Information theory methods for quantifying diagnostic heterogeneity in psychopathology

John F. Buss (bussj@iu.edu)
Indiana University Bloomington
Ashley L. Watts
University of Missouri
Lorenzo Lorenzo-Luaces
Indiana University Bloomington

Research Article

Keywords: Depression, Classification, Melancholia, Atypical

Posted Date: January 20th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2472751/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Psychiatry on November 30th, 2023.
See the published version at https://doi.org/10.1186/s12888-023-05377-5.

RESEARCH

Information theory methods for quantifying diagnostic heterogeneity in psychopathology

John F. Buss[1*], Ashley L. Watts[2] and Lorenzo Lorenzo-Luaces[1]

[*]Correspondence: bussj@iu.edu
[1] Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405, USA
[2] Department of Psychological Sciences, University of Missouri, Columbia, MO 65211, USA

Objectives: Although specifiers for a major depressive disorder (MDE) are supposed to reduce diagnostic heterogeneity, recent literature challenges the idea that the atypical and melancholic features identify more homogenous or coherent subgroups. We attempt to replicate these findings and explore whether symptom heterogeneity is reduced in depression subgroups using novel data-analytic techniques.

Methods: Using data derived from the National Epidemiological Survey on Alcohol and Related Conditions (NESARC Wave I; N = 5,749) and Sequenced Treatment Alternatives to Relieve Depression (STAR*D; N = 2,498) we computed the Hamming and Manhattan distance ratios comparing within and between individuals for the melancholic and atypical specifier subgroups.

Results: In neither of the datasets was the heterogeneity between-subgroups higher than the heterogeneity within-subgroups, suggesting that the melancholic and atypical specifiers do not create more coherent (i.e., more homogeneous) subgroups.

Conclusion: Replicating prior work, melancholic and atypical depression subtypes appear to have limited utility in reducing heterogeneity. The current study does not support the claim that symptom and course specifiers create more coherent subgroups as operationalized by similarity in symptoms and their severity.

Keywords: Depression; Classification; Melancholia; Atypical

Background

A major depressive episode (MDE) involves a combination of symptoms, including fatigue, diminished ability to concentrate, and sadness [1]. To meet the diagnostic criteria for an MDE, an individual must present five of nine possible symptoms for two weeks, and at least one symptom must be sadness or anhedonia. Some symptoms of an MDE can be met by reporting different qualitative complaints (e.g., symptom six can be met by reporting either fatigue or loss of energy), whereas other symptoms can be met by reporting complaints that differ in severity (e.g., symptom nine can be met by reporting recurrent thoughts of death or by attempting to commit suicide). Still, some items represent opposites of a behavior (e.g., symptom five can be met by psychomotor agitation, but it can also be met by psychomotor retardation, and symptom four can be met by insomnia or hypersomnia). Thus, using polythetic criteria for an MDE leads to highly heterogeneous symptom presentations to the point that two individuals with an MDE may not share a single symptom [2, 3]. Researchers often quantify diagnostic heterogeneity in symptom presentations by counting the number of unique combinations of symptoms possible and reported. For example, in the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) sample of N = 3,703 outpatients, there were 1,030 combinations of depression symptoms [2, 4].

The Diagnostic and Statistical Manual for Mental Disorders 5th edition (DSM-5) uses specifiers for depression and other diagnostic subgroups (APA, 2013). According to the DSM-5, individuals who share specifier features are more alike than individuals who do not share the specifier features and thus create "more homogeneous" subgroups. However, recent research suggests that specifiers do not create more homogeneous subgroups [5,6]. In particular, because specifiers for MDE subgroups classify individuals by adding polythetic features to the DSM criteria, the specifier subgrouping may create more *heterogeneous* subgroups. In an analysis

of heterogeneity in the melancholic and atypical specifiers in the STAR*D data, a large sample of outpatients, reduction in heterogeneity when comparing subgroups that met for a specifier (e.g., melancholic features) with those that did not (e.g., non-melancholic features) was not significantly different from zero and appeared driven by smaller sample sizes in the specifier subgroups [5].

Although the findings of Lorenzo-Luaces et al. [5] are consistent with the computational logic previously presented by Fried et al., they have not yet been replicated. Moreover, Lorenzo-Luaces et al. quantified heterogeneity using a very strict approach where two individuals were considered to have heterogeneous symptom presentations if they differed on only one symptom. This strict approach has been used in several studies [3, 7], but it treats heterogeneity as a binary variable (i.e., individuals are either the same or they are not), which is inconsistent with emerging perspectives that conceptualize and measure psychopathology continuously.

Given the limitations associated with previous studies, we sought to replicate the findings of Lorenzo-Luaces et al. [5], namely that the specifier subgroups do not reduce heterogeneity, using a large nationally-representative sample of adults (n = 5,749). Rather than rely on a simple, binary metric that indicates whether or not diagnostic combinations were 100% identical, we borrowed concepts from information theory to quantify the extent of heterogeneity on a continuum. Specifically, we used distance metrics (i.e., Hamming and Manhattan distances) in Euclidean space to characterize the relative similarity of diagnostic combinations. We hypothesized that atypical and melancholic specifiers would not reduce heterogeneity relative to depressed individuals not meeting specifier criteria. Additionally, we reanalyzed the STAR*D data (n = 2,496) to explore whether previous results were driven by the fact heterogeneity was operationalized categorically as opposed to continuously. We refer to a *diagnostic combination* as any set

of symptoms defined in the DSM-5, such that an individual meets criteria for an MDE. *Coherence* is the amount of within-group homogeneity, where greater coherence indicates greater homogeneity within a given subgroup. In contrast to coherence, we use the term *distance* to refer to the degree of heterogeneity as measured by Euclidean dimensional functions. We define *differentiation* as the ability of subgroup diagnostic criteria to define subgroups with markedly different diagnostic combinations (i.e., coherence within subgroup and greater distance between subgroups).

