Received: 20 July 2017 | Revised: 1 November 2017 | Accepted: 13 November 2017

DOI: 10.1002/da.22705

RESEARCH ARTICLE

WILEY ADAA ANXIETY AND DEPRESSION ASSOCIATION OF AMERICA

Finding treatment-resistant depression in real-world data: How a data-driven approach compares with expert-based heuristics

M. Soledad Cepeda MD, PhD | **Jenna Reps PhD** | **Daniel Fife MD** | **Clair Blacketer MPH** | **Paul Stang PhD** | **Patrick Ryan PhD**

Janssen Research and Development, Titusville, NJ, USA

Correspondence
M. Soledad Cepeda,1125 Trenton Harbourton Rd, Titusville, NJ 08560.
Email: scepeda@its.jnj.com

Background: Depression that does not respond to antidepressants is treatment-resistant depression (TRD). TRD definitions include assessments of treatment response, dose and duration, and implementing these definitions in claims databases can be challenging. We built a data-driven TRD definition and evaluated its performance.

Methods: We included adults with depression, ≥1 antidepressant, and no diagnosis of mania, dementia, or psychosis. Subjects were stratified into those with and without proxy for TRD. Proxies for TRD were electroconvulsive therapy, deep brain, or vagus nerve stimulation. The index date for subjects with proxy for TRD was the procedure date, and for subjects without, the date of a randomly selected visit. We used three databases. We fit decision tree predictive models. We included number of distinct antidepressants, with and without adequate doses and duration, number of antipsychotics and psychotherapies, and expert-based definitions, 3, 6, and 12 months before index date. To assess performance, we calculated area under the curve (AUC) and transportability.

Results: We analyzed 33,336 subjects with no proxy for TRD, and 3,566 with the proxy. Number of antidepressants and antipsychotics were selected in all periods. The best model was at 12 months with an AUC = 0.81. The rule transported well and states that a subject with ≥1 antipsychotic or ≥3 antidepressants in the last year has TRD. Applying this rule, 15.8% of subjects treated for depression had TRD.

Conclusion: The definition that best discriminates between subjects with and without TRD considers number of distinct antidepressants (≥3) or antipsychotics (≥1) in the last year.

KEYWORDS
databases, decision tree, epidemiology, machine learning, prevalence

1 | INTRODUCTION

Depression is associated with substantial morbidity, heathcare cost, mortality and family burden (Kessler, 2012; Rush et al., 2006b), and affects 5% to 8% of the adult US population annually (Blazer, Kessler, McGonagle, & Swartz, 1994; Cepeda, Stang, & Makadia, 2016; Kessler, Chiu, Demler, Merikangas, & Walters, 2005; Pratt & Brody, 2008).

No single treatment for depression is universally effective and sequential interventions are often needed (Agency for Healthcare Research and Quality, 2016; Rush et al., 2006b). About 20% of those receiving treatment do not achieve remission, while 50% have no response at all (Trivedi et al., 2006). Relapse rates increase with increases in the number of treatments required (Rush et al., 2006a). When a patient does not respond to multiple therapeutic courses of antidepressant medications, the patient is classified as having treatment-resistant depression (TRD).

Many clinical definitions of TRD exist (Berlim&Turecki, 2007), (Russell et al., 2004). The definitions range from not responding to a single

treatment to not responding to sequential treatments. The National institute of Mental Health-funded Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial, which enrolled 4,041 subjects randomized to different pharmaceutical treatments for depression, found that resistance to treatment markedly increased after the failure of two treatments at adequate dose and duration (Conway, George, &Sackeim, 2017; Rush et al., 2006a).

The use of observational healthcare databases, such as administrative claims and electronic health records, offers the opportunity to study large numbers of patients whose care reflect real world settings and it can provide real world evidence to inform medical decisions. However, implementing some of the current expert-based clinical definitions is difficult. Take, for example this definition of treatment-resistant depression: "depressions that do not resolve after antidepressant treatment in adequate doses or intensity and for a time sufficient for response" (Fava & Davidson, 1996). Implementing such a definition in claims databases is challenging because of the subjectivity involved in the assessment of adequate duration, the variability associated with daily dosing, and nonadherence that is not fully captured, and the difficulty in ascertaining why medications were replaced or stopped. Changes in medications could be due to adverse events, lack of effectiveness, or remission (Kubitz, Mehra, Potluri, Garg, &Cossrow, 2013). In the absence of information on how patients are responding to treatments in claims databases, it is difficult to ascertain the reason for the medication change, so many assumptions are used to decide whether or not a treatment failed.

