HHS Public Access

Author manuscript

Psychiatry Res. Author manuscript; available in PMC 2019 August 01.

Published in final edited form as:

Psychiatry Res. 2018 August ; 266: 228–246. doi:10.1016/j.psychres.2018.03.003.

The Use of Latent Class Analysis for Identifying Subtypes of Depression: A Systematic Review

Christine M. Ulbricht[a], **Stavroula A. Chrysanthopoulou**[a], **Len Levin**[b], and **Kate L. Lapane**[a]

[a]Department of Quantitative Health Sciences, University of Massachusetts Medical School, 368 Plantation St, Worcester, MA 01605, U.S.A.

[b]Lamar Soutter Library, University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, MA 01655, U.S.A.

Abstract

Depression is a significant public health problem but symptom remission is difficult to predict. This may be due to substantial heterogeneity underlying the disorder. Latent class analysis (LCA) is often used to elucidate clinically relevant depression subtypes but whether or not consistent subtypes emerge is unclear. We sought to critically examine the implementation and reporting of LCA in this context by performing a systematic review to identify articles detailing the use of LCA to explore subtypes of depression among samples of adults endorsing depression symptoms. PubMed, PsycINFO, CINAHL, Scopus, and Google Scholar were searched to identify eligible articles indexed prior to January 2016. Twenty-four articles reporting 28 LCA models were eligible for inclusion. Sample characteristics varied widely. The majority of articles used depression symptoms as the observed indicators of the latent depression subtypes. Details regarding model fit and selection were often lacking. No consistent set of depression subtypes was identified across studies. Differences in how models were constructed might partially explain the conflicting results. Standards for using, interpreting, and reporting LCA models could improve our understanding of the LCA results. Incorporating dimensions of depression other than symptoms, such as functioning, may be helpful in determining depression subtypes.

Keywords

depression; depression subtypes; finite mixture model; latent class analysis

1. Introduction

Successful diagnosis and treatment of depression has long thought to be impeded by uncertainty regarding the nosology of depression. With more than 1,400 possible

Corresponding author Christine M. Ulbricht, PhD, Department of Quantitative Health Sciences, University of Massachusetts Medical School, 368 Plantation Street, AS7.1078, Worcester, MA 01605 Christine.ulbricht@umassmed.edu.

Disclosures: None

combinations of diagnostic criteria for major depressive disorder alone (Ostergaard et al., 2011), symptomatology is non-specific and there is substantial variability in risk factors, severity, and illness course (Rush, 2007). This heterogeneity appears to influence treatment response, which is suboptimal despite decades of research and increasing rates of antidepressant use (Insel and Wang, 2009; National Center for Health Statistics, 2011; Wang and Insel, 2010).

While the existence of such heterogeneity is well-known, the best ways to delineate subtypes of depression, or subgroups of people who share the same features of the illness, remains to be determined (Baumeister and Parker, 2012). Numerous, not mutually exclusive, and sometimes conflicting depression subtypes have been proposed but the clinical utility of these categorizations is unclear. These subtypes have traditionally been based on differences in etiology, symptoms, time of onset, gender, and treatment response (Baumeister and Parker, 2012) and have been elucidated using a variety of variable-centered and person-centered method analytic methods. Latent variable methods, particularly latent class analysis (LCA), have been increasingly used for such purposes.

The person-centered LCA approach assumes the existence of mutually exclusive and exhaustive groups of individuals that can be differentiated by values of an unobserved latent variable. The latent variable and the resulting subgroups are based on observed indicator variables such as depression symptoms. This method is appealing because, as opposed to more traditional variable-centered methods of subgroup analysis, LCA can theoretically be used to create subgroups from a large amount of information such as the multiple facets of depression (Lanza and Rhoades, 2011). Whereas variable-centered analytic methods examine relationships among variables under the assumption that the relationships between these variables are homogenous in a population, person-centered methods examine the relationships among individuals under the assumption that the relationships between these attributes are heterogeneous in a population (Masyn, 2013). Variable-centered methods such as regression modeling are useful for describing relationship between independent and dependent variables and for predicting outcomes while person-centered approaches are useful for elucidating groups of people in order to understand the differences and similarities between people (Muthen and Muthen, 2000). While prior work has examined data-driven methods for examining symptomatic subtypes of depression (Ten Have et al., 2016; van Loo et al., 2012), to our knowledge, no published systematic review has focused exclusively on how LCA is applied to distinguish subtypes of depression. One previous systematic review examined latent variable analyses of depressive symptoms in patients with major depressive disorder but included latent class analyses, latent factor analyses, confirmatory factor analyses, and exploratory and principal component analyses (van Loo et al., 2012). The review by van Loo et al. mainly focused on presenting the results of these different analytic methods and did not examine many of the technical aspects of building the included models. The purpose of our review was to examine how LCA has been used for deriving subtypes of depression, with the objectives of: 1) describing differences in how these methods have been applied in samples of adults who had screened positive for probable depression or who had a primary diagnosis of major depressive disorder; 2) exploring similarities and differences between resulting subtypes; and 3) identifying possible problems and future directions with these methods.

2. Methods

This review was conducted according to the principles of the PRISMA Statement (Moher et al., 2009).

2.1. Search strategy

Our search strategy was developed in consultation with a research librarian (L.L.). A search for all literature reporting the use of LCA to identify subtypes of psychiatric disorders was conducted. PubMed, PsycINFO, CINAHL, Scopus (including EMBASE), and Google Scholar databases were searched to identify relevant articles indexed through January 2016. The following keywords and MeSH (Medical Subject Headings) terms were used for this search within PubMed: “Latent Class Analysis,” “LCA,” “Latent Variable,” “Depression,” and “Depressive Disorder.” The search was adapted for the other databases based upon their specific subject hierarchy or lack thereof. References cited in articles eligible for inclusion articles were also hand searched and assessed for eligibility.

2.2. Inclusion and exclusion criteria

Eligible articles: 1) were published in English; 2) developed LCA models in which the unobserved latent variable and observed indicator variables were categorical (Collins and Lanza, 2010; Masyn, 2013); 3) applied LCA to elucidate depression subtypes in people with a primary diagnosis of major depressive disorder or who screened positive as likely having depression; 4) provided details about building the latent model; 4) included adults who were 18 years or older; and 5) were not published as a review article, opinion piece, non-research letter, or commentary. Any article that did not meet all of these criteria was excluded. We defined LCA as those latent class models in which the latent and indicator variables were categorical in order to avoid both confusion over assumptions associated with various related finite mixture models, e.g., latent profile analysis, and in interpreting results from heterogeneous modeling strategies. We considered the categorical variables to be those describing qualitative differences in types of depression rather than continuous, or quantitative, differences in subtypes (Rabe-Hesketh and Skrondal, 2008; Ruscio and Ruscio, 2008). Studies had to include people who had experienced at least some depressive symptoms because we were interested in subtypes of depression and did not consider “no depression” to be an informative subtype. Furthermore, a formal diagnosis of major depressive disorder was not required due to concerns about undetected depression and diagnostic accuracy (Mitchell et al., 2009).

One author (C.M.U.) independently reviewed all titles, abstracts, and full-text article for eligibility. Any differences related to study eligibility were resolved by a second author (K.L.L.).

2.3 Data abstraction

For each included article, one reviewer (C.M.U.) used a standardized form to abstract information about each included study. Another author (K.L.L.) verified the abstracted data. Abstracted details include study design (e.g., cross-sectional, longitudinal), setting, sample characteristics (e.g., size, criteria for depression determination), means of depression assessment (instrument, time period), and resulting subtypes (e.g., number, description, prevalence). Because we were interested in understanding how resulting subtypes might differ by latent variable model, we also abstracted the following information about model building, model selection, and model interpretation.

- *Latent class indicators*: The relationship between each latent class and the observed indicator variables is the foundation for understanding the nature of the latent construct. For an indicator to be of high quality, it needs to measure the latent variable reliably (Masyn, 2013). Simulation studies suggest that the use of more indicators may reduce the number of models with poor class assignment accuracy and improve model convergence rates) (Wurpts and Geiser, 2014). We extracted the number of indicators included in the LCA model, each indicator, and the categorization of the indicators (e.g., binary, etc.).

- *Latent class membership prevalences* (γ): The latent class membership prevalences, estimated by an LCA model, represent the probability of membership in each class. The prevalence of each latent class provides information about the quality of the LCA model since the presence of a small class can indicate problems with model identification (Masyn, 2013).

- *Item-response probabilities* (ρ): The item-response probabilities estimated from an LCA model constitute the response patterns of the observed indicator items and the latent variable for each latent class (Collins and Lanza, 2010). High homogeneity, or whether or not item-response probabilities are close to 0 or 1, is a desirable feature of an LCA model (Collins and Lanza, 2010). An item-response probability approaching 0 or 1 indicates that there is a strong relationship between the indicator and the latent variable, meaning that the particular response could be determined with a high level of certainty given latent class membership. In a model with high homogeneity, the likelihood of observing a response pattern that is characteristic of a particular latent class is higher than in a model with low homogeneity. Because universal cutoff points for establishing homogeneity have not been established, we did not evaluate the quality of homogeneity in each study. Instead, we noted whether the item-response probabilities were provided and if they were presented as a table or probability profile plot.

- *Latent class separation*: Latent class separation refers to the ability to distinguish item-response probability patterns between the different classes (Masyn, 2013). In a model with a high degree of separation, a response pattern that describes one class describes only that class (Collins and Lanza, 2010). Formal criteria of what constitutes an acceptably high degree of separation are lacking and thus we stated if the item-response probabilities were provided and if so, whether they were displayed as a table or plot.

- *Software*: A number of free and commercial software packages exist for conducting latent class analysis. The packages vary in terms of options available for estimation algorithms procedures and other parameters for specifying LCA models. Because modeling options differ among the programs, we hypothesized that LCA models would differ by software and noted the software used for analyses and if any specific modeling assumptions or software options were noted.

- *Number of subtypes explored*: Building an LCA model is an iterative process which should entail an initial stage of fitting models with increasing numbers of classes to establish the appropriate number of classes (Masyn, 2013). We collected the range of classes that were explored by each study.
- *Model estimation*: LCA parameters cannot be estimated through closed-form solutions and thus are usually estimated using iterative procedures such as the expectation-maximization (EM) algorithm and the Newton-Raphson algorithm (Collins and Lanza, 2010). These procedures employ maximum likelihood estimation to find the values that are most likely observed in the indicator data. Multiple sets of random starting values should be employed for purposes of obtaining model convergence and replicating the best maximum log likelihood value (Masyn, 2013).
- *Model selection*: A correctly specified LCA model should be chosen based on a number of considerations, including fit statistics and interpretability. The relative and absolute fit of models can be evaluated using measures such as the likelihood-ratio difference test, Akaike information criterion (AIC), and Bayesian information criterion (BIC). Classification diagnostics such as entropy can also be employed. Entropy is a measure of correctly classifying individuals into latent classes. Higher values represent better latent class separation (Collins and Lanza, 2010). We extracted which measures were used to guide model selection.
- *Measurement invariance*: Evaluating the presence of measurement invariance allows the researcher to understand whether or not the latent variable constructs are the same across levels of a third variable. We extracted whether each study examined measurement invariance. If the study examined measurement invariance, we extracted what variables were evaluated, the rationale for selecting these variables, and what methods were employed to determine if measurement invariance existed.
- *Covariates/correlates/grouping variable*: The inclusion of one or more observed variables to be used as covariates or grouping variables can help describe additional features of latent classes. Two methods proposed are: classify-analyze and model-based approach. The classify-analyze approach assigns individuals to their most likely class based on their greatest posterior probability of membership and then models the association between the assigned class and covariate of interest. In the model-based approach, the actual probability of a covariate conditional on latent class membership is modeled. We extracted details pertaining to the variables explored and the modeling method employed.