Methods

NESARC

We analyzed the public-access dataset from the NIAAA-supported National Epidemiologic Survey on Alcohol and Related Conditions (NESARC), Wave I study [8]. The NESARC was a nationally administered survey of adults 18 years or older (N = 43,093) who were interviewed face-to-face using the Alcohol Use Disorder and Associated Disabilities Interview Schedule-DSM-IV (AUDADIS-IV). The NESARC sampled sociodemographic subgroups to ensure that the sample was sufficiently representative of the US population (e.g., Hispanic, Non-Hispanic Black, and young adults) with a response rate of 81%. From the total number of respondents, 7,839 met criteria for an MDE in their lifetimes. Participants were excluded from the analyses if they A) met criteria for mania or hypomania (n = 725), or B) their worst episode experienced was deemed illness or substance-induced (n = 715). After exclusion criteria were applied, 6,448 MDE cases (82.3%) remained. From this pool, participants that had missing depression symptom data were listwise deleted, leading to a final count of n = 5,749 participants (73.3%). In the NESARC, participants reported symptoms on their *worst depressive episode* within their *lifetime*. Thus data were drawn from episodes over the course of the participant's lifetime.

STAR*D

We also re-analyzed the Sequenced Treatment Alternatives to Relieve Depression (STAR*D; [9]). The STAR*D is a multi-site sequentially randomized clinical trial of 4,041 outpatients who were diagnosed with major depressive disorder (MDD) and unable to attain a satisfactory response following selective serotonin reuptake inhibitor (SSRI) treatment. Inclusion criteria included being between the ages of 18 and 75 and a diagnosis of DSM-IV unipolar and non-psychotic MDD. Exclusion criteria included a history of mania or hypomania, schizophrenia, schizoaffective disorder or psychosis, or current anorexia, bulimia, or obsessive-compulsive disorder (OCD) as assessed by the Psychiatric Diagnostic Screening Questionnaire (PDSQ) via clinical interview [10]. Depressive symptoms, including melancholic and atypical symptoms, were screened using the Inventory of Depressive Symptomatology (IDS-SR). For more information regarding the study design, please refer to the following studies [4, 9]. The original sample had data available for 4,041 patients. Of these patients, 3744 (92.7%) provided baseline data during the first measurement point of the first treatment stage. We screened out patients who did not have full symptom-level IDS data, leading to 3,717 patients (91.9%). Inclusion into the original trial required patients to meet criteria for non-psychotic MDD based on the DSM-IV. To ensure consistency, patients were screened for meeting an MDE based on the IDS, leading to $n = 2{,}496$ remaining patients (61.8%). Patients were queried on specific symptoms based on their *current* depressive episode. Thus, we derived diagnostic combinations from the STAR*D patients' current depressive episode.

Outcomes

Alcohol Use Disorder and Associated Disabilities Interview Schedule (AUDADIS-IV). In NESARC, the AUDADIS-IV [11] measures 19 symptoms of depression that are rated as either 'present' or 'absent' and coded as "1" or "2", respectively. The AUDADIS-IV covers DSM-IV criteria symptoms in

a disaggregated form. For example, it queries *both* psychomotor agitation and psychomotor retardation, whereas the DSM-IV codes psychomotor disturbances as a single symptom. In what follows, we describe our decision-making process regarding symptom inclusion in the NESARC dataset. See Lorenzo-Luaces et al. (2021; [5]) for a description of how the STAR*D symptoms were parsed.

Appetite or weight disturbances. The AUDADIS-IV contains four questions querying appetite, or weight disturbances: 1) reduced appetite, 2) reduced weight, 3) increased appetite, and 4) increased weight. To prevent over-estimating the degree of heterogeneity in the data, we combined the responses to the appetite and weight questions, thus creating two variables: 1) *decreased* appetite/weight and 2) *increased* appetite or weight. For decreased appetite/weight, we considered the person to meet the symptom whether they reported decreased appetite, decreased weight, or both. Similarly, for increased appetite or weight, we considered the person to meet the symptom whether they reported increased appetite, increased weight, or both.

Suicidal ideation. The AUDADIS-IV contains four questions pertaining to suicide: 1) death ideation (i.e., thoughts of death), 2) desire to die, 3) suicidal ideation (i.e., thoughts about killing oneself), and 4) attempted suicide. We distinguished suicidal attempts from thoughts by combining the responses to the first three questions (i.e., death ideation, desire to die, and suicidal ideation) into a symptom indicating the presence of suicidal thoughts. A person was considered to have suicidal thoughts if they expressed death ideation, desire to die, suicidal ideation, or some combination of these symptoms.

Restlessness and psychomotor agitation. The AUDADIS-IV queries an uncomfortable feeling of restlessness as well as symptoms of fidgeting and pacing as proxies for psychomotor agitation. We removed the 'feelings of restlessness' symptom when performing the analyses, as subjective feelings of restlessness do

not count towards the presence of psychomotor agitation per the DSM-5 (American Psychiatric Association, 2013).

Melancholic and atypical specifiers. The AUDADIS-IV does not query all the symptoms of melancholic and atypical depression. We categorized melancholic depression as having three symptoms from a list that included: anhedonia, psychomotor retardation/agitation, guilt, early morning awakenings, or significant weight loss. Comporting to previous NESARC analyses [12], the atypical subgroup consisted of respondents who met criteria for both hypersomnia and hyperphagia. To fully mimic the criteria for atypical depression, respondents were not classified as having an atypical specifier if they reported symptoms of anhedonia. The hierarchical rule of specifiers was applied: Participants meeting criteria for a melancholic specifier could not then meet criteria for an atypical specifier (see appendix for a list of queried symptoms and criteria rules). The STAR*D dataset used the IDS to query for all depressive symptoms, including those for the melancholic and atypical specifiers. Thus, we adhered to the DSM-5's criteria for melancholic and atypical specifiers.

Analytic strategy

Similar to previous analyses [5], we divided the NESARC and STAR*D datasets into five subgroups corresponding to the presence of melancholic and atypical specifier subgroups, as shown in Figure 1. Because we respected the hierarchical rule from DSM-5, all participants were screened for the presence of melancholia first, creating melancholic and non-melancholic subgroups. Then, all participants in the "non-melancholic" group were grouped into atypical vs. non-atypical subgroups.