Therefore, the definitions of TRD in claims databases are often complex (Kubitz et al., 2013). This complexity could be avoided if the definition was based solely on the number of different antidepressants and antipsychotics, a subject was exposed to during a specific time period. However, it is difficult to know if one definition is indeed better than another.

One approach to evaluating the performance of alternative TRD definitions is to use a reference 'gold standard' proxy for TRD, such that definiitons can be compared on their ability to identify reference cases and discriminate from noncases. A suitable 'noisy label' proxy would have high specificity, meaning patients observed with the proxy are highly likely to be classified as having the disease of interest (Agarwal et al., 2016). For TRD, a proxy for the disease could be any treatments that are recommended and consistently employed specifically in those cases when the depression has not responded to other treatment options (Conway et al., 2017). The treatments we considered were electroconvulsive therapy, deep brain stimulation, and vagus nerve stimulation (Conway et al., 2017), as these are consistently used to treat of subjects with TRD. Subjects undergoing these procedures are likely to have Stage II TRD (Conway et al., 2017). We wanted to separate depression severity from depression that is resistant to treatment as much as possible (Mathew, 2008), so we did not include suicide or suicide attempts as proxies.

We sought to build a data-driven definition of TRD and evaluate its performance relative to expert-based heuristic definitions.

2 | MATERIALS AND METHODS

2.1 | Target population

The target population was comprised of subjects who had at least one visit to a healthcare provider and who on the day of the visit: 1. Had a diagnosis of major depressive disorder or other depression diagnosis, 2. Were continuously observed in the database for at least 365 days prior to the visit, 3. Had at least one dispensing of an antidepressant in the previous 365 days, 4. Had another diagnosis of major depressive disorder or other depression diagnosis in the previous 365 days, 5. Had no prior diagnosis of mania, dementia, or psychosis, and 6. Were ≥18 years of age at the time of the visit. Appendix 1 describes the concepts used to define depression, mania, dementia, and psychosis.

The target population was stratified into two populations: those with the outcome (proxy for TRD) and those without. All patients with a procedure code on an inpatient or outpatient medical claims record for electroconvulsive therapy, deep brain stimulation, or vagus nerve stimulation within 7 days of any qualifying visit were classified as having the proxy for TRD. The first such visit where the procedure code was observed was designated as the index date for these cases. For patients without any 'TRD proxy' procedure code, the index date was selected randomly from among the qualifying visits. Appendix 2 describes the concepts used to define electroconvulsive therapy, deep brain stimulation, and vagus nerve stimulation.

2.2 | Data sources

We used three US claims databases: 1. Truven Commercial Claims and Encounters (CCAE), 2. TruvenMarketScan Medicare Supplemental Beneficiaries (MDCR), and 3. OptumInsight's de-identified ClinformaticsTMDatamart (Optum).

CCAE is an administrative health claims database reflecting an employed population and their dependents. It captures person-specific clinical utilization, expenditures, and enrollment across inpatient and outpatient medical services, and outpatient pharmacy dispensings for 127 million subjects.

MDCR is an administrative health claims database for Medicare-eligible active and retired employees and their Medicare-eligible dependents from employer-sponsored supplemental plans. Only plans where boththe Medicare-paid amounts and the employer-paid amounts are included. It has data on over 9 million subjects.

Optum is an administrative health claims database containing people who are fully insured in commercial plans and Medicare. Only members with both medical and prescription drug coverage are included ($n = 74$ million subjects). Since it has data on subjects 65 and older, we partitioned it by age to facilitate assessment of the transportability of the model.

All available years for CCAE, Optum partitioned by age, < 65 and 65 and older, and Medicare only for subjects ≥65 were used.

All the databases were converted to the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) (Stang et al., 2010). One advantage of a standardized format and content is that the same analytic code can be applied to all three databases (Voss

et al., 2015). The standardization of the content is achieved through the implementation of a standard vocabulary with source code mapping. In the OMOP, vocabulary drugs and conditions are referred to by concepts. The OMOP vocabulary provides relationships and ancestry relationships between concepts and extensive mapping to a variety of classification systems (Reich, Ryan, Stang, & Rocca, 2012), so that drugs and conditions can be grouped at specific levels of a hierarchy in a specific classification system. A series of standardized analytic tools have been developed against the OMOP CDM as part of the Observational Health Data Sciences and Informatics (OHDSI) collaborative (Hripcsak et al., 2015).