3. Results

3.1. Article selection process

Details about the article selection process are provided in Figure 1. We identified 1,465 publications in the initial searches. After removing 622 duplicates, the titles of 843 articles were reviewed. Seven hundred twenty-two studies were excluded based on this review. An additional 62 studies were excluded after reviewing 121 abstracts. The majority of articles were excluded at these stages because they did not focus on subtypes of depression. Fifty-nine full-text articles from this search and two articles identified from other sources were examined for inclusion. Twenty-four articles were ultimately eligible for inclusion. Of the 37 articles that were excluded after full-text review, 14 articles were ineligible because they included participants without depression symptoms. Seven studies used methods other than LCA with categorical indicators. Six studies were not included because they included continuous, instead of categorical, indicators in the LCA models.

3.2 Study characteristics

An overview of the 24 included articles is provided in Table 1. These articles described LCA model building using 26 samples. One article did not include any information about the name, years, location, and setting of the study (Parker et al., 1995). Articles were published from 1990-2016 (electronically published in 2015). The earliest study conducted began in 1928 (Parker and Hadzi-Pavlovic, 1993) while the most recent included participants recruited in 2011 (Sunderland et al., 2013). Two articles analyzed data from the National Comorbidity Survey (NCS) (Prisciandaro and Roberts, 2009; Sullivan et al., 1998) while another two articles used data from the 2001-2002 wave of the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) (Carragher et al., 2009; Lee et al., 2014). For both the NESARC and NCS studies, sample sizes differed within each study and it was not possible to determine the extent to which overlap occurred.

Of those reporting such details, 10 studies used only community-based samples (Alexandrino-Silva et al., 2013; Carragher et al., 2009; Lamers et al., 2012; Lee et al., 2012, 2014; Prisciandaro and Roberts, 2009; Rodgers et al., 2014a, 2014b, Sullivan et al., 2002, 1998). Fourteen studies were conducted with samples recruited solely from specialty or clinical populations (de Vos et al., 2015; Grove et al., 1987; Li et al., 2014; Parker et al., 1999, 1998, 1991, 1990; Parker and Hadzi-Pavlovic, 1993; Sneed et al., 2008; Sullivan et al., 2002; Sunderland et al., 2013; Ulbricht et al., 2015), with two of the samples being comprised of clinical trial participants (Sneed et al., 2008; Ulbricht et al., 2015). One study recruited participants from the general population and primary care and mental health care sites (Lamers et al., 2010). One study did not report how all of the participants were recruited (Parker et al., 1995) but did specify that the participants had been diagnosed with depression.

Sample sizes ranged from 61(Parker and Hadzi-Pavlovic, 1993) to 13,424 participants (Lee et al., 2014). Depression assessments varied but participants of all studies endorsed at least one depression symptom. The CIDI was the most commonly used depression assessment instrument, with eight studies using the scale to screen for depression and its items as indicators in the LCA models (Alexandrino-Silva et al., 2013; Lamers et al., 2012, 2010; Li et al., 2014; Parker et al., 1998; Prisciandaro and Roberts, 2009; Sullivan et al., 1998; Ten Have et al., 2016). The depression assessment timeframe varied from the prior week (Ulbricht et al., 2015) to lifetime, with nine articles evaluating lifetime depression (Alexandrino-Silva et al., 2013; Carragher et al., 2009; Lamers et al., 2010; Lee et al., 2014; Li et al., 2014; Parker et al., 1999; Prisciandaro and Roberts, 2009; Sullivan et al., 1998; Ten Have et al., 2016).

Treatment information was included in 10 articles (de Vos et al., 2015; Lamers et al., 2012; Lee et al., 2012, 2014; Parker et al., 1999; Sneed et al., 2008; Sullivan et al., 2002, 1998; Sunderland et al., 2013; Ulbricht et al., 2015). One of these articles excluded potential participants if they were currently taking antidepressant medication (Sneed et al., 2008). Three articles conducted the LCA on data collected prior to participants receiving treatment (de Vos et al., 2015; Sunderland et al., 2013; Ulbricht et al., 2015). Reporting was generally limited to descriptive details about mental health service (Lamers et al., 2012; Lee et al., 2014; Sullivan et al., 2002, 1998) and/or antidepressant use (Lee et al., 2012, 2014; Parker et al., 1999; Sullivan et al., 1998) although one article did examine response to treatment with citalopram as a distal outcome in the LCA model (Ulbricht et al., 2015).

3.3. Subtypes

The included articles presented 28 unique LCA models of depression subtypes (Table 1). DSM criteria depression symptoms were used as indicators in 14 of the studies (Alexandrino-Silva et al., 2013; Carragher et al., 2009; de Vos et al., 2015; Lamers et al., 2012, 2010; Lee et al., 2014; Li et al., 2014; Prisciandaro and Roberts, 2009; Rodgers et al., 2014a, 2014b, Sullivan et al., 2002, 1998; Sunderland et al., 2013; Ulbricht et al., 2015). One article used physiological measures of depression as indicators (Sneed et al., 2008). The majority of articles detailed only one best-fitting LCA model. Seven articles presented more than one model. Reasons for presenting multiple models included that the classes differed qualitatively by sex (Alexandrino-Silva et al., 2013; Rodgers et al., 2014b), by varied combinations of indicator items (Li et al., 2014), by subsamples of participants (Parker et al., 1991), or the authors not selecting only one best-fitting model to present (de Vos et al., 2015). Additionally, three articles presented models in which the classes were qualitatively the same between different subsamples but varied by prevalence (Parker et al., 1991, 1990; Rodgers et al., 2014a).

The number of classes determined by the LCA models ranged from two to seven. Models with three classes were the most common, accounting for 40% of the included models (Alexandrino-Silva et al., 2013; Lamers et al., 2010; Lee et al., 2012, 2014; Li et al., 2014; Parker et al., 1999, 1995; Prisciandaro and Roberts, 2009; Rodgers et al., 2014a, 2014b; Sunderland et al., 2013). Approximately three-fourths of the models identified at least one class that was labeled according to depression severity (Alexandrino-Silva et al., 2013; Carragher et al., 2009; de Vos et al., 2015; Lamers et al., 2012, 2010, Lee et al., 2012, 2014; Li et al., 2014; Prisciandaro and Roberts, 2009; Rodgers et al., 2014b, 2014a, Sullivan et al., 2002, 1998; Sunderland et al., 2013; Ten Have et al., 2016; Ulbricht et al., 2015). Only one article labeled LCA classes according to the presence of suicidality (Li et al., 2014). Ten models identified a class that was labeled "atypical" (Alexandrino-Silva et al., 2013; Lamers et al., 2012, 2010; Li et al., 2014; Rodgers et al., 2014a, 2014b, Sullivan et al., 2002, 1998) while eleven models had a class labeled "melancholic" (Alexandrino-Silva et al., 2013; Lamers et al., 2010; Parker et al., 1999, 1995, 1991, 1990; Parker and Hadzi-Pavlovic, 1993; Rodgers et al., 2014a). Despite the use of these DSM features when labeling classes, only three analyses included all of the DSM criteria necessary for making such a diagnoses as indicators (Table 2) (Lamers et al., 2010; Parker et al., 1999, 1995).

Various groupings of DSM depression criteria symptoms were used as latent class indicators in the majority of studies (Table 2). The total number of indicators examined ranged from 4-30. Only one study included non-symptom indicators such as executive functioning impairment, deep white matter lesions, and subcortical gray matter lesions (Sneed et al., 2008). One study did not specify which variables were used as indicators (Grove et al., 1987), while another study did not report how the indicators were categorized (Parker et al., 1995). As shown in Figure 2, sleep disturbance was depression symptom most commonly used, having been included in 86.7% of the eligible LCA models. Symptoms of the DSM atypical specifier were used the least as indicators, with 10% of the LCA models overall and 18.2% of models with atypical classes including atypical features.

Of the studies providing details about how many levels the indicators contained, two did not include at least one binary variable, instead using only 3-level (de Vos et al., 2015) or 4-level indicators (Sunderland et al., 2013) (Table 2). Studies differed widely in how many indicators were included and how the indicators were categorized, particularly for weight, appetite, sleep and psychomotor symptoms. The impact of aggregating or disaggregating such variables remains unclear, with only five LCA models elucidating classes defined by these symptoms (Alexandrino-Silva et al., 2013; Lee et al., 2012; Rodgers et al., 2014b; Sullivan et al., 2002; Ulbricht et al., 2015).

The heterogeneity observed across the many different samples also occurred when studies examined the same source population or depression assessment instrument for indicators. The two studies that drew participants from the NESARC and used similar indicators from the AUDADIS-IV both found that four-class LCA models fit the data best but the resulting classes differed slightly in labeling and prevalence (Carragher et al., 2009; Lee et al., 2014). One of the NCS studies resulted in a three-class LCA model (Prisciandaro and Roberts, 2009) while the other proposed a six-class model (Sullivan et al., 1998). These differences may be due to the studies using different eligibility criteria and CIDI items.

3.4 Model building

Approximately 92% of articles reported how many LCA models were explored (Appendix 1) (Alexandrino-Silva et al., 2013; Carragher et al., 2009; de Vos et al., 2015; Grove et al., 1987; Lamers et al., 2012, 2010, Lee et al., 2012, 2014; Li et al., 2014; Parker et al., 1998, 1995, 1991, 1999; Prisciandaro and Roberts, 2009; Rodgers et al., 2014a, 2014b; Sneed et al., 2008; Sullivan et al., 2002, 1998; Sunderland et al., 2013; Ten Have et al., 2016; Ulbricht et al., 2015). Of those that included this information, three evaluated only one LCA model with a fixed number of classes (Grove et al., 1987; Parker et al., 1999; Sneed et al., 2008). The other articles considered multiple LCA models, with the number of classes varying from

1- 15 classes, with a median of 4.5 models of varying classes being explored in each study. As displayed in Figure 3.1, Mplus was the software package used most often to conduct the LCA, with a version being used to conduct almost 60% of the studies (Alexandrino-Silva et al., 2013; Carragher et al., 2009; de Vos et al., 2015; Lamers et al., 2012, 2010, Lee et al., 2012, 2014; Li et al., 2014; Prisciandaro and Roberts, 2009; Rodgers et al., 2014b, 2014a; Sunderland et al., 2013; Ten Have et al., 2016; Ulbricht et al., 2015). The estimation method and starting values used to conduct LCA were not reported more often than not. Eighty-seven percent of articles provided a table and/or graph of item-response probabilities for the indicators in each class (Figure 3.2) (Alexandrino-Silva et al., 2013; Carragher et al., 2009; de Vos et al., 2015; Lamers et al., 2012, 2010, Lee et al., 2012, 2014; Li et al., 2014; Parker et al., 1995, 1991, 1990, 1999; Parker and Hadzi-Pavlovic, 1993; Prisciandaro and Roberts, 2009; Rodgers et al., 2014a, 2014b; Sneed et al., 2008; Sullivan et al., 2002, 1998; Ulbricht et al., 2015). Most of the studies that provided details about model selection criteria reported considering more than one factor when determining model fit. BIC was the fit statistic the most frequently used to guide model selection, with 58% of articles reporting using the traditional BIC (Figure 3.3) (Alexandrino-Silva et al., 2013; Carragher et al., 2009; de Vos et al., 2015; Lamers et al., 2012, 2010, Lee et al., 2012, 2014; Li et al., 2014; Prisciandaro and Roberts, 2009; Rodgers et al., 2014b, 2014a; Sunderland et al., 2013; Ten Have et al., 2016; Ulbricht et al., 2015).