All data were analyzed using the R programming language. All code is available at: https://osf.io/vh5qg/. Two functions from information theory, known as the Hamming and Manhattan distances, were used [13, 14]. As shown in Equation 1, the Hamming formula is a way to measure distance in a multivariate space given two

binary data strings (i.e., data containing only 0s and 1s). For every specific symptom that is not shared between any two diagnostic combinations, the Hamming distance between the diagnostic combinations will increase by 1. Since all participants in both datasets were represented by combinations containing binary data, we were able to calculate the distance between diagnostic combinations.

Equation 1: Hamming Distance Function

$$(1)\quad D_H = \sum_{i=1}^{k} |x_i - y_i|$$

$$(2)\quad x = y \Rightarrow D = 0$$

$$(3)\quad x \neq y \Rightarrow D = 1$$

$$(4)\quad R_H = \frac{D_H}{\sum_{i=1}^{k} z_i}$$

Similar to the Hamming distance, the Manhattan distance quantifies the distance between two points in an N-dimensional vector space. For our applications, Manhattan distance allows us to quantify distance in kind (i.e., symptom present vs. absent) as well as intensity (i.e., mild vs. severe presentations of the same symptom: see equation 2). A higher Manhattan distance between the diagnostic combinations of two individuals indicates a greater dissimilarity between them in the severity and kinds of symptoms. The Manhattan distance is not equivalent to a total sum score (TSS). Two combinations of symptoms can have equal TSSs and different Manhattan distances (see Appendix). The NESARC data-set lacked a continuous measure of severity for the depressive symptoms, thus a manhattan distance was only calculated for the STAR*D data-set.

Equation 2: Manhattan Distance Function

$$(1)\ D_M = \sum_{i=1}^{k} |x_i - y_i|$$

$$(2)\ \ R_M = \frac{D_M}{\sum_{i=1}^{k} vz_i}$$

To simplify interpretation, all distance measures were converted to ratios by dividing distance values by the length of the symptom space or the total number of symptoms queried, represented as z_i in Equation 1.4 and Equation 2.2. For the Manhattan ratio, all values were also divided by the maximum possible severity represented by v in Equation 2.2, or three, in this case, based on the maximum symptom severity value of the IDS.

Four separate sets of analyses were conducted. The first analysis used the NESARC dataset represented in Figure 4 to calculate Hamming distances for each specifier subgroup, where all symptoms were treated as qualitatively distinct. A second analysis using the STAR*D dataset calculated the Hamming distance ratios, including the specifier symptoms; see Figure 3. A third analysis calculated the Hamming distance using the STAR*D data set again but only counted the core depressive symptoms for each subgroup (i.e., without specifier symptoms); see Figure 3. The fourth analysis calculated the Manhattan distance ratios in the STAR*D dataset, where symptom severity was evaluated, for each of the four subgroups.

For each analysis, we calculated the within-subgroup and between-subgroup distance ratio. Within-subgroup calculations consisted of comparing each diagnostic combination to each other diagnostic combination. For example, when evaluating Subgroup "A", a diagnostic combination C_{a1} was compared to diagnostic combination C_{a2}, C_{a3}, ... C_{an}. Similarly, diagnostic combination C_{a2} was compared to C_{a3},

C_{a4}, ... C_{an}. With this method, each diagnostic combination was compared to the other diagnostic combination once. A distance was calculated between every other diagnostic combination within that subgroup and stored into a vector containing all calculated distances.

Between-subgroup distance calculations compared diagnostic combinations within a subgroup to each combination of participants not meeting subgroup criteria (e.g., non-atypical profiles with atypical profiles). For example, when evaluating Subgroup "A" against those not meeting criteria in Subgroup "B", a diagnostic combination C_{a1} was compared to diagnostic combination C_{b1}, C_{b2}, ... C_{bn}. Similarly, diagnostic combination C_{a2} was compared to C_{b1}, C_{b2}, ... C_{bn}. Each diagnostic combination in one subgroup was compared to every other diagnostic combination to those not meeting subgroup criteria once (e.g., atypical vs. non-atypical). A distance was computed for each pairing and then stored into a vector containing all distances.

Due to the size of the datasets, the within-group and between-group vectors of distances comprised millions of data points. Thus, we illustrate all analyses using box plots to avoid overcrowding the data. Two example boxplots using theoretical data are provided in Figure 2, demonstrating how within-subgroup and between-subgroup analyses may be interpreted. Panel A shows an ideal case of subgroup coherence and differentiation (i.e., where subgroups show the maximal distance between diagnostic combinations). Subgroup 1 and Subgroup 2 are approaching pure coherence, as the distance ratios approach 0; simultaneously, the two subgroups appear to be distinct, having high differentiation as the between-subgroup ratio approaches 1.

In contrast, Figure 2 Panel B displays a case of complete heterogeneity, where diagnostic combinations were generated randomly. Both within- and between-subgroup analyses exhibit nearly identical distance ratios. The between-subgroup ratio indicates that the subgroups are low in differentiation (i.e., the diagnostic

profiles in both subgroups are similar to each other), whereas the identical within-subgroup ratios indicate both subgroups are similarly heterogeneous.

Results

Symptom Endorsement

Table 1 shows the descriptive statistics representing the binary endorsement of symptoms (i.e., yes vs. no) criteria for a DSM-IV MDE within the NESARC dataset. A table of symptom endorsement for the STAR*D dataset representing the presence or absence of symptoms in the patients meeting criteria for an IDS-MDE can be found in the Appendix. In NESARC, sad mood and (94.97%) and anhedonia (87.60%) were the most frequently reported symptoms. The least endorsed symptoms were suicide attempt (11.05%) and appetite/weight increase (36.42%).

Of the individuals in the subset of the NESARC data we used, 2,384 (41.46%) met criteria for melancholic depression, and 3,365 (58.54%) met criteria for non-melancholic depression. Whereas 817 (14.21%) met criteria for atypical depression, 2,548 (44.32%) met criteria for non-atypical depression. The proportion of participants in the melancholic and atypical NESARC specifier subgroups are similar to the specifier frequencies in the STAR*D dataset: melancholic (42.23%), non-melancholic (57.77%), atypical (10.81%), and non-atypical (46.96%).

