The effects of comorbidity in defining major depression subtypes associated with long-term course and severity

K. J. Wardenaar[1], **H. M. van Loo**[1], **T. Cai**[2], **M. Fava**[3], **M. J. Gruber**[4], **J. Li**[2], **P. de Jonge**[1], **A. A. Nierenberg**[5], **M. V. Petukhova**[4], **S. Rose**[4], **N. A. Sampson**[4], **R. A. Schoevers**[1], **M. A. Wilcox**[6], **J. Alonso**[7], **E. J. Bromet**[8], **B. Bunting**[9], **S. E. Florescu**[10], **A. Fukao**[11], **O. Gureje**[12], **C. Hu**[13], **Y. Q. Huang**[14], **A. N. Karam**[15], **D. Levinson**[16], **M. E. Medina Mora**[17], **J. Posada-Villa**[18], **K. M. Scott**[19], **N. I. Taib**[20], **M. C. Viana**[21], **M. Xavier**[22], **Z. Zarkov**[23], and **R. C. Kessler**[4]

[1]Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands [2]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA [3]Department of Psychiatry, MGH Clinical Trials Network & Institute, Depression Clinical & Research Program, Massachusetts General Hospital, Boston, Massachusetts, USA [4]Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts, USA [5]Depression Clinical and Research Program and the Bipolar Clinic and Research Program, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA [6]Johnson and Johnson Pharmaceutical Research and Development, Titusville, New Jersey, USA [7]IMIM-Hospital del Mar Research Institute, Parc de Salut Mar; Pompeu Fabra University (UPF); and CIBER en Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain [8]Department of Psychiatry and Behavioral Science, Stony Brook School of Medicine, State University of New York at Stony Brook, Stony Brook, New York, USA [9]Psychology Research Institute, University of Ulster, Londonderry, UK [10]National School of Public Health, Management and Professional Development, Bucharest, Romania [11]Department of Public Health, Yamagata University School of Medicine, Yamagata, Japan [12]University College Hospital, Ibadan, Nigeria [13]Shenzhen Institute of Mental Health and Shenzhen Kangning Hospital, Guangdong Province, People's Republic of China [14]Institute of Mental Health, Peking University, Beijing, People's Republic of China [15]Department of Psychiatry and Clinical Psychology, St George Hospital University Medical Center, Department of Psychiatry and Clinical Psychology, Faculty of Medicine, Balamand University Medical School, and Institute for Development Research Advocacy and Applied Care (IDRAAC), Beirut, Lebanon [16]Research and Planning, Mental Health Services, Ministry of Health, Jerusalem, Israel [17]National Institute of Psychiatry, Calzada Mexico Xochimilco, Mexico City, Mexico [18]Universidad Colegio Mayor de Cundinamarca, Bogota, Colombia [19]Department of Psychological Medicine, University of Otago, Dunedin, New Zealand [20]Mental Health Center-Duhok, Kurdistan Region, Iraq [21]Department of Social Medicine, Federal University of Espirito Santo, Vitoria, Brazil [22]Department of Mental Health, Universidade Nova de Lisboa, Lisbon, Portugal [23]National Center of Public Health and Analyses Department Mental Health, Sofia, Bulgaria

Abstract

[*]Address for correspondence: Ronald C. Kessler, Ph.D., Department of Health Care Policy, Harvard Medical School, Boston, MA, USA. Tel. (617) 432-3587 Fax (617) 432-3588 kessler@hcp.med.harvard.edu.

Background—Although variation in long-term course of major depressive disorder (MDD) is

not strongly predicted by existing symptom subtype distinctions, recent research suggests that prediction can be improved by using machine learning methods. However, it is not known whether these distinctions can be refined by added information about comorbid conditions. The current report presents results on this question.

Methods—Data come from 8,261 respondents with lifetime DSM-IV MDD in the WHO World Mental Health (WMH) Surveys. Outcomes include four retrospectively-reported measures of persistence-severity of course (years in episode; years in chronic episodes, hospitalization for MDD; disability due to MDD). Machine learning methods (regression tree analysis; lasso, ridge, and elastic net penalized regression) followed by k-means cluster analysis were used to augment previously-detected subtypes with information about prior comorbidity to predict these outcomes.

Results—Predicted values were strongly correlated across outcomes. Cluster analysis of predicted values found 3 clusters with consistently high, intermediate, or low values. The high-risk cluster (32.4% of cases) accounted for 56.6–72.9% of high persistence, high chronicity, hospitalization, and disability. This high-risk cluster had both higher sensitivity and likelihood-ratio positive (relative proportions of cases in the high-risk cluster versus other clusters having the adverse outcomes) than in a parallel analysis that excluded measures of comorbidity as predictors.

Conclusions—Although results using the retrospective data reported here suggest that useful MDD subtyping distinctions can be made with machine learning and clustering across multiple indicators of illness persistence-severity, replication is need with prospective data to confirm this preliminary conclusion.

Patients with major depressive disorder (MDD) vary substantially in illness course and treatment response. Recognition of this variation has led researchers to search for depression subtypes defined by distinctions assessed at the beginning of treatment, such as supposed causes (e.g., postnatal depression) (Cooper *et al*., 2007; Cooper & Murray, 1995), clinical presentation (e.g., atypical or melancholic depression) (Fink *et al*., 2007; Uher *et al*., 2011), and empirically-derived (e.g., factor analysis, latent class analysis) symptom profiles (Lamers *et al*., 2012; Vrieze *et al*., 2014), in hopes that these subtypes would tap into underlying psychopathological processes that predict treatment response or course of illness (Baumeister & Parker, 2012; Carragher *et al*., 2009).

While some promising results have emerged regarding significant associations of baseline biomarkers (e.g., Pizzagalli, 2011) and psychosocial variables (e.g., Candrian *et al*., 2007) with depression treatment response, subtyping distinctions based on empirically-derived symptom profiles have been disappointing due to profile instability (Baumeister & Parker, 2012; Hasler & Northoff, 2011; van Loo *et al*., 2012). However, an alternative approach to symptom-based subtyping given the desire to predict treatment response and course of illness would be to define subtypes using recursive partitioning (Strobl *et al*., 2009; Zhang & Singer, 2010) and related machine learning methods (James *et al*., 2013; van der Laan &

Rose, 2011) that search for synergistic associations of baseline measures with subsequent outcomes.