3.5 Measurement invariance

Table 3 summarizes the studies that evaluated measurement invariance (Alexandrino-Silva et al., 2013; Lee et al., 2014; Rodgers et al., 2014b; Ulbricht et al., 2015). Of the included studies, four conducted analyses of measurement invariance. Three of these studies evaluated sex or gender (Alexandrino-Silva et al., 2013; Rodgers et al., 2014b; Ulbricht et al., 2015). The most common approach was to conduct stratified analysis by the variable under study. One study reported formal statistical testing to evaluate the extent to which the classes were qualitatively different (Ulbricht et al., 2015). Two studies reported distinct latent classes that differed by the variable of interest (Alexandrino-Silva et al., 2013; Rodgers et al., 2014b) whereas the other two reported similar classes with different prevalence across the variable under study (Lee et al., 2014; Ulbricht et al., 2015).

3.6 Correlates of latent class membership

Eighteen of the articles examined correlates of belonging to the latent classes (Table 4) (Alexandrino-Silva et al., 2013; Carragher et al., 2009; de Vos et al., 2015; Grove et al., 1987; Lamers et al., 2012, 2010, Lee et al., 2012, 2014; Li et al., 2014; Parker et al., 1999, 1998; Parker and Hadzi-Pavlovic, 1993; Rodgers et al., 2014a, 2014b, Sullivan et al., 2002, 1998; Ten Have et al., 2016; Ulbricht et al., 2015). Results across studies were conflicting, with no clear associations between the characteristics examined and class membership. The mostly commonly examined correlates of class membership were gender and age. The most frequently reported method for doing so was to assign participants to classes based on their individual posterior probabilities of class membership and then use these groups in additional analyses.

4. Discussion

While the heterogeneity of depression is established (Rush, 2007), less is known about how best to elucidate subtypes of depression to personalize treatment. LCA models have been a popular approach to this problem but results of these models are often conflicting. This systematic review identified 24 articles describing 28 LCA models, with substantial differences in these subtypes. Although LCA is frequently used, the clinical utility of the resulting latent subtypes remains unclear. The lack of implementation of these subtypes may be related to how the models are based on probabilities of class membership and there have been difficulties in being able to examine correlates of latent class membership to provide an understanding of the characteristics of subtype members. Additionally, guidance is sparse on how to validate these latent subtypes. Another possible reason for the limited adoption of LCA results is the general limited utility of additional symptom-based disorder-specific subtypes. It may be more useful to examine dimensional constructs of behavior linked to prominent symptoms across psychiatric diagnoses (Kozak and Cuthbert, 2016).

Latent classes were generally distinguished by depression severity, individual symptoms demonstrating uniquely high item-response probabilities, and according to the DSM features of atypical and melancholic depression. Despite these similarities, some inconsistencies were observed in the number of subtypes that were identified, the prevalences of the subtypes, and characteristics of the subtypes. These inconsistencies could be a product of which indicator variables were used. Models were comprised of anywhere from 4-27 indicators. While most of the articles stated that depression criteria symptoms were used as indicators, these indicators may not be truly identical since study participants were recruited over a period of 83 years. Our understanding of depression and the development of diagnostic criteria have changed substantially during this time. The differences in the results of the LCA models may also arise from how indicators were categorized. Weight and appetite changes, sleep disturbances, and psychomotor disturbances were the symptoms categorized and collapsed in the biggest variety of ways. Collapsing variables could lead to loss of detail necessary for latent class separation while including too many variables could lead to difficulties in model convergence. Furthermore, disaggregating indicators that may be correlated with each other, such as increased appetite/weight versus decreased appetite/weight, may lead to violations of the local independence assumption. While it has been suggested that distinguishing between such oppositional indicators is helpful in distinguishing depression subgroups (Ten Have et al., 2016), violating the local independence assumption may lead to issues with model identification. Methods for evaluating and relaxing this assumption, such as adding within-class correlation parameters to the model, have been proposed but the extent to which such violations are addressed when building LCA models remains unclear (Asparouhov and Muthen, 2011; Huang and Bandeen-Roche, 2004; Ten Have et al., 2016).

An additional source of inconsistencies in the LCA models included in this review is the labeling of the latent classes. The utility of latent classes relies in part on the assigned labels of the classes. Labels are subjectively assigned by investigators, a process that is theoretically based on how the item-response probabilities of each indicator variable differ between the classes. More than half of the LCA models contained at least one class labeled by depression severity. These labels were usually based on the absolute number of symptoms likely to be endorsed in one class relative to another rather than the clinical significance of the symptoms. Categorical subtypes of depression have not been consistent for predicting antidepressant treatment response. A dimensional approach may be required (Arnow et al., 2015). This may necessitate the use of methods other than LCA to determine clinically useful depression subtypes.

The labels applied to the subtypes can be misleading, a problem compounded when the item-response probabilities are not presented. Many subtypes were labeled according to depression severity (i.e., mild depression, moderate depression, severe depression) although indicators of symptom severity were not frequently modeled. Rather, severity was often used in labeling subtypes when differing numbers of depression symptoms being present distinguished the subtypes. A number of LCA models labeled subtypes with the DSM features atypical and melancholic despite few studies actually evaluating all of the DSM criteria of these diagnoses. Furthermore, latent classes are based on item-response probabilities, meaning that latent class members only have a likelihood of having each indicator symptom. These probabilities range from 0 to 1 but rarely are such extremes observed in LCA models. It is thus not known with certainty whether people likely to belong to a subtype actually experience the characteristics of that subtype. This issue is also potentially present with the models containing classes that were labeled as non-depressed despite all the articles included in this review analyzing data from participants who had at least one symptom of depression.

Many details about model specification/identification/selection were not explicitly stated. This is potentially important for understanding the subtypes because a number of decisions about the indicator variables and model selection are required when conducting LCA and software packages have different defaults and options (Haughton et al., 2009; Kongsted and Nielsen, 2017). For example, PROC LCA in SAS historically used the EM algorithm for maximum likelihood estimation but Mplus could additionally handle Bayesian estimation. The use of Bayesian estimation has been advocated for avoiding model misspecification due to violations of the local independence assumption (Asparouhov and Muthen, 2011). It is not always apparent, even after reading the technical manuals, which parameterizations and inputs were used or are even available.

Omitting details about model development might contribute to very few LCA models having been successfully replicated in different samples. This may be somewhat expected given that the fundamental principle behind this method is that the subtypes are considered to be latent because true subtype membership cannot be directly observed. Because of this, LCA is often considered to be exploratory (Lanza and Cooper, 2016). Despite their inherently exploratory nature, LCA can be a useful method for detecting common but previously unidentified patterns that can then be confirmed through other methods. Indeed, some researchers have used their initial LCA results in follow-up studies to further develop and refine hypotheses about depression subtypes and biomarkers. The results of such studies suggest that hormones (Milaneschi et al., 2017; Rodgers et al., 2015), inflammatory markers (Lamers et al., 2013), proteomics (Lamers et al., 2016), and genetic markers (Milaneschi et al., 2016, 2014) may help distinguish depression subtypes. Additionally, several recent analytic advances, such model-based methods predicting a distal outcome from subtype membership and the ability to incorporate causal inference techniques, may increase the usefulness of LCA models (Butera et al., 2014; Lanza et al., 2013).

It is important to note that the results of this review should be interpreted with several limitations in mind. The heterogeneity in how LCA was applied and the lack of transparency in how studies are reported hindered our ability to conduct a more in-depth analysis of the relationships between how models were constructed and the resulting latent subtypes. The results of this review only represent people who endorsed at least one depression symptom and thus do not necessarily reflect people currently in remission or those with subsyndromal depression. Furthermore, we limited the scope of this review to subtypes of depression and thus we did not capture studies that examined transdiagnostic subtypes of psychiatric disorder, which may prove to be a useful application of these latent variable models. Additionally, participants in half of the studies were recruited from healthcare sites, usually psychiatric clinics, which may limit the generalizability of the results. This review may also be limited by publication bias and bias in the selection of studies for inclusion. Studies with negative results, such as LCA models that did not converge, might not have been published. Omitting such research could mean that evidence disconfirming the use of LCA to elucidate depression was overlooked by this review.

In summary, this is the first review to comprehensively examine differences in LCA modeling strategies and the resulting latent depression subtypes. Given the high prevalence, morbidity, and mortality of depression, successfully diagnosing and treating the disorder is of high importance. Understanding the heterogeneity of depression to efficiently produce treatment remission has eluded experts for decades (Baumeister and Parker, 2012). Latent variable methods such as LCA hold promise for improving efforts to provide precision medicine in psychiatry but the building, interpretation and reporting of these methods are inconsistent. For transparency needed for reproducibility of scientific research, we advocate for the clear conduct and reporting of the application of LCA. Basic guidelines were published while this systematic review was being conducted, but they have yet to be widely implemented (Lubke and Luningham, 2017; Schreiber, 2016). While both publications highlight the theoretical frameworks of latent variable models and the need to provide more analytic details in manuscripts, the guidelines differ somewhat scope. The guidance by Lubke and Luningham covers latent variable mixture models broadly and emphasizes the value of designing and following an analysis plan, conducting and reporting exploratory analyses, and reporting details of modeling decisions (Lubke and Luningham, 2017). Schreiber provides reporting guidelines for LCA specifically (Schreiber, 2016). These guidelines stress including: traditional descriptive and frequency data, comprehensive evaluative information for all LCA models tested, the rationale behind modeling choices, and detailed information about modeling software (Schreiber, 2016). Whether or not such guidelines will translate into better reporting of LCA models remains to be seen. Examining the contributions of oft-ignored aspects of depression, such as functioning or cognition, could augment efforts to elucidate types of depression and other psychiatric disorders.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: This work was supported by the National Institutes of Health (R56NR015498).

References

Alexandrino-Silva C, Wang Y-P, Carmen Viana M, Bulhões RS, Martins SS, Andrade LH, 2013 Gender differences in symptomatic profiles of depression: results from the São Paulo Megacity Mental Health Survey. J. Affect. Disord. 147, 355–64. doi:10.1016/j.jad.2012.11.041 [PubMed: 23246363]

Arnow BA, Blasey C, Williams LM, Palmer DM, Rekshan W, Schatzberg AF, et al., 2015 Depression subtypes in predicting antidepressant response: A report from the iSPOT-D trial. Am. J. Psychiatry appi.ajp.2015.1. doi:10.1176/appi.ajp.2015.14020181

Asparouhov T, Muthen B, 2011 Using Bayesian priors for more flexible latent class analysis. Proc. 2011 Jt. Stat. Meet. Sect. Gov. Stat 4979–4993.

Baumeister H, Parker G, 2012 Meta-review of depressive subtyping models. J. Affect. Disord. 139, 126–40. doi:10.1016/j.jad.2011.07.015 [PubMed: 21885128]

Butera NM, Lanza ST, Coffman DL, 2014 A framework for estimating causal effects in latent class analysis: is there a causal link between early sex and subsequent profiles of delinquency? Prev. Sci. 15, 397–407. doi:10.1007/s11121-013-0417-3 [PubMed: 23839479]

Carragher N, Adamson G, Bunting B, McCann S, 2009 Subtypes of depression in a nationally representative sample. J. Affect. Disord. 113, 88–99. doi:10.1016/j.jad.2008.05.015 [PubMed: 18644628]

Collins LM, Lanza ST, 2010 Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences. John Wiley & Sons, Hoboken, NJ.

de Vos S, Wardenaar KJ, Bos EH, Wit EC, de Jonge P, 2015 Decomposing the heterogeneity of depression at the person-, symptom-, and time-level: latent variable models versus multimode principal component analysis. BMC Med. Res. Methodol 15, 88. doi:10.1186/s12874-015-0080-4 [PubMed: 26471992]

Grove WM, Andreasen NC, Young M, Endicott J, Keller MB, Hirschfeld RM, et al., 1987 Isolation and characterization of a nuclear depressive syndrome. Psychol. Med. 17, 471–84. [PubMed: 3602239]

Haughton D, Legrand P, Woolford S, 2009 Review of three latent class cluster analysis packages: Latent Gold, poLCA, and MCLUST. Am. Stat. 63, 81–91.