NESARC

In NESARC, the melancholic specifier subgroup (n=2,384) reported a total of 834 unique diagnostic combinations of melancholic depression, leading to an average ratio of 2.86 patients for each melancholic profile. However, most of the melancholic (89.13%) and non-melancholic (96.71%) diagnostic combinations were endorsed by five or fewer patients. The Hamming distances can be found in Figure 3. The results of our analyses in multivariate space within the melancholic and non-melancholic subgroups, as well as between melancholic and non-melancholic subgroups, suggest

Table 1: Endorsement of specific symptoms of DSM criteria for major depression, melancholia, and atypical specifiers in patients with MDD, MDD with melancholia features, and MDD with atypical features, as determined by the AUDADIS-IV

	AUDADIS-MDE		AUDADIS-Mel		AUDADIS-Aty	
Symptom	**%**	**(n)**	**%**	**(n)**	**%**	**(n)**
Sad mood	94.97	5460	95.51	2277	94.13	769
Anhedonia	87.60	5036	100	2384	85.68	700
Appetite/weight decrease[a]	60.08	3454	86.45	2061	17.87	146
Appetite/weight increase[b]	36.42	2094	29.07	693	100	817
Insomnia Sleep Onset	69.42	3991	84.02	2003	46.88	383
Early morning awakening	54.90	3156	82.38	1964	21.67	383
Hypersomnia[b]	46.83	2692	40.86	974	100	817
Psychomotor retardation[a]	40.76	2343	61.37	1463	30.23	247
Psychomotor agitation[a]	37.50	2156	60.74	1448	20.20	165
Fatigue	84.71	4870	84.94	2025	92.66	757
Worthlessness	62.55	3596	74.20	1769	59.73	488
Guilt[a]	58.13	3342	81.92	1953	47.49	388
Diminished concentration	84.71	4870	91.99	2193	80.78	660
Indecisiveness	75.77	4356	85.19	2031	71.60	585
Suicidal Ideation/Thoughts of Dying	59.51	3421	65.18	1554	56.55	462
Suicide Attempt	11.05	635	14.60	348	8.94	73
Desires/Thoughts of Dying	58.00	3032	64.19	1384	41.71	166

[a] also a symptom of the 'melancholic features' specifier,
[b] also a symptom of the 'atypical features' specifier

that the specifier subgroups do not increase coherence. The melancholic and non-melancholic subgroups show similar average Hamming distance ratios (M_{mel} = 0.318, SD_{mel} = 0.126 and $M_{non\text{-}mel}$ = 0.388, $SD_{non\text{-}mel}$ = 0.126), indicating few differences of within-group coherence. While the between-groups average Hamming ratio ($M_{btw\text{-}mel}$ = 0.412, $SD_{btw\text{-}mel}$ = 0.122) was also close to the within-group Hamming ratios, indicating low differentiation between the melancholic and non-melancholic subgroups. Figure 4 appears to resemble Figure 2 Panel B, whose data were generated at random. These findings suggest that the subgroups are not meaningfully different when only looking at symptom heterogeneity of diagnostic combinations.

The atypical subgroup (n=817) in NESARC reported a total of 438 unique diagnostic combinations of DSM-MDE symptoms, while the non-atypical subgroup (n=2,548) reported 1572 such combinations. The ratio of patients to combinations was (1.87) for the atypical subgroup, and (1.62) for the non-atypical subgroup, suggesting both subgroups were equally heterogeneous. Most of the atypical (95.66%) and non-atypical (97.01%) diagnostic combinations were endorsed by five or fewer patients. Similar to the melancholic and non-melancholic subgroups, the atypical and non-atypical subgroups show similar average Hamming distance ratios

($M_{aty} = 0.337$, $SD_{aty} = 0.129$ and $M_{non\text{-}aty} = 0.382$, $SD_{non\text{-}aty} = 0.124$), indicating few differences of within-subgroup coherence. While the between subgroups average Hamming ratio ($M_{btw\text{-}aty} = 0.413$, $SD_{btw\text{-}aty} = 0.122$) is close to the within-subgroup Hamming ratios, indicating low differentiation between the atypical and non-atypical subgroups.

STAR*D

In our previous analysis of STAR*D, the melancholic subgroup (n=1,053) reported a total of 646 unique profiles of depression plus the melancholic specifier symptoms, whereas the non-melancholic subgroup (n=1,443) reported 891 such profiles. The ratio of unique combinations to patients was close to equivalence, (0.61) in the melancholic subgroup and (0.62) in the non-melancholic subgroup. The atypical subgroup (n=270) reported a total of 198 unique profiles of DSM-MDE symptoms, while the non-atypical subgroup (n=1,173) reported 682 such profiles. Thus, the ratio of profiles to patients was somewhat *higher* in the atypical subgroup (0.73) than in the non-atypical subgroup (0.58, i.e., the non-atypical subgroup appeared more homogeneous).

We created boxplots to represent the within- and between-subgroup distances in multivariate space using the STAR*D dataset. The STAR*D melancholic and non-melancholic subgroups displayed similar levels of distance in multivariate space within-subgroup and when comparing between-subgroup ($M_{mel} = 0.316$, $SD_{mel} = 0.123$, $M_{non\text{-}mel} = 0.329$, $SD_{non\text{-}mel} = 0.123$, and $M_{mel\text{-}btw} = 0.371$, $SD_{mel\text{-}btw} = 0.120$), suggesting the melancholic specifier designation does not increase subgroup coherence. Similarly, the within and between-subgroup comparisons in atypical vs. non-atypical depression suggested the atypical specifier does not lead to greater coherence ($M_{aty} = 0.345$, $SD_{aty} = 0.137$, $M_{non\text{-}aty} = 0.325$, $SD_{non\text{-}aty} = 0.124$, and $M_{aty\text{-}btw} = 0.305$, $SD_{aty\text{-}btw} = 0.098$). When only focusing on the core DSM-5 depressive symptoms (i.e., ignoring the specifier symptoms), neither the melancholic

($M_{mel} = 0.282$, $SD_{mel} = 0.128$, $M_{non\text{-}mel} = 0.305$, $SD_{non\text{-}mel} = 0.128$, and $M_{mel\text{-}btw} = 0.323$, $SD_{mel\text{-}btw} = 0.120$) nor the atypical specifier appeared to increase coherence ($M_{aty} = 0.317$, $SD_{aty} = 0.130$, $M_{non\text{-}aty} = 0.297$, $SD_{non\text{-}aty} = 0.128$, and $M_{aty\text{-}btw} = 0.320$, $SD_{aty\text{-}btw} = 0.129$; see Figure 3).