2.3 | Potential definitions

We created variables that contain information on drug utilization at 3 months, 6 months, and 1 year before the index date. These variables were: number of antidepressants and antipsychotics at the ingredient level, number of antidepressants with adequate doses, and number of antidepressants with adequate duration. Adequate doses were obtained from published literature (Desseilles et al., 2011). We also included the number of psychotherapy sessions since some of the clinical definitions of TRD include not only failure of medications, but also the failure of psychotherapy (Conway et al., 2017). Finally, we included the total number of antidepressants a subject had while in the database.

In addition, we included five expert-based heuristics definitions of TRD developed for use in claims databases. Each definition states that a patient has TRD if ≥2 antidepressant failed. Failure is considered to occur when a new antidepressant or antipsychotic is added (Fife et al., 2017a). What distinguishes these definitions are (Fife et al., 2017b): 1. Differences in the maximum time when a new treatment must begin before the original treatment can be considered a failure (≤90 days, ≤180 days, or no limit); and 2. A treatment can only fail if it has been prescribed at an adequate dose. In all of these definitions, the treatments also had to be present for at least 28 days before being considered a failure.

Appendix 3 describes the ingredients used to define antidepressants and antipsychotics.

2.4 | Analysis

To evaluate which attributes, including both count variables and expert-based heuristic defintions, best discriminates TRD cases from noncases, we fit a decision tree predictive model. A decision tree is a nonparametric method that creates simple decision rules inferred from the data. These rules are easy to understand and apply. An additional advantage over traditional regression models is that it optimally selects cutoffs for continuous variables. In this study, we had numerous continuous variables that needed to be dichotomized. For example, the number of antidepressants the subject had in a year, the decision tree optimized the cutoff at ≤3 antidepressants.

A decision tree for each time period (3 months, 6 months, and 1 year before the index date) was created. Each decision tree was trained using 10 variables (see Table 2) and we limited the number of branches (depth) of the decision tree to two. The decision tree selects the variables that best discriminate between the two groups. The best definition for TRD was simply the rule extracted from the decision tree that achieved the highest discrimination.

To assess whether the model discriminated between subjects with and without evidence of TRD, we calculated the area under the receiver operating characteristic curve (AUC). The higher the AUC the better the model discriminates between the subject with and without evidence of TRD; an AUC of 0.50 means the predictive model is no better than random chance, while an AUC of 1.0 means the model can perfectly discriminate cases from noncases.

We conducted an internal validation using a 20–80 test/train split of the data and an external validation, as a measure of transportability. For the external validation, we applied the best trained model obtained in the CCAE database to the other databases and assessed the AUC. We selected the CCAE database as it had the largest sample of subjects with and without evidence of TRD.

We chose a single metric, the F1 score, to convey and compare the overall performance of the TRD definition obtained by the best model with alternative definitions.The F1 score is the weighted average of the positive predictive value and sensitivity— (Positive predictive value* Sensitivity)*2/ (Positive predictive value+ Sensitivity). The closer the F1 score is to 1, the better the performance of the definition (Powers, 2011). We selected the F1 score over accuracy because accuracy should not be used when there is an uneven distribution of the outcomes being predicted, as in this case where the number of subjects with no evidence of TRD is much larger than the number of subjects with evidence of TRD.

To describe the overall performance of the best model (the best decision rule), we calculated and reported sensitivity and specificity.

2.5 | Sampling

To run the decision tree models efficiently, we randomly sampled subjects with no proxy for TRD at a 10 to 1 ratio to subjects with the proxy for TRD. Cases and noncases were matched on age, gender, time since first antidepressant recorded in the databases to index date and time since first diagnosis of depression in the database to index date.

We used the OHDSI patient level prediction R package for the analysis (Reps, Schuemie, Suchard, Ryan, & Rijnbeek, 2017).

3 | RESULTS

A total of 33,336 subjects had no evidence of TRD, and a total of 3,566 subjects met our definition of definite TRD in the three databases. Most of the subjects with evidence of TRD were women, (Table 1).

The matching successfully balanced age, gender, and time since first diagnosis of depression recorded in the database to index date and time since first antidepressant recorded in the databases to index date between the subjects with no evidence for TRD similar to the subjects with evidence of TRD (Table 1).

