Submitted to: Expert Review in Neurotherapeutics

How depression heterogeneity hinders progress in treatment and research

Eiko I. Fried
University of Amsterdam

Abstract

Has depression research in the last two decades led to major improvements of clinical care for patients? Pharmacological innovations have not resulted in considerable progress, and although the biology of Major Depression (MD) is among the most-studied and best-funded topics, answers remain elusive. Why is that? The current editorial argues that depression heterogeneity is at the heart of the slow progress, and suggests that a novel research framework may provide an inroad to new insights. Most researchers sum disparate depression symptoms such as sadness, insomnia, and concentration problems to one sumscore representing depression severity, and use thresholds to diagnose MD. This leads to highly heterogeneous depressed samples in which patients often have few symptoms in common. Recent work suggests that this unaddressed heterogeneity explains the lack of research progress. A novel research framework—Depression Symptomics—may offer a way forward by focusing on the study of individual depression symptoms, their causal interactions, and by paying attention to differences among patients diagnosed with MD.

Introduction

Few would argue that depression research in the last two decades has led to major improvements of clinical care for patients. The tools and insights we have at our disposal today have not changed much over the years. Three examples for slow progress are antidepressants, depression assessment, and research on biological markers. Drugs most commonly prescribed for Major Depression (MD) were developed in the 80s and 90s. They only slightly outperform placebos, and their exact mechanisms remain unknown [1]. Hamilton's rating scale for depression (HRSD)—still the gold-standard in clinical trials—is nearly 60 years old [2]. And the study of MD biology has largely resulted in null-findings or small effects that are impracticable for clinical purposes, despite ever-growing samples [3,4].

A growing chorus of voices—including prominent members of the DSM-5 task force and the NIMH [5,6]—have raised concerns about the reliability and validity of MD, concluding that the clinical phenotype may hold little usefulness for research [7,8]. What exactly are the problems, and what can we do better?

Depression sum-scores don't add up

Nearly all depression research comprises several steps. A specific rating scale or clinical interview is selected among a large number of instruments [9], and symptoms such as sad mood, worthlessness, suicidal ideation, weight gain or weight loss, insomnia or hypersomnia, psychomotor problems, anxiety, genital problems, paralysis, crying, hypochondriasis, and lack of insight are assessed [10]. These diverse symptoms are added up to one sum-score that represents depression severity, and a threshold on the sum-score is used to distinguish individuals with MD from healthy controls. Researchers then investigate whether variables of interest such as biological markers or treatment response are statistically related to the sum-score, or compare patients and controls regarding such variables. Three lines of research demonstrate that this approach may obfuscate insights and hinder progress.

First, MD is a highly heterogeneous diagnosis, and two patients with a DSM-5 diagnosis may share no single symptom. Alice comes into the practice as 10am appointment reporting sadness, weight loss, psychomotor retardation, concentration problems, and suicidal ideation. Bob—our 11am appointment—suffers from anhedonia, insomnia, self-blame, fatigue, and weight gain. A recent paper identified 1,030 unique symptom profiles in 3,703 depressed patients [11], and it seems questionable to investigate risk factors, biological markers or other variables of interest in a depressed population when patients differ so considerably in their problems.

Second, it is astonishing how fuzzy the definition of MD is. The most common rating scales feature over 50 disparate depressive symptoms [10], and many scales have only little content overlap with the DSM-5 MD criteria. Such a situation is difficult to envision for medical conditions like cancer or measles. The lack of a clear definition is also apparent in the fact that many different scales exist that aim to measure the same disorder—280 by the count of Santor et al. [9]. Moreover, dozens of MD subtypes have been proposed and disregarded over the last century.

Third, the approach of summing disparate symptoms presupposes that all symptoms measure the same underlying construct, the same way that 10 math questions measure the same underlying construct "mathematical intelligence", or 10 symptoms of cancer indicate the same underlying condition cancer. For depression, the notion that one sum-score of symptoms is a good proxy for severity is inconsistent with half a century of psychometric literature: depression rating scales are not unidimensional, a psychometric fact that cannot properly be

reflected in one sum-score [12]. Nearly 60 years ago, Hamilton already insisted in his seminal 1960 paper that the sum-score be regarded only as "total crude score". His case report of 7 patients provides detailed information on four depression factors such as retarded depression and anxiety reaction, omitting information on the total crude score of patients that carries little information [2].