The latter methods have been useful in discovering stable synergistic predictors of clinical outcomes in others areas of medicine (Chang *et al*., 2012; Chao *et al*., 2012). Other than small studies of depression treatment response (Andreescu *et al*., 2008; Jain *et al*., 2013; Nelson *et al*., 2012; Rabinoff *et al*., 2011; Riedel *et al*., 2011), though, we are aware of only one previous study using machine learning to search for depression subtypes in predicting course of illness. That study, by van Loo and colleagues (van Loo *et al*., 2014), analyzed retrospectively reported data on associations of DSM-IV MDD symptoms in incident episodes with four measures of long-term illness persistence-severity in a sample of 8,261 respondents with lifetime MDD in the WHO World Mental Health (WMH) surveys. Significant subtyping distinctions were found based on the conjunction of child-adolescent onset, suicidality, and symptoms of anxiety occurring during incident depressive episodes. Respondents in the high-risk cluster (fewer than one-third of respondents) accounted for 53–71% of high persistence-severity.

The predictors in the van Loo analysis were limited, though, to variables characterizing incident episode symptoms. A question can be raised whether an expanded set of predictors might improve subtyping accuracy. In particular, information about prior lifetime comorbidities might be especially valuable given that van Loo found symptoms of anxiety to be powerful predictors of illness course and that evidence exists in the larger literature that comorbidity is related to the course of MDD (Steinert *et al*., 2014). The current report presents an expanded WMH analysis evaluating whether information about temporally primary comorbid disorders improves on the van Loo results.

METHODS

Sample

The WMH surveys are well-characterized epidemiological surveys of prevalence and correlates of commonly-occurring mental disorders (Alonso *et al*., 2013; Kessler & Üstün, 2008; Nock *et al*., 2012; Von Korff *et al*., 2009) administered in six countries classified by the World Bank as high income (Israel, Japan, New Zealand, Northern Ireland, Portugal, United States), five upper-middle income (Brazil, Bulgaria, Lebanon, Mexico, Romania), and five low/lower- middle income (Colombia, Iraq, Nigeria, Peoples Republic of China, Ukraine) (World Bank, 2009). Most surveys featured nationally representative household samples, while two (Colombia, Mexico) were representative of all urban areas in the country, one of selected states (Nigeria), and three of selected Metropolitan Areas (Brazil, Japan, Peoples Republic of China). A total of 93,167 adults (age 18+) participated. Sample sizes ranged from 2,357 (Romania) to 12,790 (New Zealand). Informed consent was obtained using procedures approved by local Institutional Review Boards. The average weighted response rate was 73.7% (55.1–96.2% range). Weights were used to adjust for differential probabilities of selection and discrepancies with population socio-demographic/ geographic distributions. Further details about WMH sampling and weighting are available elsewhere (Heeringa *et al*., 2008). The subsample considered here includes 8,261 WMH respondents who met lifetime DSM-IV criteria for MDD. (More detailed information on the

descriptive characteristics of the sample is presented by van Loo *et al.* (van Loo *et al.*, 2014).

Measures

MDD—DSM-IV MDD was assessed with the Composite International Diagnostic Interview (CIDI), Version 3.0 (Kessler & Üstün, 2004), a fully-structured diagnostic interview administered by trained lay interviewers. The CIDI translation, back-translation, and harmonization protocol required culturally competent bilingual clinicians in participating countries to review, modify, and approve key phrases describing symptoms (Harkness *et al.*, 2008). Standardized procedures for interviewer training and quality control were employed (Pennell *et al.*, 2008). The *likelihood-ratio positive* (LR+; the relative proportions of clinical cases among respondents screened positive versus others) was 8.8, which is close to the 10.0 level typically considered definitive for ruling in clinical diagnoses from fully-structured approximations (Altman *et al.*, 2000).

Respondents with lifetime DSM-IV/CIDI MDD were asked retrospective questions about age-of-onset (AOO), whether their first lifetime depressive episode "was brought on by some stressful experience" or happened "out of the blue," all DSM-IV Criterion A-D MDE symptoms for the index episode (including separate questions about irritability, weight loss and weight gain, insomnia and hypersomnia, psychomotor agitation and retardation, and about thoughts of death, suicide ideation, suicide plans, and suicide gestures-attempts), ICD-10 severity specifiers, questions to operationalize diagnostic hierarchy rule exclusions, and questions about marker symptoms of (i) a mixed episode (sleep much less than usual and still not feel tired; racing thoughts) and (ii) anxious depression (feeling nervous-anxious-worried; having sudden attacks of intense fear or panic). We also included in the initial subtyping analysis a dichotomous measure of whether either of the respondent's parents had a history of major depression based on respondent reports in the Family History Research Diagnostic Criteria Interview (Endicott *et al.*, 1978).

Four retrospective questions were asked about subsequent lifetime MDD persistence-severity: number of years since AOO when the respondent had an episode (i) lasting two weeks or longer (referred to below as *persistence*) or (ii) lasting *most days* throughout the year (referred to below as *chronicity*); (iii) whether the respondent was ever hospitalized for depression and, if so, the age of first occurrence (referred to below as *hospitalization*); and (iv) whether the respondent was currently (at the time of interview) sufficiently disabled because of his or her depression that he/she was either unable to work or had a limitation of at least 50% in the ability to perform paid work (referred to below as *disability*). These are the four outcomes considered here. The persistence and chronicity measures were divided by number of years between age-at-interview (AAI) and AOO to create continuous outcomes in the range 0–100%, while hospitalization and disability were treated as dichotomies.

Prior history of other DSM-IV/CIDI disorders—The CIDI assessed 14 other lifetime DSM-IV disorders, including three other distress disorders (generalized anxiety disorder, post-traumatic stress disorder, obsessive-compulsive disorder), five fear disorders (separation anxiety disorder, specific phobia, social phobia, panic disorder, agoraphobia),

and six externalizing disorders (attention-deficit/hyperactivity disorder, intermittent explosive disorder, oppositional-defiant disorder, conduct disorder, substance [alcohol or drug] abuse with or without dependence, and substance dependence with abuse). Age-of-onset (AOO) of each disorder was assessed using special probing techniques shown experimentally to improve recall accuracy (Knäuper *et al.*, 1999).