Huang G-H, Bandeen-Roche K, 2004 Building an identifiable latent class model with covariate effects on underyling and measured variables. Psychometrika 69, 5–32.

Insel TR, Wang PS, 2009 The STAR*D trial: revealing the need for better treatments. Psychiatr. Serv. 60, 1466–7. doi:10.1176/appi.ps.60.11.1466 [PubMed: 19880463]

Kongsted A, Nielsen AM, 2017 Latent class analysis in health research. J. Physiother. 63, 55–58. doi:10.1016/j.jphys.2016.05.018 [PubMed: 27914733]

Kozak MJ, Cuthbert BN, 2016 The NIMH Research Domain Criteria initiative: Background, issues, and pragmatics. Psychophysiology 53, 286–97. doi:10.1111/psyp.12518 [PubMed: 26877115]

Lamers F, Bot M, Jansen R, Chan MK, Cooper JD, Bahn S, et al., 2016 Serum proteomic profiles of depressive subtypes. Transl. Psychiatry 6, e851. doi:10.1038/tp.2016.115 [PubMed: 27404283]

Lamers F, Burstein M, He J, Avenevoli S, Angst J, Merikangas KR, 2012 Structure of major depressive disorder in adolescents and adults in the US general population. Br. J. Psychiatry 201, 143–50. doi:10.1192/bjp.bp.111.098079 [PubMed: 22700082]

Lamers F, de Jonge P, Nolen WA, Smit JH, Zitman FG, Beekman ATF, et al., 2010 Identifying depressive subtypes in a large cohort study: results from the Netherlands Study of Depression and

Anxiety (NESDA). J. Clin. Psychiatry 71, 1582–9. doi:10.4088/JCP.09m05398blu [PubMed: 20673552]

Lamers F, Vogelzangs N, Merikangas KR, de Jonge P, Beekman ATF, Penninx BWJH, 2013 Evidence for a differential role of HPA-axis function, inflammation and metabolic syndrome in melancholic versus atypical depression. Mol. Psychiatry 18, 692–9. doi:10.1038/mp.2012.144 [PubMed: 23089630]

Lanza ST, Cooper BR, 2016 Latent class analysis for developmental research. Child Dev. Perspect. 10, 59–64. doi:10.1111/cdep.12163

Lanza ST, Rhoades BL, 2011 latent class analysis: an alternative perspective on subgroup analysis in prevention and treatment. Prev. Sci. 14, 157–68. doi:10.1007/s11121-011-0201-1

Lanza ST, Tan X, Bray BC, 2013 Latent class analysis with distal outcomes: A flexible model-based approach. Struct. Equ. Model. 20, 1–26.

Lee C-T, Leoutsakos J-M, Lyketsos CG, Steffens DC, Breitner JCS, Norton MC, 2012 Latent class-derived subgroups of depressive symptoms in a community sample of older adults: the Cache County Study. Int. J. Geriatr. Psychiatry 27, 1061–9. doi:10.1002/gps.2824 [PubMed: 22135008]

Lee SY, Xue Q, Spira AP, Lee HB, 2014 Racial and ethnic differences in depressive subtypes and access to mental health care in the United States. J. Affect. Disord. 155, 130–7. doi:10.1016/j.jad.2013.10.037 [PubMed: 24269002]

Li Y, Aggen S, Shi S, Gao J, Tao M, Zhang K, et al., 2014 Subtypes of major depression: latent class analysis in depressed Han Chinese women. Psychol. Med. 44, 3275–88. doi:10.1017/S0033291714000749 [PubMed: 25065911]

Lubke GH, Luningham J, 2017 Fitting latent variable mixture models. Behav. Res. Ther. doi:10.1016/j.brat.2017.04.003

Masyn KE, 2013 Latent Class Analysis and Finite Mixture Modeling, in: Little TD (Ed.), The Oxford Handbook of Quantitative Methods. Oxford University Press, New York.

Milaneschi Y, Lamers F, Bot M, Drent ML, Penninx BWJH, 2017 Leptin dysregulation is specifically associated with major depression with atypical features: evidence for a mechanism connecting obesity and depression. biol. Psychiatry 81, 807–814. doi:10.1016/j.biopsych.2015.10.023

Milaneschi Y, Lamers F, Mbarek H, Hottenga J-J, Boomsma DI, Penninx BWJH, 2014 The effect of FTO rs9939609 on major depression differs across MDD subtypes. Mol. Psychiatry 19, 960–962. doi:10.1038/mp.2014.4

Milaneschi Y, Lamers F, Peyrot WJ, Abdellaoui A, Willemsen G, Hottenga J-J, et al., 2016 Polygenic dissection of major depression clinical heterogeneity. Mol. Psychiatry 21, 516–522. doi:10.1038/mp.2015.86 [PubMed: 26122587]

Mitchell AJ, Vaze A, Rao S, 2009 Clinical diagnosis of depression in primary care: a meta-analysis. Lancet 374, 609–619. doi:10.1016/S0140-6736(09)60879-5 [PubMed: 19640579]

Moher D, Liberati A, Tetzlaff J, Altman DG, 2009 Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 6, e1000097. doi:10.1371/journal.pmed.1000097 [PubMed: 19621072]

Muthen B, Muthen LK, 2000 Integrating Person-Centered and Variable-Centered Analyses: Growth Mixture Modeling With Latent Trajectory Classes. Alcohol. Clin. Exp. Res. 24, 882–891. doi:10.1111/j.1530-0277.2000.tb02070.x [PubMed: 10888079]

National Center for Health Statistics, 2011 Health, United States, 2010: With Special Feature on Death and Dying. Hyattsville, MD.

Ostergaard SD, Jensen SOW, Bech P, 2011 The heterogeneity of the depressive syndrome: when numbers get serious. Acta Psychiatr. Scand. 124, 495–6. doi:10.1111/j.1600-0447.2011.01744.x [PubMed: 21838736]

Parker G, Hadzi-Pavlovic D, 1993 Old data, new interpretation: a re-analysis of Sir Aubrey Lewis' M.D. thesis. Psychol. Med 23, 859–70. [PubMed: 8134511]

Parker G, Hadzi-Pavlovic D, Boyce P, Wilhelm K, Brodaty H, Mitchell, et al., 1990 Classifying depression by mental state signs. Br. J. Psychiatry 157, 55–65. [PubMed: 2397363]

Parker G, Hadzi-Pavlovic D, Brodaty H, Austin MP, Mitchell P, Wilhelm K, et al., 1995 Sub-typing depression, II. Clinical distinction of psychotic depression and non-psychotic melancholia. Psychol. Med. 25, 825–32. [PubMed: 7480460]

Parker G, Hadzi-Pavlovic D, Hickie I, Boyce P, Mitchell P, Wilhelm K, et al., 1991 Distinguishing psychotic and non-psychotic melancholia. J. Affect. Disord. 22, 135–48. [PubMed: 1918657]

Parker G, Hadzi-Pavlovic D, Roussos J, Wilhelm K, Mitchell P, Austin MP, et al., 1998 Non-melancholic depression: the contribution of personality, anxiety and life events to subclassification. Psychol. Med. 28, 1209–19. [PubMed: 9794028]

Parker G, Wilhelm K, Mitchell P, Roy K, Hadzi-Pavlovic D, 1999 Subtyping depression: testing algorithms and identification of a tiered model. J. Nerv. Ment. Dis. 187, 610–7. [PubMed: 10535654]

Prisciandaro JJ, Roberts JE, 2009 A comparison of the predictive abilities of dimensional and categorical models of unipolar depression in the National Comorbidity Survey. Psychol. Med. 39, 1087–96. doi:10.1017/S0033291708004522 [PubMed: 18845012]

Rabe-Hesketh S, Skrondal A, 2008 Classical latent variable models for medical research. Stat. Methods Med. Res. 17, 5–32. doi:10.1177/0962280207081236 [PubMed: 17855748]

Rodgers S, Ajdacic-Gross V, Müller M, Hengartner MP, Grosse Holtforth M, Angst J, et al., 2014a The role of sex on stability and change of depression symptom subtypes over 20 years: a latent transition analysis. Eur. Arch. Psychiatry Clin. Neurosci. 264, 577–88. doi:10.1007/s00406-013-0475-3 [PubMed: 24292327]

Rodgers S, Grosse Holtforth M, Hengartner MP, Müller M, Aleksandrowicz AA, Rössler W, et al., 2015 Serum testosterone levels and symptom-based depression subtypes in men. Front. psychiatry 6, 61. doi:10.3389/fpsyt.2015.00061 [PubMed: 25999864]

Rodgers S, Grosse Holtforth M, Müller M, Hengartner MP, Rössler W, Ajdacic-Gross V, 2014b Symptom-based subtypes of depression and their psychosocial correlates: a person-centered approach focusing on the influence of sex. J. Affect. Disord. 156, 92–103. doi:10.1016/j.jad.2013.11.021 [PubMed: 24373526]

Ruscio J, Ruscio AM, 2008 Categories and dimensions advancing psychological science through the study of latent structure. Curr. Dir. Psychol. Sci. 17, 203–207. doi:10.1111/j.1467-8721.2008.00575.x [PubMed: 19727339]

Rush AJ, 2007 The varied clinical presentations of major depressive disorder. J. Clin.Psychiatry 68 Suppl 8, 4–10.

Schreiber JB, 2016 Latent Class Analysis: An example for reporting results. Res. Social Adm. Pharm. doi:10.1016/j.sapharm.2016.11.011

Sneed JR, Rindskopf D, Steffens DC, Krishnan KRR, Roose SP, 2008 The vascular depression subtype: evidence of internal validity. Biol. Psychiatry 64, 491–7. doi:10.1016/j.biopsych.2008.03.032 [PubMed: 18490003]

Sullivan PF, Kessler RC, Kendler KS, 1998 Latent class analysis of lifetime depressive symptoms in the National Comorbidity Survey. Am. J. Psychiatry 155, 1398–406. [PubMed: 9766772]

Sullivan PF, Prescott CA, Kendler KS, 2002 The subtypes of major depression in a twin registry. J. Affect. Disord. 68, 273–84. [PubMed: 12063155]

Sunderland M, Carragher N, Wong N, Andrews G, 2013 Factor mixture analysis of DSM-IV symptoms of major depression in a treatment seeking clinical population. Compr. Psychiatry 54, 474–83. doi:10.1016/j.comppsych.2012.12.011 [PubMed: 23357125]

Ten Have M, Lamers F, Wardenaar K, Beekman A, de Jonge P, van Dorsselaer S, et al., 2016 The identification of symptom-based subtypes of depression: A nationally representative cohort study. J. Affect. Disord. 190, 395–406. doi:10.1016/j.jad.2015.10.040 [PubMed: 26546775]

Ulbricht CM, Rothschild AJ, Lapane KL, 2015 The association between latent depression subtypes and remission after treatment with citalopram: A latent class analysis with distal outcome. J. Affect. Disord. 188, 270–277. doi:10.1016/j.jad.2015.08.039 [PubMed: 26384013]

van Loo HM, de Jonge P, Romeijn J, Kessler RC, Schoevars RA 2012 Data-driven subtypes of major depressive disorder: a systematic reivew. BMC Medicine. 10:156. [PubMed: 23210727]

Wang PS, Insel TR, 2010 NIMH-funded pragmatic trials: moving on. Neuropsychopharmacology 35, 2489–90. doi:10.1038/npp.2010.161 [PubMed: 21068745]

Wurpts IC, Geiser C, 2014 Is adding more indicators to a latent class analysis beneficial or detrimental? Results of a Monte-Carlo study. Front. Psychol. 5, 920. doi:10.3389/fpsyg.2014.00920 [PubMed: 25191298]

Highlights

- Successful treatment of depression is complicated by heterogeneity in symptoms, severity, and course of the disorder.
- Latent class analysis (LCA), which can discern patterns from many depression characteristics, has been used to identify depression subtypes for many years but results have been inconsistent.
- Depression subtypes resulting from LCA models are most often distinguished by depression severity, with domains such as functioning and neurobiological measures rarely considered.
- The utility of depression subtypes resulting from LCA could be improved through standards for conducting and interpreting LCA models.