The number of antidepressants or number of antidepressants with adequate doses or adequate duration, number of antipsychotics, and

TABLE 1 Description of population in different databases

	CCAE		Optum Young		Optum Old		Medicare	
	Proxy TRD	No TRD	Proxy TRD	No TRD	Proxy TRD	No TRD	Proxy TRD	No TRD
Number of subjects	2076	19981	904	8165	192	1707	394	3483
Age in years	48	48	47	47	73	73	75	75
Males (%)	35.9	35.9	38.5	38.8	33.3	33.2	41.4	41.3
Time since first antidepressant recorded in the databases to index date (days)	1103	1077	1045	1004	1127	1076	1040	1002
Time since first diagnosis of depression recorded in the database to index date (days)	893	867	907	873	942	869	699	651
Total number of antidepressants a subject had while in the database	4.33	2.49	4.10	2.30	4.1	2.4	4.0	2.5

Notes:
TRD: Treatment-resistant depression

number of psychotherapy sessions was greater in subjects with evidence of TRD than in subjects with no evidence of TRD. The magnitude of the difference increases with time, as it is smaller at 3 months than at 12 months. Similarly, as time increases, a higher number of subjects were identified as having evidence of TRD using the expert-based heuristics definitions for TRD (Table 2).

3.1 | Decision tree

A decision tree model was run in the CCAE database for each time period (3, 6, and 12 months). Two variables were consistently selected by the model in all time periods: 1) number of antidepressants and 2) number of antipsychotics.

The AUC increased for longer time periods. We selected the decision tree model at the 12-month period as it was the best model with an AUC of 0.81. This rule states that if a subject ever had ≥1 antipsychotic or ≥3 antidepressants in the last year the subject had TRD (Table 3).

The decision tree rule had a higher F1 score (F1 = 0.44) than the other expert-based heursistic definitions (Table 4).

The specificity of the best decision tree and the the best decision rule was 0.84 and the sensitivity was 0.73.

3.2 | Transportability

The best model in CCAE (12-month time period) was applied to the other databases to assess the transportability. The transportability could change because of differences in the characteristics of the population in each database.

The characteristics of the population in the Optum and Medicare databases are described in Table 4. Similar to CCAE, subjects with evidence of TRD in the Optum and Medicare databases had a larger number of antidepressants, antipsychotics, and psychotherapy sessions (Table 5).

The model transported very well; the AUCs for Optum < 65 was 0.79, for Optum ≥65, and Medicare was 0.78.

3.3 | Rule

Clinically, the decision rule states that a patient with diagnosis of depression, who has received at least one antidepressant and is free of dementia, psychosis, or bipolar disorder will have TRD if she or she has received ≥ 3 antidepressants or ≥ 1 antipsychotic in the last year.

Applying this rule into CCAE database, we found that 15.8% of subjects with depression who are being treated with an antidepressant and had no diagnosis of mania, dementia, and psychosis had TRD.

4 | DISCUSSION

The definition that best discriminates between subjects with and without evidence of TRD in claims databases simply considers the number of distinct antidepressants and antipsychotics the subject has had in the last 12 months: ≥ 3 antidepressants or ≥ 1 antipsychotic. The more complex expert-based heuristic definitions created to discern if an antidepressant is changed because of lack of efficacy, lack of tolerability, inadequate dose, or duration do not discriminate as well.

The data-driven definition not only achieves better performance, but is simpler to understand and implement. The simplicity of the definition facilitates the implementation of the rule by others for future work and facilitates the understanding of the findings as there is less room for subjective and therefore controversial decisions. This simple definition not only has a better discriminatory ability than the complex ones, but still discriminates well when applied to new databases that have different population characteristics such as age. This transportability substantially strengthens the validity of the findings.

Our TRD definition is data driven in that a definition was not imposed on the data; we learned from the data. Our approach responds to calls for having a TRD definition that is evidence based (Conway et al., 2017). Because the rule was based on the data, the subjectivity of defining the parameters in more complex definitions disappears. For these more complex definitions, small variations in the parameters had

TABLE 2 Drug utilization in CCAE

Variable	3 months		6 months		12 months	
	Proxy TRD	No TRD	Proxy TRD	No TRD	Proxy TRD	No TRD
Mean number of antidepressants	1.81	1.15	2.16	1.29	2.62	1.50
Mean number of antipsychotics	0.58	0.07	0.69	0.08	0.83	0.09
Mean number of psychotherapies	0.38	0.17	0.47	0.20	0.58	0.25
Mean number of antidepressants with adequate dose	0.40	0.21	0.69	0.35	1.09	0.55
Mean number of antidepressants with eras of at least 30 days	0.72	0.39	1.24	0.64	1.93	1.03
Definition in which the maximum time within a new treatment must begin before the original treatment can be considered a failure is 90 days (%)	22.3	2.7	35.5	5.3	45.5	8.5
Definition in which the maximum time within a new treatment must begin before the original treatment can be considered a failure is 90 days and only considers a treatment a failure if the dose is adequate (%)	19.9	2.0	32.3	3.7	40.3	5.7
Definition in which no limit is place within a new treatment must begin before the original treatment can be considered a failure (%)	22.3	2.7	41.0	7.2	57.2	14.2
Definition in which the maximum time within a new treatment must begin before the original treatment can be considered a failure is 180 days (%)	22.3	2.7	41.0	7.2	52.7	11.9
Definition in which the maximum time within a new treatment must begin before the original treatment can be considered a failure is 180 days and only considers a treatment a failure if the dose is adequate (%)	19.9	2.0	37.3	5.0	47.0	7.9