DSM-IV organic exclusion rules and diagnostic hierarchy rules among the disorders assessed were used in making diagnoses other than for oppositional-defiant disorder, which was defined with or without conduct disorder, and substance abuse, which was defined with or without dependence. As detailed elsewhere (Haro *et al.*, 2006), generally good concordance was found between these diagnoses based on the CIDI and blinded clinical diagnoses based on reappraisal interviews with the SCID (First *et al.*, 2002).

We considered not only the 14 disorders themselves but also aggregated combinations and disaggregated subsets in predicting the course of MDD. The aggregated combinations included nested dichotomous measures of numbers of distress disorders (1+ and 2+), fear disorders (1+, 2+, and 3+), and externalizing disorders (1+, 2+, 3+). The disaggregated subsets included child-adolescent onset versus adult onset cases, where child-adolescent onset was defined as AOO less than or equal to 18 years of age. Given differences in the age-of-onset distributions of mental disorders (Kessler *et al.*, 2007), the proportion of cases with child-adolescent versus adult onsets varies across disorders.

Analysis methods

We followed van Loo and colleagues (van Loo *et al.*, 2014) in predicting persistence and chronicity in the subsamples of respondents where AAI-AOO was either 10+ years (chronicity) or 15+ years (persistence) based on preliminary inspection carried out by van Loo showing that outcome scores stabilized after these cut-points. The chronicity and persistence models both used a Poisson link function. Proportional hazards survival analysis was used, in comparison, to predict first hospitalization among respondents who were not hospitalized at AOO. Logistic regression analysis, finally, was used to predict current disability in the total sample. The sample size used to predict the four outcomes varied widely because of these differences in sample. Only the 2,869 respondents with first onset of depression 15+ years prior to interview were included in the persistence analysis. Only the 3,958 respondents who had a first onset of depression 10+ years prior to interview were included in the chronicity analysis. Only the 6,465 respondents who were not hospitalized in conjunction with their incident episode of depression were included in the analysis of subsequent first hospitalization. All 8,261 respondents, in comparison, were included in the analysis of current disability.

The analyses were carried out in two steps designed to generate optimal prediction equations for the outcomes using a number of different machine learning search methods. In the first step, regression trees were estimated to find important interactions among the temporally primary comorbid disorders and between these disorders and the symptoms of MDD in incident episodes. The WMH weights were used in generating these trees. In order to minimize risk of over-fitting, 100 trees were estimated, each in a separate bootstrapped pseudo-sample (Hastie *et al.*, 2009; Strobl *et al.*, 2009; Zhang & Singer, 2010). The R-

package *rpart* (Thernau *et al*., 2012) was used for this purpose. Inspection of the frequencies with which unique terminal interactions (i.e., sub-samples defined by the conjunction of the dichotomous predictors selected to optimize prediction of the outcome) occurred across the 100 trees was inspected and further analysis was limited to the subset of interactions that appeared in at least 10% of trees.

In the second step, a separate dummy variable for each temporally primary comorbid disorder, a dummy predictor for each of the terminal interactions found in 10% or more of trees, and an offset term for the predicted values of the outcome from the models estimated in the earlier van Loo analysis (van Loo *et al*., 2014) were used to predict each outcome in multiple regression analyses. The central difficulty in estimating second-step models was that the predictors were highly inter-correlated, leading to coefficient instability when all predictors were included in a single equation. The classic way to address this problem is with stepwise regression (Draper & Smith, 1981), but this method over-fits and performs poorly when used to predict in new samples (Berk, 2003). A number of machine learning methods have been developed to improve on stepwise regression. We used one of these methods, *penalized regression*. This method trades off a certain amount of bias to reduce overfitting by shrinking coefficients (either constraining the sum of absolute values and/or the variances of nonzero standardized regression coefficients in the equations) by including fixed values of the penalties as constants and estimating model parameters under the constraints of these penalties (Berk, 2008).

Three different penalized regression models were used in our analysis. These three differ in the mixing parameter penalties (MPPs) used in estimation. One of these three was the lasso penalty (MPP=1), which favors a sparse model that forces coefficients for all but one predictor in each strongly correlated set to zero. A second was the ridge penalty (MPP=0), which uses proportional coefficient shrinkage to retain all predictors in the model. The third was the elastic net penalty, which uses simulation to vary MPP in the range 0.0–1.0 to select a penalty value with the best cross-validated fit (Zou & Hastie, 2005). The R-package *glmnet* (Friedman *et al*., 2010) was used to estimate all the penalized regression models. Glmnet implemented external 10-fold cross-validation to select the MPP yielding best overall model fit. Internal cross-validation was then used to select the coefficient in front of the penalty. This means that rather than choosing the predictors to retain by setting an entry or exit p-value as in stepwise regression, simulation is used to select coefficients to maximize overall model fit under the penalty in cross validation. Coefficients were exponentiated to create incidence density ratios (IDRs) for predictors of persistence and chronicity, hazard ratios (HRs) for predictors of hospitalization, and odds-ratios (ORs) for predictors of disability.

The coefficients in the three penalized regression models were then used to generate predicted values for each outcome for each respondent. Importantly, these predicted values were generated for all 8,261 respondents even though three of the four equations were estimated using subsamples based on the idea that we wanted to predict the likelihood that each respondent would eventually have each outcome regardless of how long ago the respondent had a first depressive episode.

Inspection of a correlation matrix among the predicted values in the total sample documented high correlations, suggesting that the predicted values could be combined to develop a summary measure of risk across all outcomes. This was done by transforming the predicted values from each equation to percentiles and using these transformed scores as input to k-means cluster analysis. This clustering partitioned the sample into subtypes with similar multivariate profiles of predicted scores across outcomes. The R-package *stats* (R Core Team, 2013) was used for this purpose, with 100 random starts generated for each number of clusters to avoid local minimization problems.

Inspection of observed (as opposed to predicted) mean dichotomized outcome scores across clusters was used to determine the optimal number of clusters to retain in predicting the outcomes based on *area under the receiver operating characteristic curve* (AUC; the proportion of times a randomly selected respondent with the outcome and a randomly selected respondent without the outcome could be differentiated correctly by cluster membership). Once this optimal number of clusters was determined, operating characteristics of a dichotomous screening scale that distinguished respondents in the cluster with the highest risk of the outcomes from other respondents were calculated for each outcome. Included here were measures of *sensitivity* (SENS; the percent of all respondents with the adverse outcome who were in the high-risk cluster), *positive predictive value* (PPV; risk of the adverse outcome among respondents in the high-risk cluster), and LR+ (relative proportions of cases in the high-risk cluster versus other clusters having the adverse outcomes).