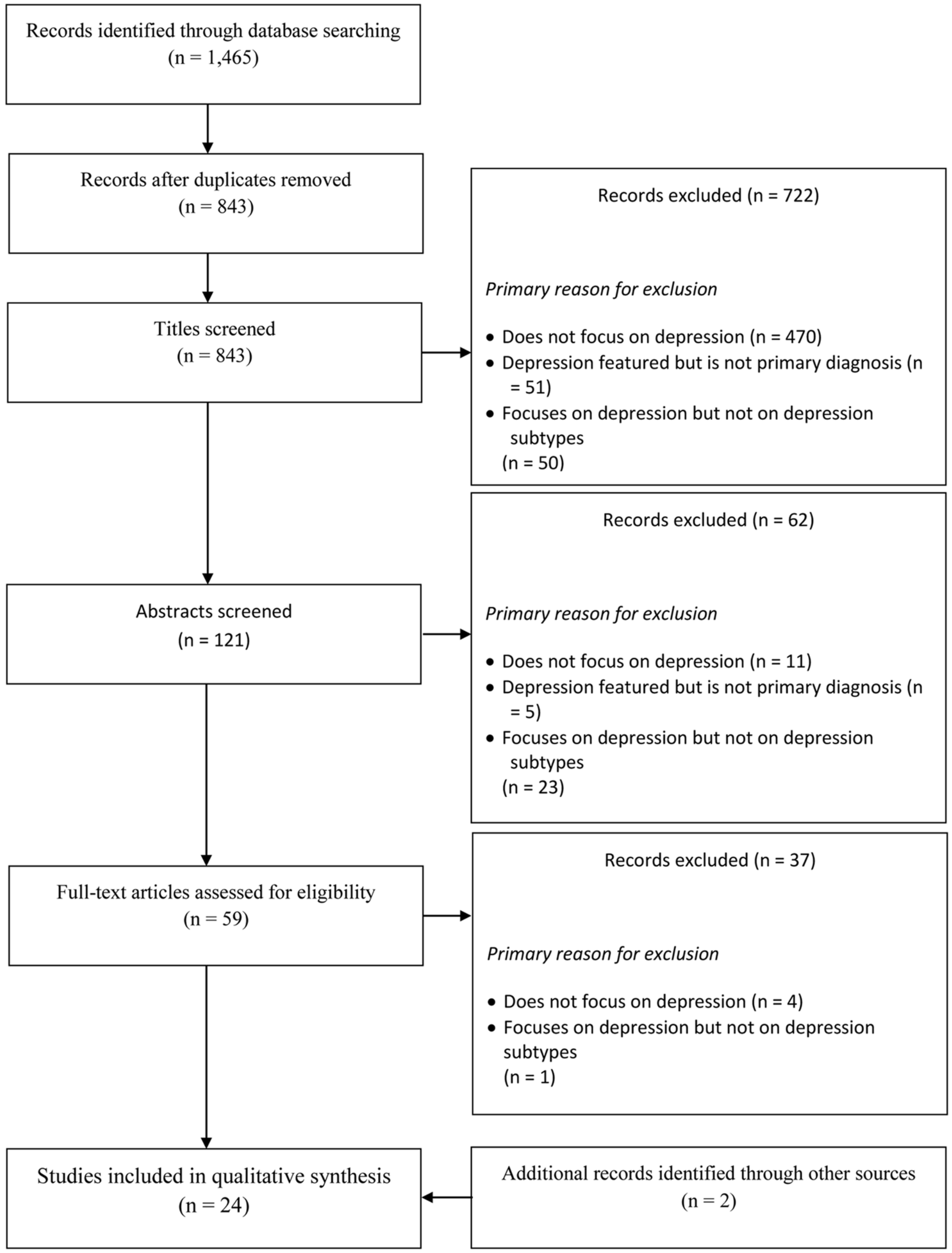


Figure 1.
PRISMA flow diagram of literature search and study selection process

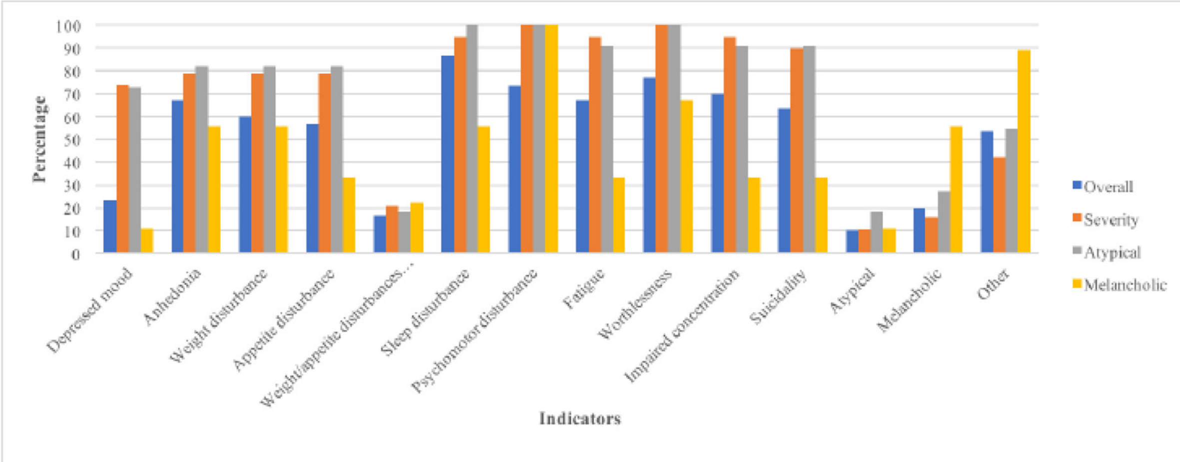
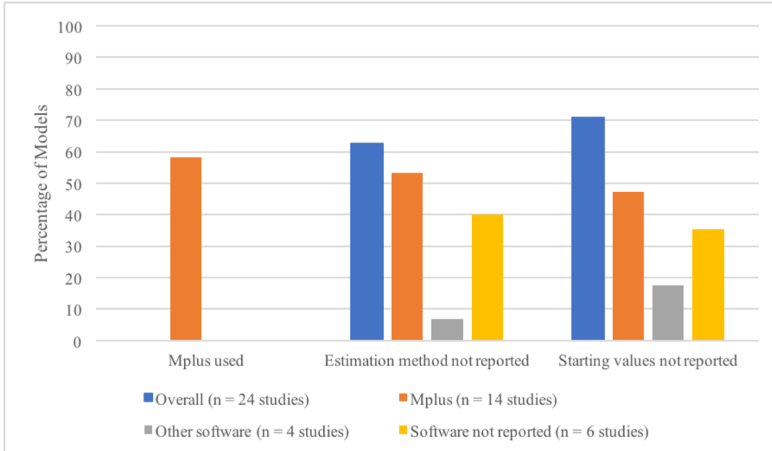


Figure 2.

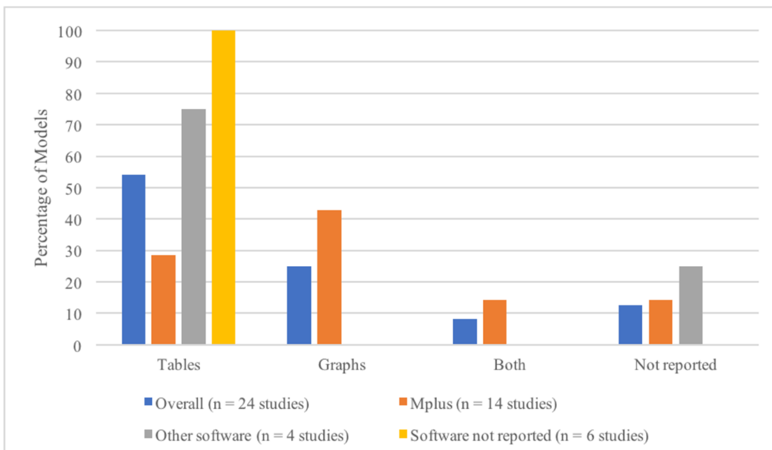
Percentage of models evaluating depression symptoms as indicators by labels for resulting latent classes described by depression severity, atypical features, and melancholic features[a]

[a]Categories of latent classes are not mutually exclusive because some models had classes distinguished by severity, atypical features, and/or melancholic features.

3.1 Model estimation



3.2 Model specification



3.3 Model selection

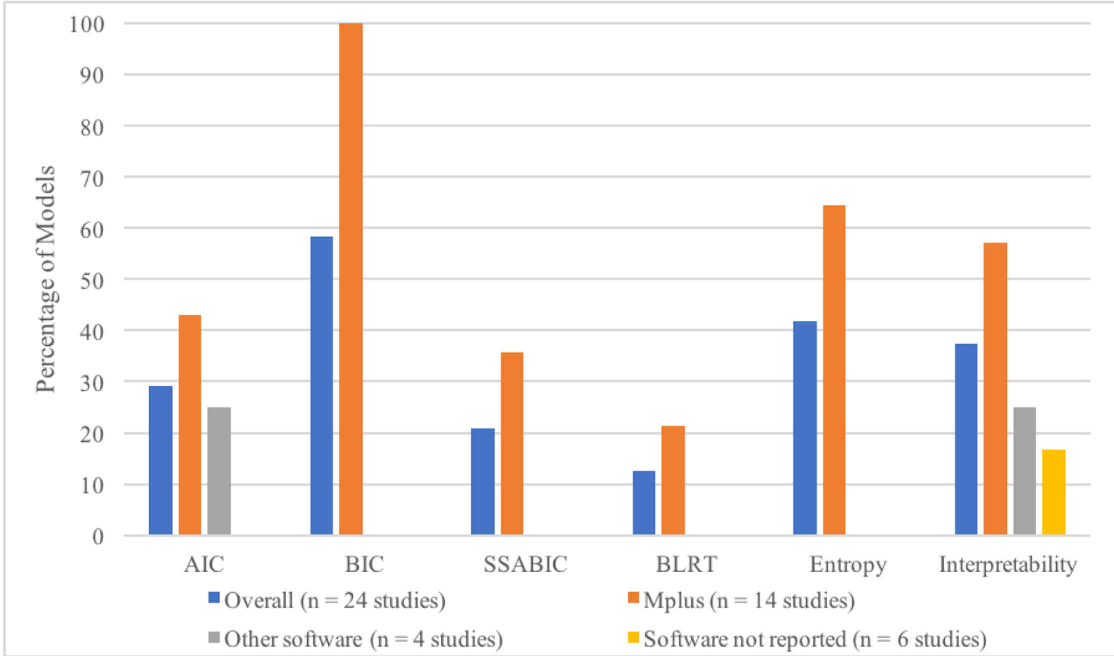


Figure 3.
Model Characteristics by Software

Table 1.