The corresponding figures focusing on Manhattan distances can be found in the Appendix. The trends in the STAR*D Manhattan results mirrored the Hamming distance calculations: there was no evidence that the specifier symptoms reduced heterogeneity, even when we operationalized severity along a continuum.

Discussion

We examined whether the melancholic and atypical specifiers for MDD quantifiably reduced the heterogeneity in the relevant specifiers' subgroups. To achieve this, we derived symptom combinations for each participant and computed distances (i.e., Manhattan's and Hamming's) as measures of heterogeneity, defined here by coherence, in STAR*D and NESARC wave 1. Using the DSM-5's specifier criteria, our analyses did not create more coherent subgroups in either symptom severity or symptom endorsement.

Strengths and Limitations

Several limitations of the current analysis are worth considering. First, patients were excluded from the STAR*D dataset if they reported psychosis, met criteria for anorexia, bulimia, substance dependence, primary OCD, or had prior non-response to citalopram. The only exclusion criteria applied to the NESARC dataset were a lifetime history of mania and hypomania and an illness or a substance-induced MDE, in contrast to the STAR*D's extensive exclusion criteria. Thus the current results may not generalize to patients with bipolar depression, medication-induced depression, and depression due to a general medical condition. Secondly, NESARC did not query for all additional specifier symptoms required for the atypical and melancholic criteria, which forced us to use proxy definitions. Although prior

studies have used these proxy definitions and found that they may be valid for the melancholic and atypical specifiers, there may have been misclassifications relative to relying exclusively on the DSM. Third, our results do not indicate whether melancholic and atypical subgroups are valid clinical constructs that "carve nature at its joints", nor do our results inform whether they are useful in terms of predicting metrics of interest (e.g., treatment outcomes). Finally, we did not examine whether the specifier subgroups are biologically-homogeneous constructs (e.g., as indexed by predictive biomarkers).

Despite these limitations, our study has notable strengths. First, we tested a long-standing assumption of the DSM: that specifier subgroups reduce heterogeneity. Second, we used two large and well-characterized samples that complemented each other in terms of weaknesses (see our earlier discussion). Finally, we moved beyond prior work that has relied on counting symptom diagnostic combinations without quantifying heterogeneity between individuals with continuous metrics. Prior analyses have used metrics requiring 100% agreement in all symptoms to count individuals as being homogeneous. As depression and other forms of psychopathology appear to be better characterized by a continuum of severity, at least between individuals, the similarity is better represented on a symptom continuum rather than categorically (i.e., same profile vs. not the same profile).

Implications

Developing valid subtypes for specific psychopathology may have many benefits, including 1) elucidating specific etiologic mechanisms, 2) creating prescriptive categories that may be used by treatment-matching algorithms, 3) identifying different clinical phenomena (e.g., risk factors, prognosis), 4) and creating more coherent subgroups of patients. However, our results do not indicate that DSM-5 atypical and melancholic specifiers create more coherent subgroups of patients. Although the melancholic and atypical subtypes have been long-rooted in historical

contexts and preserved through the editions of the DSM, the evidence supporting their construct validity is weak, and there is inconsistent evidence of biological correlates of melancholic and atypical specifiers [15–17]. Additionally, the predictive validity of the melancholic and atypical subtypes is likewise inconclusive, at least in matching to cognitive-behavioral therapy vs. SSRIs. [15, 18, 19].

Moreover, the current DSM's definitions of the atypical and melancholic features may not accurately capture their putative hallmark features. In the case of melancholia, a significant divergence in defining the construct between researchers and the DSM-5 is apparent. Many proponents claim psychomotor retardation, and mood non-reactivity as the main components of melancholia [20] whereas others argue that an endogenous onset of depression is melancholia's hallmark feature [21, 22]. Indeed, though melancholic depression has been historically conceptualized as an endogenous presentation of the disorder, the DSM-5 does not query patients for endogeneity. One avenue for future work may be to propose theoretical accounts of melancholic or atypical depression [23], specifying whether they are better understood as networks of reinforcing symptoms, interactions of latent vulnerabilities (e.g., thought disorder X psychomotor disturbances X detachment), or clusters of symptoms that are differentially aggregated across people. Alternatively, specific symptoms themselves may indicate more homogeneous subgroups. For example, both positive affectivity and sleep disturbances appear significant in predicting symptom change during treatment and may be suitable candidate endophenotypes to pursue [24–27]. Indeed, taking a symptoms-based approach may help disassemble the potentially relevant biomarkers of the melancholic and atypical subtypes. For example, whether elevated cortisol levels in the morning indicate the presence of melancholia or the elevated cortisol levels indicate the melancholic *symptom* 'early morning awakening' has not been explored.

Nevertheless, research on potential depressive subtypes appears to assume a latent variable model rather than a symptom or network-focused approach [28–30]. Researchers have shown that individual depressive symptoms have differing heritability [31] and correlate differentially with clinical validators (e.g., prognosis, comorbidities; [32]). Further, there is a burgeoning discussion surrounding the etiologies and biological mechanisms associated with specific symptoms. Researchers have proposed that neurovegetative depressive symptoms (e.g., sleep disturbances, psychomotor changes) but not cognitive symptoms (e.g., impaired concentration) have strong associations with inflammation biomarkers, thus alluding to some HPA axis dysfunction [33]. Diagnostic heterogeneity, thus, may correspond to etiological heterogeneity. Whether depressive symptoms should be given equal weight in a diagnosis of depression has not been explored and is worth further scrutiny.