All analyses used the WMH weights to adjust for differential probabilities of selection in generating samples. All prediction equations additionally included dummy predictor variables for country to adjust for between-country differences in outcomes. The effects of weights but not geographic clustering were taken into consideration in cross-validations. Standard errors of operating characteristics were estimated using the design-based Taylor series linearization method (Wolter, 1985), which accounted for the effects of both weights and clustering, using R-package *survey* (Lumley, 2004).

RESULTS

Machine learning models

The only terminal interactions emerging repeatedly in regression trees involved number of fear disorders without regard to AOO. Nested dichotomies for number of fear disorders (1+, 2+, 3+) were consequently included as dummy predictor variables in the penalized regression analyses. The best-fitting penalized regression model for each outcome was an elastic net with MPP=0.1. This means the coefficients are especially hazardous to interpret because many highly correlated predictors remain in the model with proportional coefficient shrinkage to maximize overall model fit at the expense of interpretability of individual coefficients. Nonetheless, as individual-level predicted values are very similar in the sparser lasso model compared to the optimal elastic net model across outcomes (*r*=.78–.98), it is possible to inspect lasso coefficients to get some sense of predictor importance.

All but one of the 22 predictors retained under the lasso penalty across outcomes had positive coefficients, indicating increased persistence-severity associated with prior disorders. (Table 1) The proportion of retained coefficients was highest for the fear disorder interactions (42%; 5/12), next highest for the distress disorders (33%; 8/24), next highest for individual fear disorders (15%; 6/40), and lowest for externalizing disorders (6%; 2/48 positive associations, 1/48 negative associations). The 5 retained predictors of hospitalization had HRs in the range 1.09–1.34. The 5 positive retained predictors of disability had ORs in the range 1.09–1.84. The one negative predictor had an OR of 0.83. The 7 retained predictors of persistence had IDRs in the range 1.03–1.14. The 3 retained predictors of chronicity had IDRs in the range 1.10–1.16. The relatively modest size of these coefficients reflects strong inter-correlations among retained predictors.

Cluster analysis

Predicted values of each outcome were calculated for each respondent in the total sample based on the ridge, lasso, and optimal elastic net model coefficients. Spearman correlations among these predicted values were in the range .76–.98. Principal axis exploratory factor analysis across outcomes showed that the correlations were consistent with the existence of a single underlying factor (factor loadings in the range .77–.96). (Detailed results are available on request). Based on these results, k-means cluster analysis of transformed (to percentiles) predicted outcome scores was used to search for empirically-derived multivariate clusters defining subtypes with differential outcome risk.

As in the earlier van Loo analysis (van Loo *et al.*, 2014), inspection of mean percentile scores for solutions in the range between three and eight clusters showed that all solutions defined one class that had the highest mean scores on all outcomes, a second class that had the lowest mean scores on all outcomes, and between one and six other classes that had intermediate mean scores on all outcomes. (Figure 1a–1f) Based on this observation, alternative three-cluster solutions were constructed from the original four-cluster through eight-cluster solutions by collapsing the intermediate clusters in each solution. AUC was then compared for the original three-cluster solution and the alternative collapsed three-cluster solutions to predict dichotomized versions of the outcomes (top 10 percentiles of persistence and chronicity along with yes-no measures of hospitalization and disability). None of the collapsed higher-order cluster solutions had a higher AUC than the original three-cluster solution on any outcomes (.68 for persistence, 62 for chronicity, .70 for hospitalization, and .73 for disability). The distribution of cluster membership in the three-cluster solution was 32.4% in the high-risk cluster, 35.6% in the intermediate-risk cluster, and 32.0% in the low-risk cluster.

Associations of cluster membership with course of illness

Concentration of risk of adverse outcomes in the high-risk cluster was examined in the sub-sample of respondents with AAI-AOO of 15+ years because, as noted in the section on analysis methods, this was the most restrictive subsample used in the analyses. Cross-tabulations with dichotomized versions of outcomes show that the proportions of all cases with positive scores on the dichotomized outcomes occurring in the high-risk cluster (i.e., estimates of SENS) were 60.6% for persistence, 56.6% for chronicity, 61.8% for

hospitalization, and 72.9% for disability. (Table 2) The high-risk cluster also captured 55.8% of all cases with any of the four adverse outcomes and 70.5% of those with two or more of those outcomes. Between 4.6% (disability) and 18.8% (high persistence) of cases in the high-risk cluster experienced each of the four adverse outcomes (PPV), while 40.7% experienced at least one and 13.8% more than one of these outcomes. These proportions were in the range 2.6–5.4 times as high as among cases not classified as being in the high-risk cluster (LR+).

Improvement in prediction due to considering comorbidity

Comparison of the operating characteristics of the high-risk cluster based on the current analysis with the high-risk cluster based on the earlier van Loo analysis (van Loo *et al.*, 2014) allowed assessment of the extent to which prediction improved when we added information about temporally primary comorbid disorders. (Detailed results are available on request.) Three observations are noteworthy. First, the percent of respondents classified in the high-risk cluster increased from 31.2% to 32.4% in the current analysis (a 3.8% increase on the base in the earlier analysis). Second, SENS increased more than the 3.8% proportional increase in size of the high-risk cluster, with a 7.7% increase for having any adverse outcome (i.e., 1–55.8/51.8 based on SENS being 55.8% in the current analysis vs. 51.8% in the earlier analysis) and an 8.7% increase for having two or more of these outcomes (i.e., 1–70.5/64.8 based on SENS being 70.5% in the current analysis vs. 64.8% in the earlier analysis). Third, LR+ *increased* as a result of the higher increases in SENS than relative prevalence, with LR+ of 2.6 for having any of the 4 outcomes vs. 2.4 in the earlier analysis and LR+ of 4.8 for having more than one of these outcomes vs. 4.1 in the earlier analysis.