Notes:
TRD: Treatment-resistant depression

TABLE 3 Different rules in CCAE by time period

Variable	3 months	6 months	12 months
AUC test	**0.79**	**0.79**	**0.81**
Decision rule for TRD	1. > = 2 AD last 3 months and > = 1 AP (TRD)	1. > = 2 AD last 6 months and > = 1 AP (TRD)	1. > = 2 AD past 12 months and > = 1 AP (TRD)
	2. 0 AP and > = 4 AD ever (TRD)	2. 0 AP and > = 4 AD ever (TRD)	2. 0 AP and > = 3 AD past 12 months (TRD)
	3. 0 AP and < 4 AD ever (non-TRD)	3. 0 AP and < 4 AD ever (non-TRD)	3. 0 AP and < 3 AD past 12 months (non-TRD)
	4. > = 1 AP and < 2 AD past 3 months (TRD)	4. > = 1 AP and < 2 AD past 6 months (TRD)	4. > = 1 AP and < 2 AD past 12 months (TRD)

Notes:
AD: Antidepressant, AP: Antipsychotic, TRD: Treatment-resistant depression

a major impact on the number of subjects who develop TRD (Fife et al., 2017b). To mention just few: 1. Deciding what treatment duration is appropriate before you call a treatment a failure: Is it 4, 6, or 8 weeks? 2. Deciding how long a treatment should be tried before you call it a failure when a new treatment is added: Is it 90 days, 180 days, or is no limit necessary?. 3. Deciding how to deal with treatments of ≤30 days of duration: Is it that the treatment did not work at all or that it was stopped because of lack of tolerability? And 4. Deciding what dose is appropriate when a subject is on more than one antidepressant.

Clinicians face similar problems when trying to decide whether or not the patient they are treating has TRD (Fava, 2003). First, there is subjectivity in deciding whether a treatment has had an adequate duration, what an adequate dose is when the patient is on more than one antidepressant, or how to assess dose adjustment due to tolerability. Second, there are challenges involved in ascertaining whether inadequate response to antidepressant treatment is due to lack of adherence (not taking the medication as prescribed) or true therapeutic failure. Third, it is not clear how to incorporate the impact of patient

TABLE 4 Perfomance comparison of the data-driven definitions and the expert-based heuristic definitions

Definitions of TRD	F1 score
Decision rule-data driven: ≥ 3 antidepressants or ≥ 1 antipsychotic in 1 year	0.44
Definition in which the maximum time within a new treatment must begin before the original treatment can be considered a failure is 90 days	0.40
Definition in which the maximum time within a new treatment must begin before the original treatment can be considered a failure is 90 days and only considers a treatment a failure if the dose is adequate	0.41
Definition in which no limit is placed within a new treatment must begin before the original treatment can be considered a failure	0.39
Definition in which the maximum time within a new treatment must begin before the original treatment can be considered a failure is 180 days	0.39
Definition in which the maximum time within a new treatment must begin before the original treatment can be considered a failure is 180 days and only considers a treatment a failure if the dose is adequate	0.42