Beyond the construct validity of the atypical and melancholic subtypes, the use of the Hamming and Manhattan distance ratios to calculate coherence is quite similar to the Jaccard index used in prior studies quantifying heterogeneity [34]. However, unlike the Jaccard index, using Manhattan distances allows for coherence to be evaluated along a continuum adding an axis of symptom severity. Regardless, given the push to identify new depressive endophenotypes, the use of a diagnostic heterogeneity measurement should be considered when making comparisons between diagnostic categories and systems in the future. Furthermore, even if identifiable endophenotypes create subgroups low in coherence between diagnostic combinations, measures of heterogeneity could be adapted to the presence of biological correlates (e.g., determining elevated cortisol levels).

Conclusion

The current study does not support the claim that melancholic and atypical depressive specifiers reduce diagnostic heterogeneity, as operationalized by Euclidean distance metrics. The use of distance functions in Euclidean space is a valuable

metric when assessing the utility of psychopathology's current and future diagnostic systems. Future research should further assess the utility of heterogeneity metrics and other potential measures for quantifying symptom heterogeneity, severity, and depressive symptom time course.

Declarations

Ethics Approval and Consent

Not Applicable

Consent for Publication

Not Applicable

Data Availability Statement

The data that support the findings of this study are openly available in NIMH data archive, https://nda.nih.gov/. The code for the analyses can be found here: https://osf.io/v8sbe/.

Conflict of interest

We report no conflict of interest relevant to the current publication.

Funding

Not Applicable

Author's contributions

First Author: JB is credited as the first author of this study with the creation of the methods, running of analyses, and drafting of the manuscript. LL is credited with the conception of the study, providing revisions to the scientific content, and providing stylistic/grammatical revisions to the manuscript. AW is credited with aiding in providing content revisions to the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank the investigators of the STAR*D study for collecting and making available this data and the participants for giving their time towards the advancement of clinical science.

Author details

[1] Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405, USA. [2] Department of Psychological Sciences, University of Missouri, Columbia, MO 65211, USA.

References

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-5. 5th ed. Washington, DC: Autor; 2013.
2. Fried EI, Nesse RM. Depression is not a consistent syndrome: an investigation of unique symptom patterns in the STAR* D study. Journal of Affective Disorders. 2015;172:96–102.
3. Zimmerman M, Ellison W, Young D, Chelminski I, Dalrymple K. How many different ways do patients meet the diagnostic criteria for major depressive disorder? Comprehensive Psychiatry. 2015;56:29–34.
4. Rush AJ, Fava M, Wisniewski SR, Lavori PW, Trivedi MH, Sackeim HA, et al. Sequenced treatment alternatives to relieve depression STAR* D: rationale and design. Controlled Clinical Trials. 2004;25(1):119–142.
5. Lorenzo-Luaces L, Buss JF, Fried EI. Heterogeneity in major depression and its melancholic and atypical specifiers: a secondary analysis of STAR* D. BMC psychiatry. 2021;21(1):1–11.
6. Fried EI, Coomans F, Lorenzo-Luaces L. The 341 737 ways of qualifying for the melancholic specifier. The Lancet Psychiatry. 2020;7(6):479–480.
7. Fried EI, Nesse RM. Depression sum-scores don't add up: why analyzing specific depression symptoms is essential. BMC medicine. 2015;13(1):1–11.
8. Chen CM, Dufour MC, Yi Hy. Alcohol consumption among young adults ages 18–24 in the United States: Results from the 2001–2002 NESARC survey. Alcohol Research & Health. 2004;28(4):269.
9. Fava M, Rush AJ, Trivedi MH, Nierenberg AA, Thase ME, Sackeim HA, et al. Background and rationale for the sequenced treatment alternatives to relieve depression (STAR* D) study. Psychiatric Clinics of North America. 2003;.
10. Zimmerman M, Mattia JI. A self-report scale to help make psychiatric diagnoses: the Psychiatric Diagnostic Screening Questionnaire. Archives of General Psychiatry. 2001;58(8):787–794.
11. Grant BF, Dawson DA, Stinson FS, Chou PS, Kay W, Pickering R. The Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (AUDADIS-IV): reliability of alcohol consumption, tobacco use, family history of depression and psychiatric diagnostic modules in a general population sample. Drug and alcohol dependence. 2003;71(1):7–16.
12. Case SM, Sawhney M, Stewart JC. Atypical depression and double depression predict new-onset cardiovascular disease in US adults. Depression and anxiety. 2018;35(1):10–17.
13. Bookstein A, Kulyukin VA, Raita T. Generalized hamming distance. Information Retrieval. 2002;5(4):353–375.
14. Reynolds BE. Taxicab geometry. Pi Mu Epsilon Journal. 1980;7(2):77–88.
15. Sani G, Tondo L, Undurraga J, Vázquez GH, Salvatore P, Baldessarini RJ. Melancholia: does this ancient concept have contemporary utility? International Review of Psychiatry. 2020;32(5-6):466–470.
16. Łojko D, Rybakowski JK. Atypical depression: current perspectives. Neuropsychiatric Disease and Treatment. 2017;13:2447.
17. Day CV, Williams LM. Finding a biosignature for melancholic depression. Expert review of neurotherapeutics. 2012;12(7):835–847.
18. Uher R, Dernovsek MZ, Mors O, Hauser J, Souery D, Zobel A, et al. Melancholic, atypical and anxious depression subtypes and outcome of treatment with escitalopram and nortriptyline. Journal of Affective Disorders. 2011;132(1-2):112–120.
19. Imai H, Noma H, Furukawa TA. Melancholic features (DSM-IV) predict but do not moderate response to

antidepressants in major depression: an individual participant data meta-analysis of 1219 patients. European Archives of Psychiatry and Clinical Neuroscience. 2021;271(3):521–526.