DISCUSSION

Data limitations include use of retrospective reports based on fully-structured diagnostic interviews that included only a limited set of predictors. An especially important limitation regarding predictors is that personality disorders were not assessed, as personality disorders are known to predict depression treatment outcome (Candrian *et al.*, 2007). Nor did we consider other predictors of depression persistence-severity other than depressive symptoms and comorbidity (e.g., gender, other socio-demographics, stress exposure) because of our interest in focusing on symptoms-comorbidity in defining subtypes. Future extensions should include a wider range of predictors and outcomes based on clinical assessments and prospective designs. The machine learning methods used here, although designed to minimize model over-fitting, were not completely conservative because they failed to adjust for within-country geographic clustering. This means that predictor effects might not hold up as well in other samples as if data had been based on simple random samples. At the other extreme, the tree and penalized regression methods do not exhaust all available machine learning methods. As a result, other synergistic associations might be discovered if future studies used a wider range of integrated machine learning methods (van der Laan *et al.*, 2007).

In interpreting predictor coefficients it is important to appreciate that machine learning methods are designed to maximize overall model prediction at the expense of coefficient accuracy. The predictors selected under the Lasso penalty, in particular, are the ones that best represent the joint effects of the larger sets of predictors to which they are related. It is consequently important not to over-interpret the importance of the specific predictors retained in these models. Within the context of this and the other limitations noted above, we found that temporally primary fear and distress disorders are much more consistent predictors of long-term major depression persistence-severity than are externalizing disorders. And we found synergistic effects among fear disorders, but not distress disorders or between fear and distress disorders. The failure to find synergistic effects among distress disorders is striking given that major depression is a distress disorder (Clark & Watson, 2006).

We found that multivariate predictions are strongly correlated across outcomes and that distinct subsets of respondents had high or low predicted risk across all outcomes. This high-risk cluster included one-third of respondents and accounted for 56.6–72.9% of adverse outcomes. This clustering was stronger than in the van Loo analysis, documenting the importance of comorbidity in predicting long-term depression persistence-severity. Although the finding of strong correlations among predicted scores across outcomes is not surprising given that all outcomes assessed long-term disorder persistence-severity, the finding of distinct high-risk and low-risk clusters was not preordained by these high bivariate associations.

Although a number of previous epidemiological studies examined baseline predictors of long-term depression course in treatment samples (Judd *et al.*, 2013; Penninx *et al.*, 2008) or community samples (Bottomley *et al.*, 2010; Patten *et al.*, 2010; Rhebergen *et al.*, 2012), none searched for depression subtypes among predictors. As noted in the introduction, though, other subtyping analyses similar to those reported here have been carried out, including analyses to predict treatment response (Andreescu *et al.*, 2008; Jain *et al.*, 2013) and naturalistic patterns of remission among patients in treatment (Riedel *et al.*, 2011) as well as in the placebo control group of a depression clinical trial (Nelson *et al.*, 2012). In addition, a number of recent clinical studies used methods similar to ours to predict onset of suicidality during the course of depression treatment (Musil *et al.*, 2013; Rabinoff *et al.*, 2011; Seemuller *et al.*, 2009) or after termination of treatment (Ilgen *et al.*, 2009). Our results suggest that similar efforts in prospective samples might be able to document subtypes that predict long-term persistence-severity of depression.

In considering the extension of our analyses to prospective studies, it is important to recognize that we failed to find strong higher-order predictor profiles based on complex trees despite the sample being quite large. Taken together with the results of a recent systematic review that failed to find support for stable symptom-based MDD subtypes defined by internal consistency (e.g., factor analysis, latent class analysis) (van Loo *et al.*, 2012), this argues against the existence of strong MDD subtypes defined *exclusively* by synergistic associations among symptoms or comorbidities. However, broader MDD predictive subtypes might be found by focusing more closely within the high-risk cluster identified in our analysis. Importantly, this cluster is not a classical subtype because it was

discovered by examining multivariate predicted outcome scores rather than the predictors themselves. Many different predictor combinations could lead to the same predicted outcome scores. This means that more work is needed to identify subtypes *within* the high-risk cluster by considering multivariate profiles among the predictors that determine cluster membership. This would require other methods than used in the current report. In addition, it would be useful if future studies expanded the outcomes beyond the four considered here to gain more insight into the range over which prediction occurs. Such an investigation could be carried out informally using the simple correlational methods employed here or more formally using machine learning methods developed to discover genetic *master regulators* (Chan & Kyba, 2013; Rabinowitz & Silhavy, 2013; Rayner *et al*., 2013; Ryu *et al*., 2007).

Acknowledgments

REFERENCES

Alonso, J.; Chatterji, S.; He, Y. The Burden of Mental Disorders: Global Perspectives from the WHO World Mental Health Surveys. New York: Cambridge University Press; 2013.

Altman, D.; Machin, D.; Bryant, T.; Gardner, S., editors. Statistics with Confidence: Confidence intervals and statistical guidelines, Second Edition. London: BMJ Books; 2000.

Andreescu C, Mulsant BH, Houck PR, Whyte EM, Mazumdar S, Dombrovski AY, Pollock BG, Reynolds CF 3rd. Empirically derived decision trees for the treatment of late-life depression. American Journal of Psychiatry. 2008; 165:855–862. [PubMed: 18450930]

Baumeister H, Parker G. Meta-review of depressive subtyping models. Journal of Affective Disorders. 2012; 139:126–140. [PubMed: 21885128]

Berk, RA. Regression Analysis: A Constructive Critique. Newbury Park, CA: Sage; 2003.

Berk, RA. Statistical Learning from a Regression Perspective. New York, NY: Springer; 2008.

Bottomley C, Nazareth I, Torres-Gonzalez F, Svab I, Maaroos HI, Geerlings MI, Xavier M, Saldivia S, King M. Comparison of risk factors for the onset and maintenance of depression. British Journal of Psychiatry. 2010; 196:13–17. [PubMed: 20044653]

Candrian M, Farabaugh A, Pizzagalli DA, Baer L, Fava M. Perceived stress and cognitive vulnerability mediate the effects of personality disorder comorbidity on treatment outcome in major depressive disorder: a path analysis study. Journal of Nervous and Mental Disease. 2007; 195:729–737. [PubMed: 17984772]

Carragher N, Adamson G, Bunting B, McCann S. Subtypes of depression in a nationally representative sample. Journal of Affective Disorders. 2009; 113:88–99. [PubMed: 18644628]

Chan SS, Kyba M. What is a master regulator? Journal of Stem Cell Research and Therapy. 2013; 3:e114.