Notes:
TRD: Treatment resistant depression

TABLE 5 Drug utilization in optum and medicare 12 months before the index date

	Optum Young		Optum Old		Medicare	
Variable	Proxy TRD	No TRD	Proxy TRD	No TRD	Proxy TRD	No TRD
Mean number of antidepressants	2.50	1.45	2.44	1.39	2.48	1.43
Mean number of antipsychotics	0.74	0.08	0.58	0.06	0.73	0.10
Mean number of psychotherapies	0.47	0.18	0.47	0.12	0.38	0.13
Mean number of antidepressants with adequate dose	1.04	0.58	1.10	0.53	1.08	0.55
Mean number of antidepressants with eras of at least 30 days	1.88	1.03	1.42	0.91	1.84	0.95
Definition in which the maximum time within a new treatment must begin before the original treatment can be considered a failure is 90 days (%)	40.3	8.2	32.3	5.9	39.6	6.9
Definition in which the maximum time within a new treatment must begin before the original treatment can be considered a failure is 90 days and only considers a treatment a failure if the dose is adequate (%)	36.2	5.2	29.7	3.8	35.8	4.7
Definition in which no limit is placed within a new treatment must begin before the original treatment can be considered a failure (%)	54.5	13.3	47.9	10.4	50.3	11.5
Definition in which the maximum time within a new treatment must begin before the original treatment can be considered a failure is 180 days (%)	49.3	11.0	42.2	8.5	45.7	9.7
Definition in which the maximum time within a new treatment must begin before the original treatment can be considered a failure is 180 days and only considers a treatment a failure if the dose is adequate (%)	44.5	7.3	39.6	5.6	41.4	6.5

responses to treatment in previous depression episodes. And fourth, the assessment requires that the patient has good recall of previous treatment regimens. Therefore, the rule we created to identify subjects with TRD could also be used clinically and not only to study TRD in healthcare databases. It is very simple, it only requires the number of antidepressants and antipsychotics taken in the prior year.

The indication of TRD in this study was based on proxies. Although the therapies included as proxies are recognized treatments for TRD,(Health Quality Kolar, 2017; Ontario, 2004) these therapies can also be used for the treatment of patients with severe depression and imminent risk of suicide, who may not have TRD. Furthermore, in clinical practice not all patients who have TRD receive these therapies.

So, the proxy has imperfect specificity and imperfect sensitivity. The impact of having an imperfect proxy is measurement error and likely leads to an underestimation of the rate of TRD.

Since we used electroconvulsive therapy, deep brain stimulation, or vagus nerve stimulation as proxies for TRD, this definition of TRD likely represents stage II TRD, where the degree of resistance warrants more invasive or higher risk antidepressant treatments (Conway et al., 2017). As our model was based on severe cases of TRD, it may impact the generalizability of the model. Nonetheless, we found that 16% of subjects who have depression treated with antidepressants had TRD using our definition, this estimate is within the prevalence range of TRD in subjects with depression suggesting that the model is indeed generalizable.

The published prevalences of TRD can change substantially depending on how TRD is defined and the setting. It varies from 9% when the requirement is a failure to respond to a third antidepressant (Rush et al., 2006a) to 50% when the metric is clinical response to an antidepressant (Fava & Davidson, 1996; Nemeroff, 2007). The prevalence of TRD is lower in primary care settings and higher in tertiary care centers (Nemeroff, 2007). In this study, subjects from both primary and tertiary centers were included. To define TRD we used what is equivalent to treatment failures, the number of treatments.

The subjects in these databases represent families with private insurance and Medicare. Thus, the findings may not be generalizable to subjects with means-tested public insurance.

5 | CONCLUSION

In summary, we created an evidence-based definition that discriminates nicely between subjects with and without proxies for TRD, a rule that is transportable to many other claims databases with different patient characteristics. The decision tree rule states that a subject with pharmaceutically treated depression who has had ≥ 3 distinct antidepressants or ≥ 1 antipsychotic in a year is classified as having TRD. Approximately, 16% of subjects who have depression treated with antidepressants had TRD.

CONFLICT OF INTEREST

All authors are employees of Janssen Research & Development, LLC. Janssen Research & Development, LLC has an interest in depression and treatment resistant depression.

ORCID

M. Soledad Cepeda MD, PhD
http://orcid.org/0000-0002-5159-6217

REFERENCES

Agarwal, V., Podchiyska, T., Banda, J. M., Goel, V., Leung, T. I., Minty, E. P., …Shah, N. H. (2016). Learning statistical models of phenotypes using noisy labeled training data. *Journal of the American Medical Informatics Association*, *23*(6), 1166–1173. https://doi.org/10.1093/jamia/ocw028

Agency for Healthcare Research and Quality, (AHRQ). (2016). Treatment-Resistant Depression: A Narrative and Systematic Review of Definitions and Methods in Clinical Research Studies. *Technology Assessment Protocal*,Retrieved from https://www.ahrq.gov/sites/default/files/wysiwyg/research/findings/ta/topicrefinement/trdepression-protocol-amendment.pdf

Berlim, M. T., & Turecki, G. (2007). What is the meaning of treatment resistant/refractory major depression (TRD)? A systematic review of current randomized trials. *European Neuropsychopharmacology*, *17*(11), 696–707. https://doi.org/10.1016/j.euroneuro.2007.03.009

Blazer, D. G., Kessler, R. C., McGonagle, K. A., & Swartz, M. S. (1994). The prevalence and distribution of major depression in a national community sample: The National Comorbidity Survey. *American Journal of Psychiatry*, *151*(7), 979–986.