20. Martino DJ, Szmulewicz AG, Valerio MP, Parker G. Melancholia: an attempt at definition based on a review of empirical data. The Journal of Nervous and Mental Disease. 2019;207(9):792–798.
21. Parker G. Melancholia. American Journal of Psychiatry. 2005;162(6):1066–1066.
22. Parker G. A case for reprising and redefining melancholia. The Canadian Journal of Psychiatry. 2013;58(4):183–189.
23. Robinaugh D, Haslbeck J, Waldorp L, Kossakowski J, Fried EI, Millner A, et al. Advancing the network theory of mental disorders: A computational model of panic disorder. PsyArXiv. 2019;.
24. Modell S, Lauer CJ. Rapid eye movement (REM) sleep: an endophenotype for depression. Current psychiatry reports. 2007;9(6):480–485.
25. Gottesmann C, Gottesman I. The neurobiological characteristics of rapid eye movement (REM) sleep are candidate endophenotypes of depression, schizophrenia, mental retardation and dementia. Progress in neurobiology. 2007;81(4):237–250.
26. Hasler G, Drevets WC, Manji HK, Charney DS. Discovering endophenotypes for major depression. Neuropsychopharmacology. 2004;29(10):1765–1781.
27. Fisher AJ, Reeves JW, Lawyer G, Medaglia JD, Rubel JA. Exploring the idiographic dynamics of mood and anxiety via network analysis. Journal of abnormal psychology. 2017;126(8):1044.
28. Herane-Vives A, Young AH, Wise T, Aguirre J, de Angel V, Arnone D, et al. Comparison of short-term (saliva) and long-term (hair) cortisol levels in out-patients with melancholic and non-melancholic major depression. BJPsych open. 2020;6(3).
29. Dougherty LR, Klein DN, Olino TM, Dyson M, Rose S. Increased waking salivary cortisol and depression risk in preschoolers: the role of maternal history of melancholic depression and early child temperament. Journal of Child Psychology and Psychiatry. 2009;50(12):1495–1503.
30. Michopoulos I, Zervas IM, Pantelis C, Tsaltas E, Papakosta VM, Boufidou F, et al. Neuropsychological and hypothalamic–pituitary-axis function in female patients with melancholic and non-melancholic depression. European Archives of Psychiatry and Clinical Neuroscience. 2008;258(4):217–225.
31. Jang KL, Livesley WJ, Taylor S, Stein MB, Moon EC. Heritability of individual depressive symptoms. Journal of affective disorders. 2004;80(2-3):125–133.
32. Lux V, Kendler K. Deconstructing major depression: a validation study of the DSM-IV symptomatic criteria. Psychological medicine. 2010;40(10):1679–1690.
33. Majd M, Saunders EF, Engeland CG. Inflammation and the dimensions of depression: a review. Frontiers in Neuroendocrinology. 2020;56:100800.
34. Olbert CM, Gala GJ, Tupler LA. Quantifying heterogeneity attributable to polythetic diagnostic criteria: theoretical framework and empirical application. Journal of Abnormal Psychology. 2014;123(2):452.

Figure 1: Melancholic and Atypical subgroups of patients derived from the IDS on the STAR*D (Panel A) and AUDADIS-IV on the NESARC (Panel B) data-sets

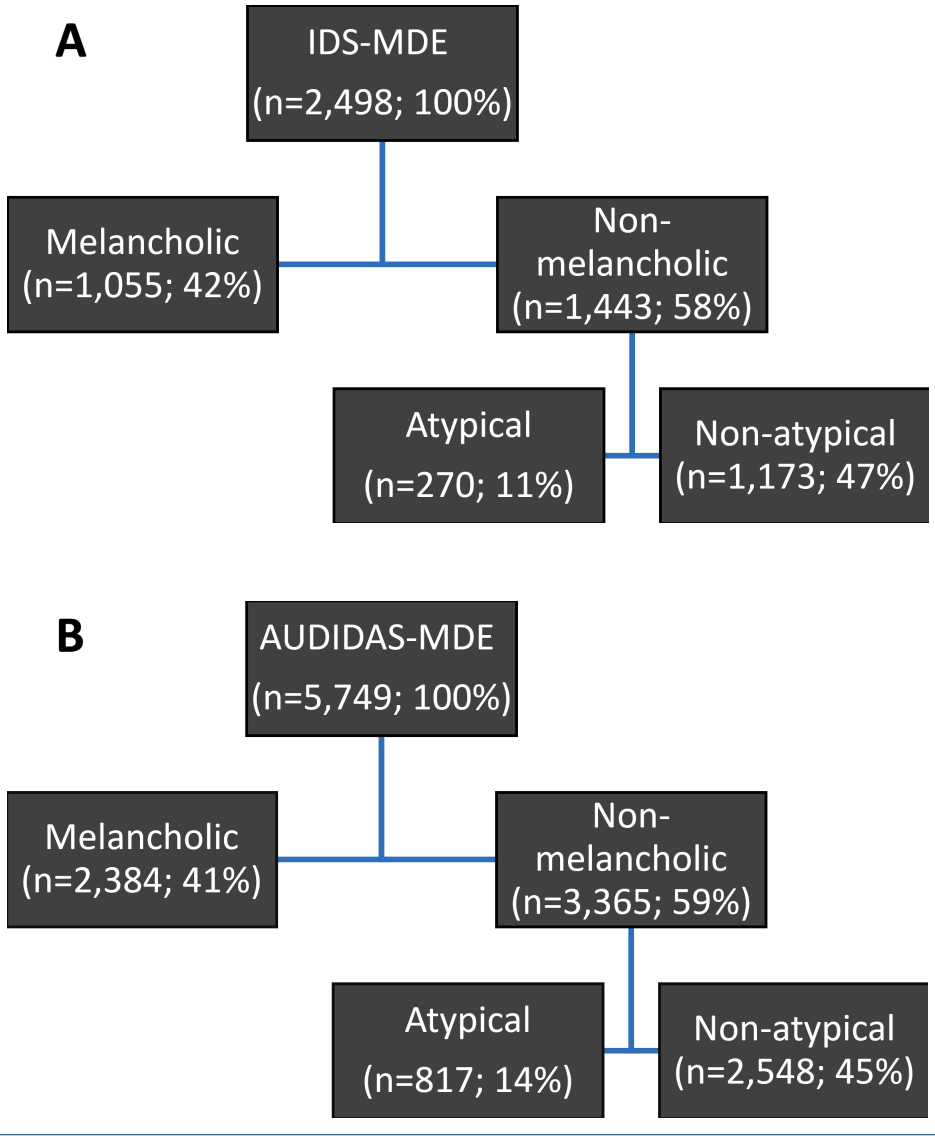


Figure 2: Illustration of distance ratios indicating ideal inner-group coherence and between-group differentiation between subgroup profiles (Panel A) and heterogenous subgroup profiles generated using random data (Panel B)