Chang YJ, Chen LJ, Chung KP, Lai MS. Risk groups defined by Recursive Partitioning Analysis of patients with colorectal adenocarcinoma treated with colorectal resection. BMC Medical Research Methodology. 2012; 12:2. [PubMed: 22214198]

Chao ST, Koyfman SA, Woody N, Angelov L, Soeder SL, Reddy CA, Rybicki LA, Djemil T, Suh JH. Recursive partitioning analysis index is predictive for overall survival in patients undergoing spine stereotactic body radiation therapy for spinal metastases. International Journal of Radiation Oncology • Biology • Physics. 2012; 82:1738–1743.

Clark LA, Watson D. Distress and fear disorders: an alternative empirically based taxonomy of the "mood" and "anxiety" disorders. British Journal of Psychiatry. 2006; 189:481–483. [PubMed: 17139030]

Cooper C, Jones L, Dunn E, Forty L, Haque S, Oyebode F, Craddock N, Jones I. Clinical presentation of postnatal and non-postnatal depressive episodes. Psychological Medicine. 2007; 37:1273–1280. [PubMed: 17349101]

Cooper PJ, Murray L. Course and recurrence of postnatal depression. Evidence for the specificity of the diagnostic concept. British Journal of Psychiatry. 1995; 166:191–195. [PubMed: 7728362]

Draper, N.; Smith, H. Applied Regression Analysis, Second Edition. New York, NY: Wiley; 1981.

Endicott, J.; Andreasen, N.; Spitzer, RL. Family History Research Diagnostic Criteria. New York, NY: Biometrics Research, New York State Psychiatric Institute; 1978.

Fink M, Rush AJ, Knapp R, Rasmussen K, Mueller M, Rummans TA, O'Connor K, Husain M, Biggs M, Bailine S, Kellner CH. DSM melancholic features are unreliable predictors of ECT response: a CORE publication. Journal of ECT. 2007; 23:139–146. [PubMed: 17804986]

First, MB.; Spitzer, RL.; Gibbon, M.; Williams, JBW. Structured Clinical Interview for DSM-IV Axis I Disorders, Research Version, Non-patient Edition (SCID-I/NP). New York, NY: Biometrics Research, New York State Psychiatric Institute; 2002.

Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software. 2010; 33:1–22. [PubMed: 20808728]

Harkness, J.; Pennell, BE.; Villar, A.; Gebler, N.; Aguilar-Gaxiola, S.; Bilgen, I. Translation procedures and translation assessment in the world mental health survey initiative. In: Kessler, RC.; Üstün, TB., editors. The WHO World Mental Health Surveys: Global Perspectives on the Epidemiology of Mental Disorders. New York: Cambridge University Press; 2008. p. 91-113.

Haro JM, Arbabzadeh-Bouchez S, Brugha TS, de Girolamo G, Guyer ME, Jin R, Lepine JP, Mazzi F, Reneses B, Vilagut G, Sampson NA, Kessler RC. Concordance of the Composite International Diagnostic Interview Version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO World Mental Health surveys. International Journal of Methods in Psychiatric Research. 2006; 15:167–180. [PubMed: 17266013]

Hasler G, Northoff G. Discovering imaging endophenotypes for major depression. Molecular Psychiatry. 2011; 16:604–619. [PubMed: 21602829]

Hastie, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. New York, NY: Springer; 2009.

Heeringa, SG.; Wells, EJ.; Hubbard, F.; Mneimneh, ZN.; Chiu, WT.; Sampson, NA.; Berglund, PA. Sample designs and sampling procedures. In: Kessler, RC.; Üstün, TB., editors. The WHO World Mental Health Surveys: Global Perspectives on the Epidemiology of Mental Disorders. New York, NY: Cambridge University Press; 2008. p. 14-32.

Ilgen MA, Downing K, Zivin K, Hoggatt KJ, Kim HM, Ganoczy D, Austin KL, McCarthy JF, Patel JM, Valenstein M. Exploratory data mining analysis identifying subgroups of patients with depression who are at high risk for suicide. Journal of Clinical Psychiatry. 2009; 70:1495–1500. [PubMed: 20031094]

Jain FA, Hunter AM, Brooks JO 3rd, Leuchter AF. Predictive socioeconomic and clinical profiles of antidepressant response and remission. Depression and Anxiety. 2013; 30:624–630. [PubMed: 23288666]

James, G.; Witten, D.; Hastie, T.; Tibshirani, R. An Introduction to Statistical Learning: With Applications in R. New York, NY: Springer; 2013.

Judd LL, Schettler PJ, Coryell W, Akiskal HS, Fiedorowicz JG. Overt irritability/anger in unipolar major depressive episodes: past and current characteristics and implications for long-term course. JAMA Psychiatry. 2013; 70:1171–1180. [PubMed: 24026579]

Kessler RC, Amminger GP, Aguilar-Gaxiola S, Alonso J, Lee S, Ustun TB. Age of onset of mental disorders: a review of recent literature. Current Opinion in Psychiatry. 2007; 20:359–364. [PubMed: 17551351]

Kessler RC, Üstün TB. The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). International Journal of Methods in Psychiatric Research. 2004; 13:93–121. [PubMed: 15297906]

Kessler, RC.; ÜstüTB, n, editors. The WHO World Mental Health Surveys: Global Perspectives on the Epidemiology of Mental Disorders. New York, NY: Cambridge University Press; 2008.

Knäuper B, Cannell CF, Schwarz N, Bruce ML, Kessler RC. Improving accuracy of major depression age-of-onset reports in the US National Comorbidity Survey. International Journal of Methods in Psychiatric Research. 1999; 8:39–48.

Lamers F, Burstein M, He JP, Avenevoli S, Angst J, Merikangas KR. Structure of major depressive disorder in adolescents and adults in the US general population. British Journal of Psychiatry. 2012; 201:143–150. [PubMed: 22700082]

Lumley T. Analysis of complex survey samples. Journal of Statistical Software. 2004; 9:1–19.

