapy (Crowther et al. 2015).	We appreciate the reviewer's comments The references cited indicate potentially important areas for future research, but they did not meet our selection criteria, so we are not able to include this articles.
Public Reviewer, Charles R. Conway, MD	Discussion/ Conclusion	I concur with the authors of the well-written report that the field of psychiatry has yet to come to a consensus on a definition of TRD. Certainly, the movement is towards defining TRD as having failed two adequate dose-duration trials of medications/psychotherapy/ ECT.	We appreciate and agree with the reviewer's comments.

Commentator & Affiliation	Report Section	Comment	Response
		In my opinion, given the horrible nature of TRD, and the consequences of failing to successfully treat TRD (morbidity and mortality via suicide), there should be more emphasis on the urgency of developing and covering treatments for this population. The report makes the mistake of the "perfect is the enemy of the good" in setting up a very difficult/impossible goal of psychiatry coming to a uniform consensus on TRD, without adequate emphasis on the urgent need to study and reimburse treatments that treat the condition. TRD is a deadly illness (estimated 15% die by suicide) and recent evidence from CDC that rates of suicide have been rising in the past decade; recent studies demonstrate that, on average, a U.S. Veteran dies by suicide every hour. Studies using well-characterized TRD patients with well-documented histories of treatment failure of adequate duration must go forward NOW. We do not have the luxury of waiting for the perfect.	
Peer Reviewer #1	Discussion/ Conclusion	This is a very dense report to digest and unfortunately my time was limited. I think the review authors did an excellent job of summarizing and critically appraising the questions they set out to address. I would note that they did not identify studies that compared ACROSS different measures used to assess depression (even though there may not be studies that assessed this in TRD patients as the instruments are the same. There is some increasing concern different measures are not comparable. I am not sure this issue was considered.	This is an important topic, but the comparative perspective for measures used in different studies was not a key question for the report.
Peer Reviewer #2	Discussion/ Conclusion	The Discussion/Conclusion portion is excellent—my favorite part of the report. (Of course it is; it's where we can begin to think through how we will fill gaps in understanding, treatment, and lived experience.) The	We appreciate the reviewer's comments.

Commentator & Affiliation	Report Section	Comment	Response
		implications are clearly stated. What's more, and of special note to us at the Depression and Bipolar Support Alliance, is that they align with the general considerations about treatment for major depressive disorder (MDD)—as opposed to TRD alone as a distinct subset…. The implications—that a clear, consensus definition of TRD is needed; that identification of the best and most patient-oriented outcome measures (such as well-being, disability, function) is vital; and that considerations of course of illness including level of severity and treatment duration and adherence are needed—are clear. Limitations—lack of evidence identifying risk factors, insufficient data to assess prognostic factors, and especially inconsistent reporting of definitions of and outcomes related to TRD—were also clearly presented. … The future research section is particularly well-executed and very clear and specific. The only thing I'd have liked to see is something pointing more directly to the issue of multiple co-morbidities' likely presentation along with TRD. Page ES-12, lines 33-36 does get at the importance of testing the effectiveness of interventions in real-world settings, but I'd very much like to see co-occurring conditions make its way onto the future research roster somehow, since I view it as a key under-researched issue that is of paramount importance to people with TRD.	We thank the reviewer for the comments. We agree that this point is important. We have added psychiatric comorbidities as one of the key factors to be considered as a potential confounder in the Research Recommendations on page ES-3rd paragraph, and 2nd full paragraph, in the Discussion of the main report.

Commentator & Affiliation	Report Section	Comment	Response
Peer Reviewer #3	Discussion/ Conclusion	The discussion and implications of the findings are clearly presented. Unfortunately, the major findings serve to highlight how incomplete and inconsistent the existing data base is. That only 17% of studies actually conform to the most well-validated tenets of TRD is striking, and concerning. The conclusions that the field needs to adopt a uniform set of definitions and approach to monitoring the trajectory of TRD, its natural course and response to interventions, e.g. by way of patient registries and electronic health records, are well taken. However, in addition to this "effectiveness" approach, it is evident from the review of the literature in the report that the field must also pursue more basic "efficacy" and "translational" research to better understand TRD…Intervention studies should not only include more modalities, e.g. psychotherapies, as is well documented in the report, but needs to include an "experimental therapeutics" or translational approach, whereby the targets of the treatments are identified and their engagement is confirmed.	We appreciate the reviewer's comments. The reviewer makes a thoughtful and innovative point about the breadth and depth of possible future research. Although some of this is beyond what we can definitively conclude or "recommend" from the evidence at hand in this technology assessment, we do mention these ideas in the discussion, which the nominator and sponsor can themselves consider. Specifically, we have added a half paragraph in Research Recommendations regarding translational approaches to address this point.
Peer Reviewer #1	Clarity and Usability	This report has done an excellent job of summarizing a wide variety of literature. The conclusions are HIGHLY relevant and those doing research in this area should take note.	We appreciate the reviewer's comments.
Peer Reviewer #2	Clarity and Usability	The report is well structured and makes its points quite clearly; there's no mistaking the foundational need of a clear, consensus definition of TRD and the adoption of outcome measures for TRD that are consistent, comparable, and complete.	We appreciate the reviewer's comments.

Commentator & Affiliation	Report Section	Comment	Response
		I don't really question the relevance but do somewhat question the importance and feasibility of the conclusions to policy and especially practice. To have the standard definition; sure. But to envision large-scale monitoring--when a standard definition is not yet in place and delivery of care to people who likely have TRD is occurring in such diverse and, sadly, fractured systems--seems like so potentially heavy a lift that I question its inclusion. I acknowledge that examples of other illness states' structures for tracking and monitoring are offered, but I wonder if there's another policy or practice recommendation (perhaps around integration and or specific screening and triage initiatives) that might be more approachable to already overwhelmed systems and clinicians.	This observation is thoughtful, but we think that in the era of electronic health records and the move to harmonize measures used in these records, and with some already successful examples (e.g., the OMERACT project), such monitoring is feasible and necessary to successfully address TRD.
Peer Reviewer #3	Clarity and Usability	The report is thorough and well-organized. This is actually quite an accomplishment, considering the lack of uniformity in the underlying studies and clinical reports. Indeed, the conclusions are sound but of necessity rather limited, in that the main new information gleaned was just how deep the lack of a unified vision - or even common definitions - go in the field. The call for new efforts at common understanding, e.g. agreement on definition of each of the components of TRD, common data elements in describing disease and response to treatment, etc. make good sense. However, there seem to be some basic issues that require further attention that are not focused on in the report. First and foremost, depression - treatment-resistant or otherwise - remains an incredibly heterogeneous disorder …. So in addition to the wise recommendations in the draft report, renewed	We appreciated the reviewer's comments. We agree that this point is important, but the issue of etiology is not one that we were tasked to address in this report, and we did not consider it in our data collection, so we do not

Commentator & Affiliation	Report Section	Comment	Response
		attention at the "front end" is imperative, e.g. efforts at better characterizing the pathophysiology - both biological and psychological - of depression, ideally including the development of biomarkers and a scientifically-based delineation of meaningful subtypes of depression. Recommended efforts such as patient registries will be more meaningful if the vast universe of treatment-resistant depressed patients can be meaningfully tracked separately based on defined criteria for each subtype.	believe we can conclude this (albeit important) point.
	Figures	N/A	
	References	N/A	
	Appendix	N/A	