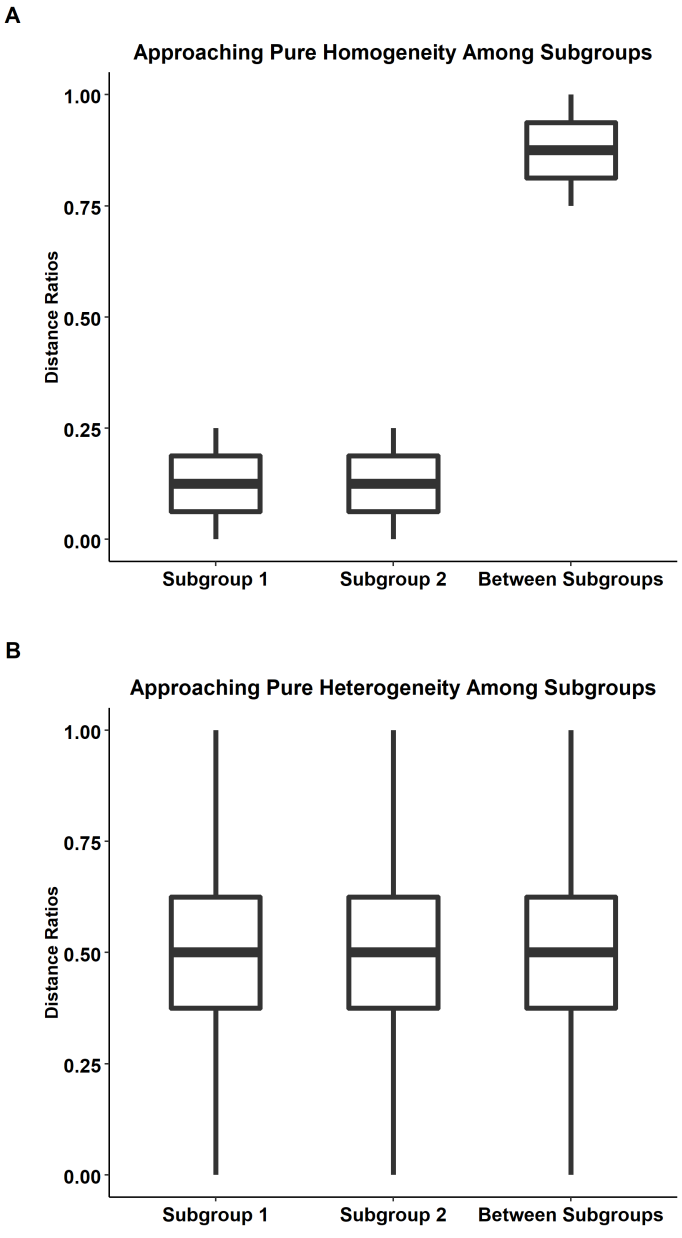


Figure 3: Hamming distance ratios of diagnostic combinations for patients meeting MDD criteria queried by the IDS-SR within the STAR*D data-set, with specifier symptoms, and without specifier symptoms

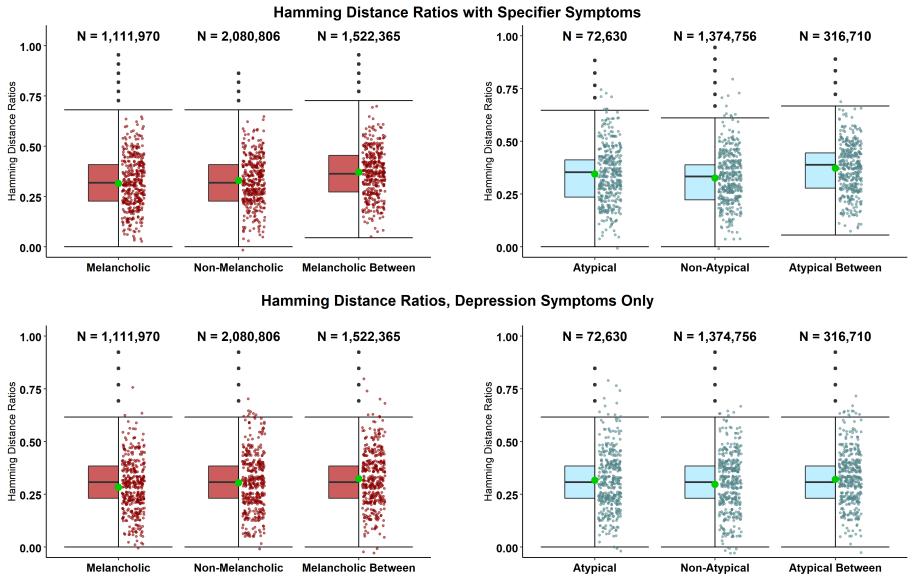
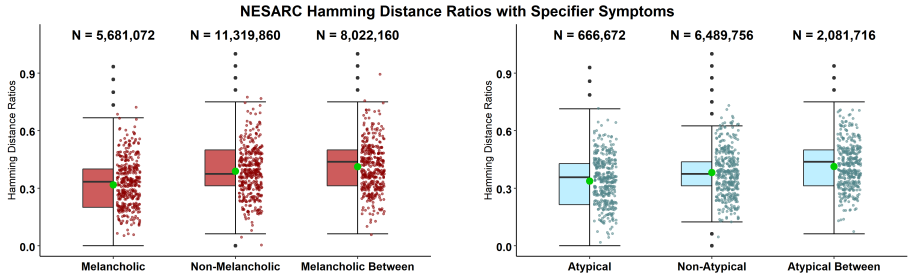


Figure 4: Hamming distance ratios of diagnostic combinations for patients meeting MDD criteria queried by the AUDADIS-IV within the NESARC data-set



Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AppendixNESARC.docx