Musil R, Zill P, Seemuller F, Bondy B, Meyer S, Spellmann I, Bender W, Adli M, Heuser I, Fisher R, Gaebel W, Maier W, Rietschel M, Rujescu D, Schennach R, Moller HJ, Riedel M. Genetics of

emergent suicidality during antidepressive treatment--data from a naturalistic study on a large sample of inpatients with a major depressive episode. European Neuropsychopharmacology. 2013; 23:663–674. [PubMed: 23063133]

Nelson JC, Zhang Q, Deberdt W, Marangell LB, Karamustafalioglu O, Lipkovich IA. Predictors of remission with placebo using an integrated study database from patients with major depressive disorder. Current Medical Research and Opinion. 2012; 28:325–334. [PubMed: 22292447]

Nock, MK.; Borges, G.; Ono, Y. Suicide: Global Perspectives from the WHO World Mental Health Surveys. New York, NY: Cambridge University Press; 2012.

Patten SB, Wang JL, Williams JV, Lavorato DH, Khaled SM, Bulloch AG. Predictors of the longitudinal course of major depression in a Canadian population sample. Canadian Journal of Psychiatry. 2010; 55:669–676.

Pennell, B-E.; Mneimneh, Z.; Bowers, A.; Chardoul, S.; Wells, JE.; Viana, MC.; Dinkelmann, K.; Gebler, N.; Florescu, S.; He, Y.; Huang, Y.; Tomov, T.; Vilagut, G. Implementation of the world mental health surveys. In: Kessler, RC.; Üstün, TB., editors. The WHO World Mental Health Surveys: Global Perspectives on the Epidemiology of Mental Disorders. New York, NY: Cambridge University Press; 2008. p. 33-57.

Penninx BW, Beekman AT, Smit JH, Zitman FG, Nolen WA, Spinhoven P, Cuijpers P, De Jong PJ, Van Marwijk HW, Assendelft WJ, Van Der Meer K, Verhaak P, Wensing M, De Graaf R, Hoogendijk WJ, Ormel J, Van Dyck R. The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. International Journal of Methods in Psychiatric Research. 2008; 17:121–140. [PubMed: 18763692]

Pizzagalli DA. Frontocingulate dysfunction in depression: toward biomarkers of treatment response. Neuropsychopharmacology. 2011; 36:183–206. [PubMed: 20861828]

R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.

Rabinoff M, Kitchen CM, Cook IA, Leuchter AF. Evaluation of quantitative EEG by classification and regression trees to characterize responders to antidepressant and placebo treatment. Open Medical Informatics Journal. 2011; 5:1–8. [PubMed: 21603560]

Rabinowitz JD, Silhavy TJ. Systems biology: metabolite turns master regulator. Nature. 2013; 500:283–284. [PubMed: 23925111]

Rayner BS, Figtree GA, Sabaretnam T, Shang P, Mazhar J, Weaver JC, Lay WN, Witting PK, Hunyor SN, Grieve SM, Khachigian LM, Bhindi R. Selective inhibition of the master regulator transcription factor egr-1 with catalytic oligonucleotides reduces myocardial injury and improves left ventricular systolic function in a preclinical model of myocardial infarction. Journal of the American Heart Association. 2013; 2:e000023. [PubMed: 23902638]

Rhebergen D, Lamers F, Spijker J, de Graaf R, Beekman AT, Penninx BW. Course trajectories of unipolar depressive disorders identified by latent class growth analysis. Psychological Medicine. 2012; 42:1383–1396. [PubMed: 22053816]

Riedel M, Moller HJ, Obermeier M, Adli M, Bauer M, Kronmuller K, Brieger P, Laux G, Bender W, Heuser I, Zeiler J, Gaebel W, Schennach-Wolff R, Henkel V, Seemuller F. Clinical predictors of response and remission in inpatients with depressive syndromes. Journal of Affective Disorders. 2011; 133:137–149. [PubMed: 21555156]

Ryu B, Kim DS, Deluca AM, Alani RM. Comprehensive expression profiling of tumor cell lines identifies molecular signatures of melanoma progression. PloS One. 2007; 2:e594. [PubMed: 17611626]

Seemuller F, Riedel M, Obermeier M, Bauer M, Adli M, Mundt C, Holsboer F, Brieger P, Laux G, Bender W, Heuser I, Zeiler J, Gaebel W, Jager M, Henkel V, Moller HJ. The controversial link between antidepressants and suicidality risks in adults: data from a naturalistic study on a large sample of in-patients with a major depressive episode. International Journal of Neuropsychopharmacology. 2009; 12:181–189. [PubMed: 18662490]

Steinert C, Hofmann M, Kruse J, Leichsenring F. The prospective long-term course of adult depression in general practice and the community. A systematic literature review. Journal of Affective Disorders. 2014; 152–154:65–75.

Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychological Methods. 2009; 14:323–348. [PubMed: 19968396]

Thernau T, Atkinson B, Ripley B. Rpart: Recursive Partioning. R Package 4.1-0. 2012

Uher R, Dernovsek MZ, Mors O, Hauser J, Souery D, Zobel A, Maier W, Henigsberg N, Kalember P, Rietschel M, Placentino A, Mendlewicz J, Aitchison KJ, McGuffin P, Farmer A. Melancholic, atypical and anxious depression subtypes and outcome of treatment with escitalopram and nortriptyline. Journal of Affective Disorders. 2011; 132:112–120. [PubMed: 21411156]

van der Laan MJ, Polley EC, Hubbard AE. Super learner. Statistical Applications in Genetics and Molecular Biology. 2007; 6 Article 25.

van der Laan, MJ.; Rose, S. Targeted Learning: Causal Inference for Observational and Experimental Data. New York, NY: Springer; 2011.

van Loo HM, Cai T, Gruber MJ, Li J, de Jonge P, Petukhova M, Rose S, Sampson NA, Schoevers RA, Wardenaar KJ, Wilcox MA, Al-Hamzawi A, Andrade LH, Bromet EJ, Bunting B, Fayyad J, Florescu SE, Gureje O, Hu C, Huang Y, Levinson D, Medina-Mora ME, Nakane Y, Posada-Villa J, Scott KM, Xavier M, Zarkov Z, Kessler RC. Major depressive disorder subtypes to predict long-term course. Depression and Anxiety. 2014 e-publication ahead of print.

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

Von Korff, MR.; Scott, KM.; Gureje, O. Global Perspectives on Mental-Physical Comorbidity In the WHO World Mental Health Surveys. New York, NY: Cambridge University Press; 2009.