General description of eligible studies

Reference	Study Description		Depression Assessment Instruments & Time Frame	LCA Indicators	Results of Final LCA Model
	Study Name, Year, Location, Setting	Sample	Subtypes identified (prevalence)		
Alexandrino-Silva et al., 2013	São Paulo Megacity Mental Health Survey 2005-2007 Brazil Community	*N* = 1,207 Participants with indication of depression symptoms > several days for ≥2 weeks	WMH-CIDI translated & adapted to Brazilian-Portuguese Worst episode, lifetime	Considered 34 DSM-IV/ICD-10 depression symptoms Included 21 measures of criteria symptoms & 9 measures of anxiety, sociability	3 classes Women • Mild (41.1%) • Melancholic (39.3%) • Atypical (19.5%) Men • Mild (40.1%) • Melancholic/ psychomot or retarded (40.4%) • Agitated depression (19.6%)
Carragher et al., 2009	NESARC Wave 2001-2002 U.S.A. Community	*N* = 12,180 Participants with lifetime occurrence of a 2-week period of depressed mood or loss of interest in activities	AUDADIS-IV Lifetime occurrence of a 2-week period of depressed mood or loss of interest in activities	7 measures of DSM-IV depression criteria symptoms	4 classes • Non-depressed (18.3%) • Psychosomatic (30.6%) • Cognitive-emotional (10.2%) • Severely depressed (40.9%)
de Vos et al., 2015	*Study name & year not reported* Netherlands Day-care depression unit	*N* = 147 Outpatient participants	QIDS-SR $_{16}$ Not explicitly stated, but standard time frame for instrument is previous 1-2 weeks	12 measures of DSM-IV criteria symptoms	2 classes[a] • Low severity (28.4%) • High severity (71.6%)
Grove et al., 1987	NIMH-CRB Collaborative Study of the Psychobiology of Depression – Clinical *Years not reported* U.S.A. Academic medical centers	*N* = 512 Participants without a history of mania/ hypomania who started study in a major depressive episode defined by Research Domain Criteria	SADS Worst level during episode	Considered 36 affective symptoms from the SADS, 28 of which covaried positively per Golden's requirements. 17 of these pass Golden's mathematical consistency tests & 11 had poor separation or did not fit model	2 classes • Nuclear (48.4%) • Nonnuclear (51.6%)
Lamers et al., 2010	Baseline NESDA 2004-2007 Netherlands General population, primary care sites, mental health care sites	*N* = 818 Participants with diagnosis of MDD or minor depression in past month	CIDI, lifetime version 2.1 for 9 DSM-IV depression symptoms; 6 items from the 30-item IDSSR to address DSM-IV atypical & melancholic depression specifiers Lifetime	10 measures of DSM-IV depression criteria symptoms, 6 measures of atypical/ melancholic features	3 classes • Moderate (29.1%) • Severe melancholic (46.3%) • Severe atypical (24.6%)
Lamers et al., 2012	NCS-R 2001-2003 U.S.A. Community	*N* = 805 Participants diagnosed 12-month MDD	CIDI 3.0 Current	10 measures of DSM-IV depression criteria symptoms comprising 14 symptoms because weight and appetite were separated	4 classes • Moderate (14.6%) • Moderate typical (24.8%) • Severe typical (44.9%) • Atypical (15.6%)

	Study Description		Depression Assessment Instruments & Time Frame		Results of Final LCA Model
Reference	**Study Name, Year, Location, Setting**	**Sample**	**Subtypes identified (prevalence)**	**LCA Indicators**	
Lee et al., 2012	Cache County Memory Study 1995 U.S.A. Community	*N* = 400 Participants ≥ 65 years of ages with ≥1 current depressive symptom at baseline, 3-year or 7-year follow-up	Modified DIS Current	9 measures of depression symptoms	3 classes • Minor depression (21.5%) • Major depression (61.5%) • Minor or major depression + psychomotor symptoms (17.0%)
Lee et al., 2014	NESARC wave 1 2001-2002 U.S.A. Community	*N* = 13,424 Participants who endorsed lifetime criterion A DSM-IV MDD symptoms	AUDADIS-IV Lifetime	7 measures of DSM-IV criteria symptoms	4 classes[b] • Mild (17.0-25.0%) • Cognitive (10.0-13.0%) • Psychosomatic (26.0-32.0%) • Severe (37.0-41.0%)
Li et al., 2014	CONVERGE study *Years not reported* China Mental health centers & psychiatric departments	*N* = 6,008 Women ages 30-60 years with DSM-IV major depression	CIDI, WHO lifetime version 2.1, Chinese version; SCID-P items; VATSPUD items; medical records Lifetime worst episode	9 measures of DSM-IV depression criteria symptoms	3 classes • Moderate (6.4%) • Non-suicidal (8.0%) • Severe (85.6%)
				14 measures of DSM-IV depression criteria symptoms	4 classes • Moderate typical (13.6%) • Non-suicidal (8.7%) • Severe typical (70.4%) • Atypical (7.4%)
				18 measures of DSM-IV depression criteria symptoms, 7 measures of melancholic features, & 2 measures of Beck's cognitive trio	6 classes • Mild (5.8%) • Moderate typical (19.5%) • Non-suicidal low guilt (10.9%) • Low suicidal high guilt (17.0%)) • Severe typical (39.5%) • Atypical (7.3%)
Parker et al., 1990	*Study name not reported* *Years not reported* Australia Inpatient & outpatient psychiatric facilities	*N* = 262 Depressive patients	Clinician rating assessments; self-reported data on depression & anxiety symptoms, life stressors, Zung depression scale *Time frame not reported*	Considered: 2 symptoms (guilt and nihilism) 1 item capturing 'endogenous' quality 27 signs which captured general attitude; severity of mood; responsiveness to the interviewer; retardation and agitation as expressed facially, in posture, speech, and in movement; reactivity; speech; concentration; and insight Included core items: unresponsive to interviewer, dull/ inattentive, fixed/ immobile face, self-preoccupied, inability	2 classes • Non-melancholia (43.0%) • Melancholia (57.0%)
	Specialized Mood Disorders Unit	*N* = 202 Depressive patients not in remission	Semi-structured clinical interview, Newcastle scale, Zung depression scale, General Health Questionnaire, 21-item HAM-D *Time frame not reported*		2 classes • Non-melancholia (67.0%) • Melancholia (33.0%)

	Study Description		**Depression Assessment Instruments & Time Frame**		**Results of Final LCA Model**
Reference	**Study Name, Year, Location, Setting**	**Sample**	**Subtypes identified (prevalence)**	**LCA Indicators**	
				to be cheered by interviewer, slumped posture, immobility, slowed movements, slowed speech, mute or reduced speech, poverty of associations, impaired insight, nihilistic, observable anxiety, endogenous quality.	
Parker et al., 1991	*Study name not reported* *Years not reported*	*N* = 136 34 psychotic depressives and 102 endogenous depressives	Semi-structured interview, Zung depression scale, 21-item HAM-D, Newcastle depression index, Core diagnostic system Episode	Hallucination or delusion, constipation, no diurnal mood variation, sustained depressive content, high core score (based on core variables from Parker 1990)	2 classes • Non-psychotic melancholia (64.0%) • Psychotic melancholia (36.0%)
	Australia Specialized Mood Disorders Unit	From this group, patients with psychotic melancholic depression were age- and sex-matched to endogenous depressives.			2 classes • Non-psychotic melancholia (50%) • Psychotic melancholia (50%)
Parker et al., 1993	Sir Aubrey Lewis' MD thesis 1928-1929 London, England Maudsley Hospital	*N* = 61 Patients with depression	Clinician examination *Time frame not reported*	Considered 21 items: 7 psychotic disturbance measures, 8 psychotic features, & 6 endogeneity measures Included 14 items: 5 psychotic disturbance measures, 6 psychotic features, & 3 endogeneity measures	2 classes • Melancholia (46.0%) • Non-melancholia (54.0%)
Parker et al., 1995	*Study name not reported* *Years not reported* *Location not reported*	*N* = 327 Diagnosed with depression	*Not reported*	3 measures of psychomotor disturbances, 11 measures of endogeneity and/or psychoticism	3 classes • Non-melancholia (40.0%) • Melancholia (35.0%) • Psychotic depression (25.0%)
Parker et al., 1998	*Study name not reported* *Years not reported* Australia Tertiary referral Mood Disorders Unit and other hospitals	*N* = 185 Participants with non-melancholic DSM-III-R MDE for < 2 years	Clinical interview, Newcastle Index, CIDI 1.1, Beck Depression Inventory, mood/affect & anxiety symptom checklists, 17-item HAM-D, CORE psychomotor disturbance item *Time frame not reported*	Considered 13 items (6 measures of disordered personality, 5 measures of anxiety, & 2 measures of life event stressors) Included 11 items (6 measures of disordered personality & 5 measures of anxiety)	4 classes • Neither disordered personality nor anxiety (39.0%) • Anxiety without disordered personality (27.0%) • Disordered personality without anxiety⁻ (17.0%) • Disordered personality & anxiety (17.0%)
Parker et al., 1999	*Study name not reported* *Years not reported* Australia	*N* = 269 Participants with DSM-III-R MDE for < 2 years	Semi-structured interviews, Newcastle endogeneity scale, HAM-D, CORE scores Lifetime & current	Considered 13 symptoms and 3 scales for agitation, retardation, and non-interactiveness	3 classes • Nonmelancholic (55.0%) • Melancholic (34.0%)

	Study Description		**Depression Assessment Instruments & Time Frame**		**Results of Final LCA Model**
Reference	**Study Name, Year, Location, Setting**	**Sample**	**Subtypes identified (prevalence)**	**LCA Indicators**	
	Inpatient & outpatient hospital sites				• Putative psychotic (11.0%)
Priscian daro & Roberts, 2009	NCS *Study year not reported* U.S.A. Community	*N* = 1,402 Participants with ≥3 co-occurring depressive symptoms, including either sad mood or anhedonia, during lifetime.	CIDI 2-week period during lifetime when experienced "some degree of symptomatology"	9 measures of DSM-III-R criteria symptoms	3 classes • Moderate depression (51.3%) • Cognitive-affective distress (22.9%) • Severe depression (25.8%)
Rodgers et al., 2014a	Zurich Study 1988-2008 Switzerland Community	1998: *N* = 192; 1999: *N* = 184; 2008: *N* = 146 Participants who: 1) were recruited in 1978 as young adults 2) completed study interviews in 1988, 1999, and 2008 answered affirmatively to depression screening questions	SPIKE Previous 12 months	15 measures of DSM-IV major depression A criteria	3 classes[c,d] • Severe typical (21.2-22.3%) • Severe atypical (14.6-54.3%) • Moderate (23.4-64.1%)
Rodgers et al., 2014b	ZInEP 2010-2012 Switzerland Community	*N* = 816 Participants aged 20-41 years who endorsed the first or second filter question of the Mini-SPIKE depression section	Mini-SPIKE Past 12 months	17 measures of DSM-IV depression criteria symptoms including 3 measures of atypical features	Men[d] 5 classes • Moderate (25.2%) • Psychomotor retarded (4.3%) • Severe irritable/ angry-rejection sensitive (30.3%) • Severe typical (22.8%) • Severe atypical (17.4%) Women 3 classes • Moderate (41.8%) • Severe typical (35.7%) • Severe atypical (22.6%)
Sneed et al., 2008	*Years & locations not reported*				2 classes
	Neurocognitive Outcomes of Depression in the Elderly (NCODE) Prospective cohort	*N* = 150 Participants aged ≥60 years, with DSM-IV single episode or recurrent major depression; who were treated in outpatient, academic medical setting; & had MMSE scores ≥25	DIS *Time frame not reported*	1 measure of late-onset depression, 1 measure of executive dysfunction, 2 measures of hyperintensity burden	• Vascular depressed (49.0%) • Non-vascular depressed (51.0%)
	Old-Old Study Multi-site RCT of citalopram vs. placebo	*N* = 97 Participants aged ≥75 years, with DSM-IV single episode or recurrent major depression, not living in residential setting,	HAM-D *Time frame not reported*		• Vascular depressed (48.0%) • Non-vascular depressed (52.0%)