Cepeda, M. S., Stang, P., & Makadia, R. (2016). Depression is associated with high levels of C reactive protein and low levels of exhaled nitric oxide: Results of a large population based study on NHANES. *The Journal of Clinical Psychiatry*

Conway, C. R., George, M. S., & Sackeim, H. A. (2017). Toward an Evidence-Based, Operational Definition of Treatment-Resistant Depression: When Enough Is Enough. *JAMA Psychiatry*, *74*(1), 9–10. https://doi.org/10.1001/jamapsychiatry.2016.2586

Desseilles, M., Witte, J., Chang, T. E., Iovien, N., Dording, C. M., Ashih, H., …Mischoulon, D. (2011). Assessing the adequacy of past antidepressant trials: A clinician's guide to the antidepressant treatment response questionnaire. *Journal of Clinical Psychiatry*, *72*(8), 1152–1154. https://doi.org/10.4088/JCP.11ac07225

Fava, M. (2003). Diagnosis and definition of treatment-resistant depression. *Biological Psychiatry*, *53*(8), 649–659. https://doi.org/10.1016/S0006-3223(03)00231-2

Fava, M., & Davidson, K. G. (1996). Definition and epidemiology of treatment-resistant depression. *The Psychiatric Clinics of North America*, *19*(2), 179–200.

Fife, D., Feng, Y., Wang, M. Y., Chang, C. J., Liu, C. Y., Juang, H. T., … Wang, B. (2017a). Epidemiology of pharmaceutically treated depression and treatment resistant depression in Taiwan. *Psychiatry Research*, *252*, 277–283. https://doi.org/10.1016/j.psychres.2017.03.006

Fife, D., Reps, J., Cepeda, M. S., Stang, P., Blacketer, M., & Singh, J. (2017b). Treatment resistant depression incidence estimates from studies of health insurance databases depend strongly on the details of the operating definition. Submitted. Unpublished data

Health Quality Ontario. (2004). Repetitive transcranial magnetic stimulation for the treatment of major depressive disorder: An evidence-based analysis. *Ontario Health Technology Assessment Series*, *4*(7), 1–98.

Hripcsak, G., Duke, J. D., Shah, N. H., Reich, C. G., Huser, V., Schuemie, M. J., … Ryan, P. B. (2015). Observational Health Data Sciences and Informatics (OHDSI): Opportunities for observational researchers. *Studies in Health Technology and Informatics*, *216*, 574–578.

Kessler, R. C. (2012). The costs of depression. *The Psychiatric Clinics of North America*, *35*(1), 1–14.

Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*(6), 617–627. https://doi.org/10.1001/archpsyc.62.6.617

Kolar, D. (2017). Current status of electroconvulsive therapy for mood disorders: A clinical review. *Evidence-Based Mental Health*, *20*(1), 12–14. https://doi.org/10.1136/eb-2016-102498

Kubitz, N., Mehra, M., Potluri, R. C., Garg, N., & Cossrow, N. (2013). Characterization of treatment resistant depression episodes in a cohort

of patients from a US commercial claims database. *Plos One*, *8*(10), e76882.https://doi.org/10.1371/journal.pone.0076882

Mathew, S. J. (2008). Treatment-resistant depression: Recent developments and future directions. *Depression and Anxiety*, *25*(12), 989–992. https://doi.org/10.1002/da.20540

Nemeroff, C. B. (2007). Prevalence and management of treatment-resistant depression. *Journal of Clinical Psychiatry*, *68 Suppl8*, 17–25.

Powers, D. M. W. (2011). Evaluation: From precision, recall and f-measure to ROC, informedness, markedness& correlation. *Journal of Machine Learning Technologies*, *2*(1), 37.

Pratt, L. A., & Brody, D. J. (2008). Depression in the United States household population, 2005–2006. *NCHS Data Brief*, (7), 1–8.

Reich, C., Ryan, P. B., Stang, P. E., & Rocca, M. (2012). Evaluation of alternative standardized terminologies for medical conditions within a network of observational healthcare databases. *Journal of Biomedical Informatics*, *45*(4), 689–696. https://doi.org/10.1016/j.jbi.2012.05.002

Reps, J., Schuemie, M., Suchard, M. A., Ryan, P. B., & Rijnbeek, P. (2017). Patient level prediction: Package for patient level prediction using data in the OMOP Common Data Model.