In sum, lumping disparate symptoms to a sum-score—and lumping patients with very different symptoms into one category—result in loss of important information. These common research practices have contributed to the lack of progress in depression research.

Moving forward: Depression Symptomics

The novel research framework *Depression Symptomics* promises a way forward. Depression Symptomics aims to tackle widely acknowledged shortcomings of sum-scores and depression diagnosis by focusing on three objectives: (1) studying individual symptoms, (2) the causal relations among symptoms, and (3) personalized processes of patients.

1. Studying individual symptoms

Recent work has shown that specific MD symptoms differ from each other in crucial aspects: They differ in their impact on impairment of functioning, their response to specific life events, their relations with biological markers, and have different risk factors [13,14]. There is also evidence that symptoms differ in their response to antidepressant treatment [15]—which makes sense when we consider that the most common side-effects of antidepressants are the very symptoms used to measure depression (e.g., fatigue, insomnia or hypersomnia, weight and appetite changes, and sexual dysfunction). When investigating changes of symptom sum-scores over time in clinical trials, patients' improvement on some symptoms is likely concealed by increases in other symptoms.

This provides further evidence that sum-scores conceal differential effects, and suggest that symptom-based investigations may offer potential for genetic and neuroimaging studies that have so far largely resulted in disappointing findings. The same holds for clinical trials: We should study what *specific* symptoms improve and worsen in response to specific antidepressants. This may enable us to predict based on baseline symptom-profiles which participants will likely respond positively to what kind of drug.

2. Studying associations among symptoms

Symptoms of depression can interact with each other in complex dynamical systems, such as sadness → insomnia → fatigue → concentration problems. It is widely acknowledged that depressed patients are often trapped in situations of reinforcing problems. Such causal influences among symptoms are ignored in the study of sum-scores, and have recently become a topic of detailed study in an emerging field termed "network approach to psychopathology" [16].

Specifically, studies have shown that certain depression symptoms seem to be more relevant than others in symptom networks of patients, that symptom networks of depressed people show different characteristics than those of controls, and that symptom networks may emit so-called 'early warning signals' before healthy people transition into depression (review [16]). This network approach opens up new possibilities of timely prevention and intervention.

3. Personalized medicine

Recent technological developments such as smart-watches and other wearable devices allow for daily monitoring of depressed patients [17], and new statistical models enable us to estimate

idiographic and nomothetic processes at the same time [18]. Personalized medicine—paying closer attention to differences among patients—may offer an important inroad to overcome limitations of the highly heterogeneous depression phenotype. Different fields of application come to mind: Improved prediction of depression onset in samples at-risk of developing depression, improved prediction of depression relapse, and improved prediction of treatment [19].

Future outlook

MD diagnosis provides little clinical utility: It lacks treatment specificity, a clear clinical presentation, and precise diagnostic boundaries, and has high comorbidity rates and a very low inter-rater reliability [7,8]. Head of the DSM-5 task force David Kupfer concluded that "the relatively low reliability of major depressive disorder [...] is a concern for clinical decision-making" [5], and prior NIMH director Thomas Insel stated that the DSM's main weakness is its "lack of validity" [6].

Depression Symptomics conceptualizes individual symptoms as fundamental building blocks of mental disorders that provide an untapped source of important and clinically relevant information. This is consistent with the NIMH strategic plan for mood disorder research and the NIMH Research Domain Criteria (RDoC) that aim to study and identify more homogeneous and reliable endophenotypes [6,20]. Embracing complexity was a key requirement to recent insights in biology, economics, and environmental sciences, and Depression Symptomics may change our perspective of MD from an unwieldy and cumbersome disorder to a phenotype where attention to specific symptoms and their interactions—and attention to differences between patients—are crucial for scientific and clinical progress.

1. Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. PLoS Med. 2008;5: e45. doi:10.1371/journal.pmed.0050045
2. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry. 1960;23: 56–62. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=495331&tool=pmcentrez&rendertype=abstract
3. Cai N, Bigdeli TB, Kretzschmar W, Li Y, Liang J, Song L, et al. Sparse whole-genome sequencing identifies two loci for major depressive disorder. Nature. 2015;523: 588–91. doi:10.1038/nature14659
4. Fried EI, Kievit RA. The volumes of subcortical regions in depressed and healthy individuals are strikingly similar: a reinterpretation of the results by Schmaal et al. Mol Psychiatry. Nature Publishing Group; 2016;21: 724–725. doi:10.1038/mp.2015.199
5. Kupfer DJ, Kraemer HC. Field Trial Results Guide DSM Recommendations. In: Huffington Post [Internet]. 2013. Available: http://www.huffingtonpost.com/david-j-kupfer-md/dsm-5_b_2083092.html
6. Insel TR. Transforming Diagnosis. In: National Institute of Mental Health [Internet]. 2013. Available: http://www.nimh.nih.gov/about/director/2013/transforming-diagnosis.shtml
7. Fried EI. Problematic assumptions have slowed down depression research: why symptoms, not syndromes are the way forward. Front Psychol. 2015;6: 1–11. doi:10.3389/fpsyg.2015.00309
8. Parker G. Beyond major depression. Psychol Med. 2005;35: 467–74. Available: http://www.ncbi.nlm.nih.gov/pubmed/15856717
9. Santor DA, Gregus M, Welch A. Eight Decades of Measurement in Depression. Measurement. 2006;4: 135–155. doi:10.1207/s15366359mea0403
10. Fried EI. The 52 symptoms of major depression. J Affect Disord. Elsevier; 2017;208: 191–197. doi:10.1016/j.jad.2016.10.019
11. Fried EI, Nesse RM. Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. J Affect Disord. Elsevier; 2015;172: 96–102. doi:10.1016/j.jad.2014.10.010
12. Fried EI, van Borkulo CD, Epskamp S, Schoevers RA, Tuerlinckx F, Borsboom D. Measuring Depression Over Time . . . or not? Lack of Unidimensionality and Longitudinal Measurement Invariance in Four Common Rating Scales of Depression. Psychol Assess. 2016;28: 1354–1367. doi:10.1037/pas0000275
13. Fried EI, Nesse RM. Depression sum-scores don't add up: why analyzing specific depression symptoms is essential. BMC Med. 2015;13: 1–11. doi:10.1186/s12916-015-0325-4
14. Jokela M, Virtanen M, Batty GD. Inflammation and specific symptoms of depression. JAMA psychiatry. 2016;73: 1–6. doi:10.1001/jamapsychiatry.2015.1977
15. Hieronymus F, Emilsson JF, Nilsson S, Eriksson E. Consistent superiority of selective serotonin reuptake inhibitors over placebo in reducing depressed mood in patients with major depression. Mol Psychiatry. Nature Publishing Group; 2015; 1–8. doi:10.1038/mp.2015.53
16. Fried EI, van Borkulo CD, Cramer AOJ, Boschloo L, Schoevers RA, Borsboom D. Mental disorders as networks of problems: a review of recent insights. Soc Psychiatry Psychiatr Epidemiol. 2016;1: 1–32. doi:10.1007/s00127-016-1319-z
17. Bos FM, Schoevers RA, aan het Rot M. Experience sampling and ecological momentary assessment studies in psychopharmacology: A systematic review. Eur Neuropsychopharmacol. 2015;25: 1853–1864. doi:10.1016/j.euroneuro.2015.08.008

18. Schuurman NK, Ferrer E, de Boer-Sonnenschein M, Hamaker EL. How to Compare Cross-Lagged Associations in a Multilevel Autoregressive Model. Psychol Methods. 2016;21: 206–221. doi:10.1037/met0000062
19. Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. The Lancet Psychiatry. Elsevier Ltd; 2016;366: 1–8. doi:10.1016/S2215-0366(15)00471-X
20. National Institute of Mental Health. Breaking ground, breaking through: The Strategic Plan for Mood Disorders Research (NIH Publication No. 03–5121). Washington, DC: National Institutes of Health; 2003.