Reference	Study Description: Study Name, Year, Location, Setting	Study Description: Sample	Depression Assessment Instruments & Time Frame: Subtypes identified (prevalence)	LCA Indicators	Results of Final LCA Model
		had HAM-D_{24} ≥ 24 & had MMSE ≥ 19			
Sullivan et al., 1998	NCS *Years not reported* U.S.A. Community	*N* = 2,836 Participants whose worst lifetime depressive episodes lasted ≥ 2 weeks, were associated with help-seeking or impairment, & had ≥1 contemporaneous depressive symptoms	CIDI Worst lifetime episode	14 measures of DSM-III-R major depression criteria symptoms	6 classes (prevalence in NCS) • Minimal symptoms (8.0%) • Mild typical (6.0%) • Mild atypical (4.0%) • Intermediate (11.0%) • Severe typical (4.0%) • Severe atypical (2.0%)
Sullivan et al., 2002	Virginia Twin Registry *Year not reported* U.S.A. Population-based twin registry	*N* = 2,941 Participants who endorsed ≥ 1 depression symptom within prior year	18 questions about symptoms, 14 of which represent the DSM-III-R A criteria for major depression Symptom must have lasted for ≥ 5 days in prior year	14 measures of DSM-III-R major depression criteria symptoms	7 classes • Minor typical (5.0%) • Typical (6.0%) • Atypical (3.0%) • Non-appetitive (4.0%) • Mood only (17.0%) • Overeat (2.0%) • Agitated (6.0%)
Sunderland et al., 2013	*Study name not reported* 2009-2011 Australia & New Zealand Online clinical treatment program	*N* = 1,165 Patients referred by clinician	PHQ-9 Prior 2 weeks	9 measures of DSM-IV depression criteria symptoms	3 classes • Low responders (27.9%) • Intermediate responders (42.4%) • High responders (29.7%)
ten Have et al., 2016 [epub 2015]	NEMESIS-2 Baseline: 2007-2009 Follow-up: 2010-2012 The Netherlands Community	*N* = 1,388 Participants, ages 18-64 years, who reported lifetime key depression symptom at baseline	CIDI 3.0 Lifetime	Considered 10 measures of depression symptoms, 2 measures of anxiety, 1 measure of irritability & 1 measure of racing thoughts Included all but racing thoughts	4 classes • Mild depression (19.0%) • Moderate depression without anxiety (23.6%) • Moderate depression with anxiety (29.3%) • Severe depression with anxiety (28.0%)
Ulbricht et al., 2015	STAR*D level 1 2001-2004 U.S.A. Outpatient primary care & psychiatric sites	*N* = 2,772 Clinical trial participants with moderate to severe non-psychotic depression	QIDS-SR_{16} Previous 1-2 weeks	16 measures of DSM-IV depression criteria symptoms	4 classes[e] • Mild (27.0-37.0%) • Moderate (21.0-24.0%) • Severe with increased appetite (13.0-22.0%) • Severe with insomnia (26.0-31.0%)

Table 2.

Indicators used in LCA models by depression assessment instrument

Reference	Total number	Categorization	Number of Indicators Representing Each Symptom												
			DSM Diagnostic Criteria										Additional DSM Specifiers		Other
			Depressed	Anhedonia	Weight disturbance	Appetite disturbance	Sleep disturbance	Psychomotor disturbance	Fatigue	Guilt/ worthlessness	Impaired concentration	Suicidality	Atypical	Melancholic	
CIDI Only															
Alexandrino-Silva et al., 2013															
Full set	34	Binary	0	0	2	2	2	4	1	3	2	5	0	0	13[i]
Reduced set	30	Binary	0	0	2	2	2	4	1	3	2	5	0	0	9[ii]
Lamers et al., 2010	16														
	12	Binary	1	1	0	0	0	0	1	1	1	1	3	3	0
	2	3 levels	0	0	1	1	0	0	0	0	0	0	0	0	0
	2	4 levels	0	0	0	0	1	1	0	0	0	0	0	0	0
Lamers et al., 2012	10														
	6	Binary	1	1	0	0	0	0	1	1	1	1	0	0	0
	4	3 levels	0	0	1	1	1	1	0	0	0	0	0	0	0
Li et al., 2014															
Set 1	9	Binary	1	1	1		1	1	1	1	1	1	0	0	0
Set 2	14	Binary	1	1	2	2	2	2	1	1	1	1	0	0	0
Set 3	27	Binary	1	1	2	2	2	2	1	2	3	2	0	3	6[iii]
Prisciand aro & Roberts, 2009	9	Binary	1	1	1		1	1	1	1	1	1	0	0	0
Sullivan et al., 1998	15	Binary	1	1	2	2	2	2	1	1	1	1	0	0	1[iv]
ten Have et al., 2016															
Full set	14														
	9	Binary	1	1	0	0	0	0	1	1	0	1	0	0	4[v]
	4	3 levels	0	0	1	1	1	1	0	0	0	0	0	0	0

Reference	Total number	Categorization	Number of Indicators Representing Each Symptom												
			DSM Diagnostic Criteria									Additional DSM Specifiers		
			Depressed	Anhedonia	Weight disturbance	Appetite disturbance	Sleep disturbance	Psychomotor disturbance	Fatigue	Guilt/ worthlessness	Impaired concentration	Suicidality	Atypical	Melancholic	Other
	1	4 levels	0	0	0	0	0	0	0	0	1	0	0	0	0
Refined set	13														
	8	Binary	1	1	0	0	0	0	1	1	0	1	0	0	3[vi]
	4	3 levels	0	0	1	1	1	1	0	0	0	0	0	0	0
	1	4 levels	0	0	0	0	0	0	0	0	1	0	0	0	0
AUDADIS-IV Only															
Carragher et al., 2009	7	Binary	0	0	1	1	1	1	1	1	1	0	0	0	0
Lee et al., 2014	7	Binary	0	0	1		1	1	1	1	1	1	0	0	0
QIDS-SR$_{16}$ Only															
de Vos et al., 2015	12	3 levels	1	1	1	1	2	2	1	1	1	1	0	0	0
Ulbricht et al., 2015	16	Binary	1	1	2	2	4	2	1	1	1	1	0	0	0
Other Assessment Instruments															
Grove et al., 1987	17-36 [exact number unclear]	Binary	*Not reported*												
Lee et al., 2012	9	Binary	1	1	1		1	1	1	1	1	1	0	0	0
Parker et al., 1990	30	Binary	0	0	0	0	0	2	0	1	0	0	0	0	See Table 1
Parker et al., 1991	5	Binary	0	0	0	0	0	1	0	0	0	0	0	1	3[vii]
Parker et al., 1993															
Full set	21	Binary	0	1	2	0	0	7	0	0	0	0	0	1	10[viii]
Refined set	14	Binary	0	1	1	0	0	5	0	0	0	0	0	0	*ix*
Parker et al., 1995	14	Binary	0	2	1		1	2	0	2	0	0	0	2	4[x]
Parker et al., 1998															
Full set	13	Binary	0	0	0	0	0	0	0	0	0	0	0	0	13[xi]
Refined set	11	Binary	0	0	0	0	0	0	0	0	0	0	0	0	11[xii]

Reference	Total number	Categorization	Number of Indicators Representing Each Symptom												
			DSM Diagnostic Criteria										Additional DSM Specifiers		Other
			Depressed	Anhedonia	Weight disturbance	Appetite disturbance	Sleep disturbance	Psychomotor disturbance	Fatigue	Guilt/ worthlessness	Impaired concentration	Suicidality	Atypical	Melancholic	
Parker et al., 1999	16	3 levels	0	2	1		1	4	0	2	0	0	0	2	4 [xiii]
Rodgers et al., 2014a	15	Binary	1	1	2	2	2	2	1	1	1	1	1	0	0
Rodgers et al., 2014b	17	Binary	1	1	2	2	2	2	1	1	1	1	1	0	2 [xiv]
Sneed et al., 2008	4	Binary	0	0	0	0	0	0	0	0	0	0	0	0	4 [XV]
Sullivan et al., 2002	14	Binary	1	1	2	2	2	2	1	1	1	1	0	0	0
Sunderland et al., 2013	9	4 levels	1	1	0	1	1	1	1	1	1	1	0	0	0

[i] slowed thinking, racing thoughts, loss of confidence, feeling worse than others, irritability, anxiety, social isolation, being less talkative, crying, difficulty coping with responsibilities, insomnia (but feel alert and rested), panic

[ii] slowed thinking, racing thoughts, loss of confidence, feeling worse than others, irritability, anxiety, social isolation, being less talkative

[iii] loss of sexual drive, irritability/anger, hopeless, cry a lot, helpless, nervous/jittery/anxious

[iv] gender

[v] anxiety, panic attacks, irritability, racing thoughts

[vi] anxiety, panic attacks, irritability

[vii] hallucination or delusion, constipation, sustain depressive content

[viii] delusions of poverty & ruin, conviction would never get well, hypochondriacal delusions, ideas of influencing others, ideas of reference or of persecution, delusions of “pronounced apprehensive cast”, perception disorders, denial of illness, self-reproach, lack of response to reassurance

[ix] delusions of poverty & ruin, ideas of influencing others, ideas of reference or of persecution, perception disorders, denial of illness, self-reproach, lack of response to reassurance

[x] constipation, hallucinations, delusions, CORE non-interactiveness scale

[xi] inadequate personality; disordered personality; eccentric personality style; dramatic personality style; sensitive personality style; anxious personality style; nervy; worrier; tense; anxious; lifetime anxiety disorder; DSM acute stressor; consensus antecedent stressor

[xii] inadequate personality; disordered personality; eccentric personality style; dramatic personality style; sensitive personality style; anxious personality style; nervy; worrier; tense; anxious; lifetime anxiety disorder

[xiii] delusions, hallucinations, constipation, CORE noninteractiveness

[xiv] irritable/angry, hypersensitive to critical remarks

[xv]late-onset depression, executive dysfunction, deep white matter lesions, subcortical gray matter lesions

Table 3.

Evaluation of Measurement Invariance

Reference	Variables evaluated	Rationale for selecting variables	How evaluated	Findings
Alexandrino-Silva et al., 2013	Gender	Literature supporting gender differences in prevalence of depressive subtypes.	Using KNOWNCLASS option in Mplus, multi-group LCA was conducted and confirmed distinct symptom profiles by gender. LCA then conducted in separate samples for men and women.	Both men and women had a 3-class solution but the classes for women were melancholic, atypical, and mild. The classes for men were melancholic/ psychomotor retarded, agitated, and mild.
Lee et al., 2014	Racial/ethnic groups	Literature supporting less help-seeking behavior and greater reports of physical symptoms than psychological symptoms among racial/ethnic minorities relative to U.S.- born whites.	Stratified analysis on racial/ ethnic groups.	Qualitatively, the 4 latent variable classes (mild, cognitive, psychosomatic, severe) were the same by race/ethnicity. The prevalence estimates differed by race/ethnicity.
Rodgers et al., 2014b	Sex and gender role orientation	Literature documenting depressive symptom presentation in women, a hypothesis about a masculine depressive syndrome, and concerns about gender role orientation.	Stratified LCA analyses by sex.	Five-class solution was chosen for men (severe irritable/angry rejection sensitive, severe atypical, severe typical, psychomotor retarded, moderate) and a 3-class solution for women (severe atypical, severe typical, moderate).
Ulbricht et al., 2015	Sex	Literature supporting different presentation of depression by sex.	Difference G^2 likelihood ratio test to see if item-response probabilities were the same in a model that constrained the item-response probabilities to be the same by sex and a model allowing item response probabilities to differ by sex.	Qualitatively, the latent classes were the same by sex. Prevalence of each class differed by sex.