Rush, A. J., Trivedi, M. H., Wisniewski, S. R., Nierenberg, A. A., Stewart, J. W., Warden, D., … Fava, M. (2006a). Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: A STAR*D Report. *American Journal of Psychiatry*, *163*(11), 1905–1917. https://doi.org/10.1176/ajp.2006.163.11.1905

Rush, A. J., Trivedi, M. H., Wisniewski, S. R., Stewart, J. W., Nierenberg, A. A., Thase, M. E., … Team Star D. Study (2006b). Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs for depression. *New England Journal of Medicine*, *354*(12), 1231–1242. https://doi.org/10.1056/NEJMoa052963

Russell, J. M., Hawkins, K., Ozminkowski, R. J., Orsini, L., Crown, W. H., Kennedy, S., …Rush, A. J. (2004). The cost consequences of treatment-resistant depression. *Journal of Clinical Psychiatry*, *65*(3), 341–347.

Stang, P. E., Ryan, P. B., Racoosin, J. A., Overhage, J. M., Hartzema, A. G., Reich, C., … Woodcock, J. (2010). Advancing the science for active surveillance: Rationale and design for the Observational Medical Outcomes Partnership. *Annals of Internal Medicine*, *153*(9), 600–606. https://doi.org/10.7326/0003-4819-153-9-201011020-00010

Trivedi, M. H., Rush, A. J., Wisniewski, S. R., Nierenberg, A. A., Warden, D., Ritz, L., … Team Star D. Study (2006). Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: Implications for clinical practice. *American Journal of Psychiatry*, *163*(1), 28–40. https://doi.org/10.1176/appi.ajp.163.1.28

Voss, E. A., Makadia, R., Matcho, A., Ma, Q., Knoll, C., Schuemie, M., …Ryan, P. B. (2015). Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. *Journal of the American Medical Informatics Association*, *22*(3), 553–564. https://doi.org/10.1093/jamia/ocu023

How to cite this article: Cepeda MS, Reps J, Fife D, Blacketer C, Stang P, Ryan P. Finding treatment-resistant depression in real-world data: How a data driven approach compares with expert-based heuristics. *Depress Anxiety*. 2018;35:220–228. https://doi.org/10.1002/da.22705

APPENDIX 1: SNOMED CONCEPTS USED IN THE STUDY TO DEFINE DEPRESSION, DEMENTIA, MANIA, AND PSYCHOSIS

Depression	Dementia	Mania	Psychosis
Adjustment disorder with depressed mood	Alcohol amnestic disorder	Bipolar disorder	Delusional disorder
Depressive disorder	Amnesia	Mania	Psychotic disorder
Dysthymia	Amnestic disorder	Recurrent manic episodes	Schizophrenia
Adjustment disorder with depressed mood	Cerebral degeneration associated with generalized lipidosis		
Depressive disorder	Cerebral degeneration in childhood		
Dysthymia	Cerebral lipidosis		
	Communicating hydrocephalus		
	Degenerative brain disorder		
	Huntington's chorea		
	Normal pressure hydrocephalus		
	Obstructive hydrocephalus		
	Organic mental disorder		
	Organic personality disorder		
	Presbyophrenic psychosis		

APPENDIX 2: SNOMED CONCEPTS USED TO DEFINE ELECTRO-CONVULSIVE THERAPY, DEEP BRAIN STIMULATION, AND VAGUS NERVE STIMULATION

Electroconvulsive therapy
Neurostimulation of brain tissue
Cranial nerve neurostimulator electrode procedures

APPENDIX 3: LIST OF ANTIDEPRESSANTS AND ANTIPSYCHOTICS

Antipsychotics	Antidepressants	Antidepressants	Antidepressants	Antidepressants	Antidepressants
Aripiprazole	Amitriptyline	Doxepin	Levomilnacipran	Protriptyline	Trimipramine
Clozapine	Amoxapine	Duloxetine	Maprotiline	Reboxetine	Venlafaxine
Olanzapine	Bupropion	Escitalopram	Milnacipran	Selegiline	Vilazodone
Paliperidone	Citalopram	Fluoxetine	Mirtazapine	Sertraline	Vortioxetine
Quetiapine	Clomipramine	Fluvoxamine	Moclobemide	Tianeptine	
Risperidone	Desipramine	Imipramine	Nortriptyline	Tranylcypromine	
Ziprasidone	Desvenlafaxine	Isocarboxazid	Paroxetine	Trazodone	