Table 4.

Correlates/grouping variables explored

Reference	Method	Correlates							
		Current Age	Gender	Race/Ethnicity	Education	Marital Status	Psychiatric Comorbidities	Depression Characteristics	Other
Alexandrino-Silva et al., 2013	Multinomial al logistic regression (unclear if classify-analyze or if conducted within LCA)	ns	+ (As grouping variable)	na	ns	ns	Women: bipolar spectrum (+), any anxiety disorder (+), alcohol & drug dependence (+), nicotine dependence (ns), dysphoric disorder (+), nicotine dependence (ns) Men: bipolar spectrum (ns), any anxiety disorder (+), nicotine dependence (+)	na	Employment (ns), income (ns) Women: disability (ns), childhood adversity (ns)
Carragher et al., 2009	Calculated odds ratios and confidence intervals after including covariates in the LCA model	−	+/−	−	−	+	Lifetime mood disorder (+), lifetime personality disorder (+), lifetime anxiety disorder (+), alcohol abuse diagnosis (last 12 months) (ns), alcohol dependence diagnosis (last 12 months) (ns), alcohol abuse and dependence diagnosis (last 12 months) (+), any drug use disorder diagnosis (last 12 months) (ns), lifetime nicotine dependence diagnosis (+), any drug use	Family history (+)	Total personal income (−), urbanicity (−), negative life events (+)

Reference	Method	Correlates							
		Current Age	**Gender**	**Race/ Ethnicity**	**Education**	**Marital Status**	**Psychiatric Comorbidities**	**Depression Characteristics**	**Other**
							disorder diagnosis (last 12 months) (ns)		
de Vos et al., 2013	*No additional analyses of correlates reported*	na							
Grove et al., 1987	Classify-analyze with t-tests and χ^2 tests of "validating characteristics"	ns	ns	na	ns	ns		Age at onset (ns), number of hospitalizations (ns), number inpatient (*)	Occupation (*), childhood parental loss (ns)
Laniers et al., 2010	Classify-analyze with univariate analyses & multinomial logistic regression models.	ns	+	Na	ns	na	Panic disorder with agoraphobia (ns), panic disorder without agoraphobia (ns), social phobia (ns), agoraphobia (ns), GAD (ns), alcohol dependence (ns)	Age at onset (−), number of depressive episodes, family history (+), duration of depressive disorder (+), manic symptoms (ns)	Neuroticis m (+), extraversi on (−), negative life events(ns), childhood trauma (+/−), overall functionin g (+/−), physical activity (ns), current smoking (+/−), pain (ns), body mass index (+), somatic comorbidit ies (ns), metabolic syndrome (+), waist circumfere nce (+), triglycerid es (+), HDL cholesterol (ns), hypertensi on (ns), blood glucose (ns)
Lamers et al., 2012	Classify-analyze; posterior probabilitie s used to assign participants to most likely class; χ^2 and F-tests conducted to examine differences between classes	ns	*	na	na	na	Mania (ns), hypomania (*), dysthymia (*), GAD (ns), panic disorder (ns), social phobia (*), agoraphobia (*), specific phobia (*), substance use disorder (ns), any binge eating disorder (ns)	# of symptoms (*); # of episodes (*); age at onset (*); severity (*); family history of depression (ns) & mania (*); treatment in the past year for behavioral/ emotional problems: any mental healthcare (*), any healthcare (*)	Functional impairment (*), BMI (*), somatic disorders: heart attack (ns) & heart disease (ns), high blood pressure, diabetes (ns), migraine (ns), other headaches (ns)
Lee et al., 2012	Multinomi al logistic regression with pseudo-class draws with posterior probability based on class membership with multiple imputations.	ns	ns	na	ns	ns	na	Onset (ns), current antidepressant use (−), # of episodes (−), duration of current episode (ns)	Recent bereaveme nt (+), number of vascular health problems (ns), functional disability (−)

Reference	Method	Correlates							
		Current Age	**Gender**	**Race/ Ethnicity**	**Education**	**Marital Status**	**Psychiatric Comorbidities**	**Depression Characteristics**	**Other**
Lee et al., 2014	Classify-analyze; compared subtypes by service use	na	na	Separate models by race 9	na	na	na	n	Types of mental health service use: consulting counselors, doctors, and other professionals; being hospitalized overnight; emergency room visits; being prescribed medication to improve mood
Li et al., 2014	Compared differences in external validators with χ^2 and ANOVAs?	Sets 1-3: *	NB: only included women	na	na	na	GAD (sets 1-3: *), panic disorder (sets 1-3:*), phobia (sets 1-3:*)	Age of onset (set 1,3: ns; 2:*), duration of worst episode (set 1:ns; 2-3:*), # of episodes (set 1,3:*; 2:ns), dysthymia (set 1:ns; 2-3:*), family history (set 1-2:ns; 3:*)	BMI (set 1: ns; 2-3:*), neuroticis m (sets1-3:*), stressful life events (set 1:ns; 2-3:*), childhood sexual abuse (set 1:ns; 2-3:*)
Parker et al., 1990	*No additional analyses reported*	na							
Parker et al., 1991	*No additional analyses reported*	na							
Parker et al., 1993	*Classify-analyze, crude odds ratios*	ns	ns	na	na	ns	na	Family history (ns), course (+)	Personality style (ns)
Parker et al., 1995	*No analyses of correlates*	na							
Parker et al., 1998	Classify-analyze χ^2 and F tests to compare variables between latent classes	*	na	na	ns	*	Pre-morbid anxiety: childhood social phobia (*), behavioral inhibition (*), Costello-Comrey trait anxiety score (ns), total current anxiety symptom score (*). Drug & alcohol history: past use of anxiolytics for > 1 year (ns), past dependence on anxiolytic drugs (*), past	History: age at first episode (*), # of lifetime episodes (*), lifetime duration (ns), # of hospitalizations (ns). Current episode: duration (ns), GAF severity (*), Hamilton severity (*), Beck severity (*), mean Newcastle score (ns), mean CORE score (ns). Suicide/self-injury: History of attempts (*), age at first attempt (ns), self-injury (*), age at first self-injurious act (*significance not reported*)	Occupational status (ns) Family history of anxiety: mother-anxiety state (ns), father-anxiety state (ns), first-degree relative treated for nerves (ns). Family history of alcohol problems: mother and/or father (*), siblings (ns), first-degree relatives (*).

Reference	Method	Correlates							
		Current Age	Gender	Race/ Ethnicity	Education	Marital Status	Psychiatric Comorbidities	Depression Characteristics	Other
							excessive alcohol intake for > 1 year (ns).		
Parker et al., 1999	Classify-analyze dropping one of the latent classes χ^2 and t-tests to compare variables between latent classes	*	na	na	na	na	Previous manic/ hypom anic episode (*)	Family history (ns), age at first episode (*), # of lifetime episodes (ns), duration of lifetime episodes (ns), # of hospital admissions (*), time off work during lifetime (*), HAM-D (*), ECT (*), TCA treatment response (ns), ECT treatment response (ns), Newcastle Index (*)	Life event stressors (*), GAF (*)
Prisciand aro & Roberts, 2009	*No additional analyses of correlates reported*	na							
Rodgers et at., 2014a	Chi-square tests, Fisher's exact tests, Kruskal-Wallis tests, & multinomial logistic regression models	na	Analytic method unclear but significant differences found, with a higher proportion of women in the severe atypical class	na	na	na	na	na	na
Rodgers et al., 2014b	Chi-square tests, Fisher's exact tests, Kruskal-Wallis tests, & multinomial logistic regression models with latent classes as dependent & correlates as independent variables	Women & men: −	[originally included as covariate but then LCA models were run separately for men and women]	na	Women: + Men: ns	na	Women: Lifetime: affective disorder (+), anxiety disorder (+), psychosis syndromes (ns), alcohol/drug abuse/ dependence (ns), eating disorders (+/ −), personality disorder (ns); global severity index (−) Men: Lifetime: affective disorder (+), anxiety	Past year: MDD, dysthymia, male depressive syndrome (men only)	Women: urbanicity (ns), gender role orientation (ns) Men: urbanicity (+), gender role orientation (−)

Reference	Method	Correlates: Current Age	Gender	Race/ Ethnicity	Education	Marital Status	Psychiatric Comorbidities	Depression Characteristics	Other
							disorder (+), psychosis syndromes (na), alcohol/drug abuse/ depend ence (+), eating disorders (+), personality disorder (ns); global severity index (−)		
Sneed et al., 2008	*No additional analyses of correlates reported.*	na							
Sullivan et al., 1998	Classify-analyze χ^2 and t-tests to compare variables between latent classes Logistic and multiple regression models adjusted for age, race, & gender	na	ns	ns	na	ns	Bipolar 1 disorder (+), conduct disorder (+), antisocial personality (+), panic disorder (+), agoraphobia without panic disorder (+), GAD (+), social phobia (+/−), simple phobia (+), PTSD (+/−), alcohol dependence (+), drug dependence (+); parental history of major depression (ns) & alcohol/drug dependence (+)	Age at onset (ns), number of episodes (+), duration of longest episode (+), saw mental health worker (+), saw other professional (+), took medication (+), interfered with working/ socializing (+/−), hospitalization (+), number of consequences (+/−)	Personality traits: conformity (+), dependency (+), extraversion (ns), neuroticism (+), external locus of control (+), internal locus of control (ns), openness (ns), self-esteem (−), self-reliance (ns)
Sullivan et al., 2002	Linear and logistic regression models	The structure of the models conducted is not readily apparent from the manuscript. It is unclear if the latent class was treated as the dependent or the independent variables. As a result, the results are not summarized here.							
Sunderland et al., 2013	*No additional analyses reported*	na							
ten Have et al., 2016 Baseline	Classify-analyze Evaluated baseline correlates & outcomes with descriptive statistics	*	ns	na	*	na	Lifetime: any anxiety disorder (*), panic disorder (*), social phobia (*), specific phobia (*),	Lifetime: any mood disorder (*), major depression (*),dysthymia (*), bipolar disorder (*), number of episodes (*).	12-month: somatic disorder (*), cardiovasc ular disease (ns), diabetes (ns), Parents' mental health problems (*)

Reference	Method	Correlates							
		Current Age	**Gender**	**Race/ Ethnicity**	**Education**	**Marital Status**	**Psychiatric Comorbidities**	**Depression Characteristics**	**Other**
							GAD (*),any substance abuse disorder (*), alcohol abuse (ns), alcohol dependence (*), drug abuse (*), drug dependence (*), any impulse disorder (*), any Axis-1 disorder (*), antisocial personality disorder (ns), borderline personality disorder (*), any NEMESIS-2 disorder (*)	Age at onset (ns), persistency (*).	
Ulbricht et al., 2015	Examined correlates with multinomial logistic regression Evaluated measurement invariance with gender Explored LCA model with treatment remission as distal outcome	Women: − Men: −	As a grouping variable	Women : +/− Men: ns	na	na	Women: GAD (+), PTSD (+), bulimia (+), social phobia (+), any other psychiatric comorbidity (+/−) Men: GAD (+), PTSD (+/−), bulimia (+), social phobia (+), any other psychiatric comorbidity (+)		

Abbreviations & symbols: na = not assessed; ns = assessed but association not significant; * = significant association between covariates and latent classes in t-test, χ^2, F-test, Kruskal-Wallis test or ANOVA; + = positive association between covariate and latent class membership in logistic regression model; - = negative association between covariate and latent class membership in logistic regression model; GAD = generalized anxiety disorder; GAF = Global Assessment of Functioning; OCD = obsessive compulsive disorder; PTSD = post-traumatic stress disorder.