Vrieze E, Demyttenaere K, Bruffaerts R, Hermans D, Pizzagalli DA, Sienaert P, Hompes T, de Boer P, Schmidt M, Claes S. Dimensions in major depressive disorder and their relevance for treatment outcome. Journal of Affective Disorders. 2014; 155:35–41. [PubMed: 24210628]

Wolter, KM. Introduction to Variance Estimation. New York: Springer-Verlag; 1985.

World Bank. Data & Statistics, Country Groups by Income. 2009

Zhang, H.; Singer, BH. Recursive Partitioning and Applications, Second Edition. New York, NY: Springer; 2010.

Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society, Series B. 2005; 67:301–320.
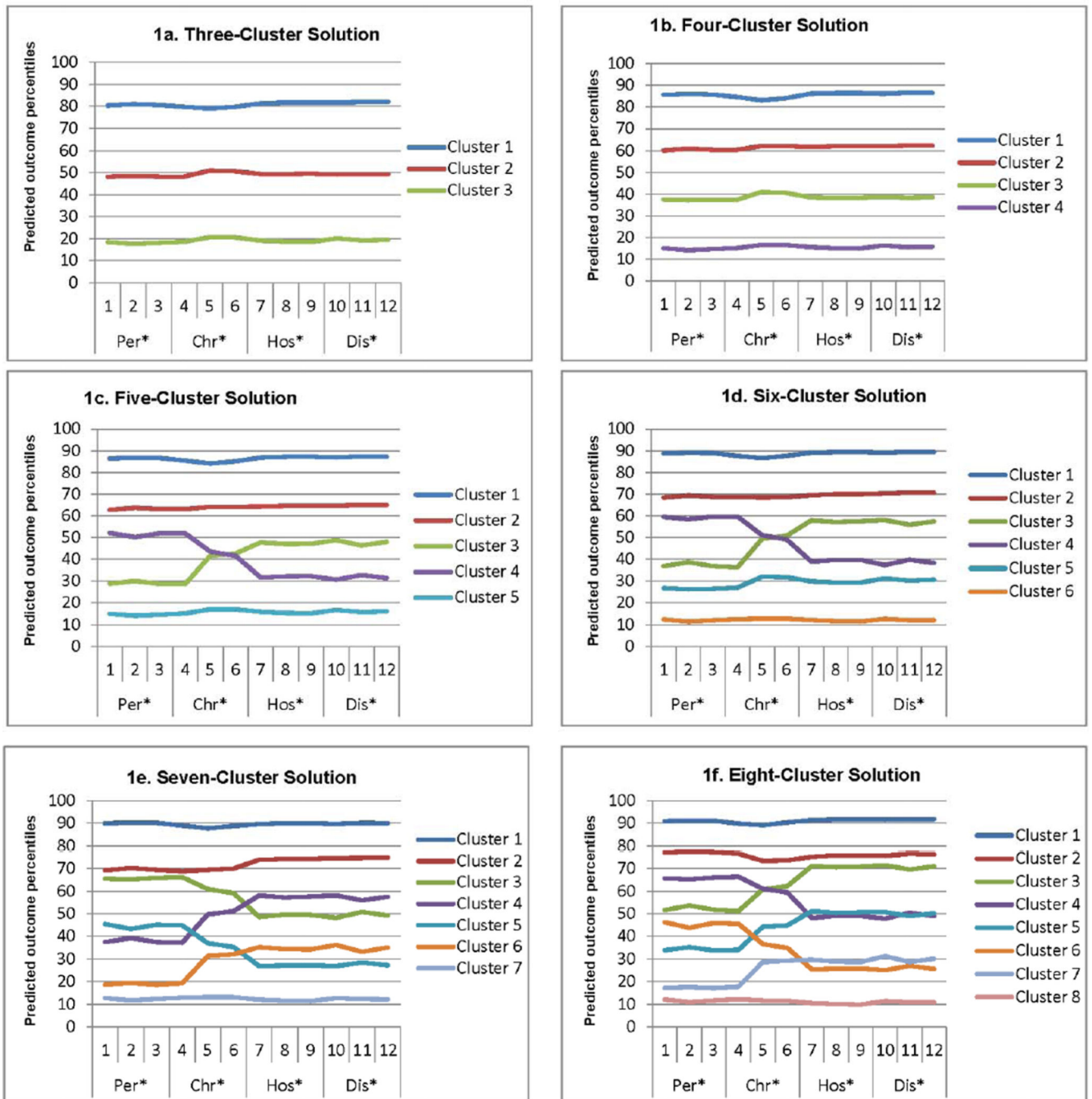
Figure 1. Mean predicted outcome scores in the three-cluster through eight-cluster k-means[a]
*Per = the percentile-transformed predicted score on the persistence outcome variable in the lasso(1), ridge (2). and elastic net (3) models: Chr = the percentile-transformed predicted score on the chronicity outcome in the lasso (4), ridge (5). and elastic net (6) models; Hos = the percentile-transformed cumulative predicted probability of hospitalization in the lasso (7), ridge (&). and elastic net (9) models; Dis = the percentile-transformed predicted probability of disability in the lasso (10), ridge (11), and elastic net (12) models.

[a]k-means cluster analysis of percentile-transformed predicted scores on the four outcomes for all respondents based on the Lasso GLM

Table 1

Lasso penalized regression coefficients to predict subsequent adverse MDD outcomes based on incident episode characteristics[a] and prior comorbidity prior to the first depressive episode[b]

	Persistence IDR	Chronicity IDR	Hospitalization HR	Disability OR
I. Comorbid fear disorders				
1 fear disorder	1.05		1.09	
2 fear disorders			1.14	1.41
3 fear disorders	1.04	1.00		
II. Fear disorders[c]				
Specific phobia				
Total				1.24
Social phobia				
AOO 18years	1.12	1.14		
Panic disorder				
Total			1.34	
AOO 18years	1.13			1.10
Agoraphobia				
Total				1.87
III. Distress disorders[d]				
Generalized anxiety disorder				
Total	1.06	1.10		
AOO 18years			1.16	1.54
Post-traumatic stress disorder				
Total	1.04			
AOO 18years	1.14	1.16		
Obsessive-compulsive disorder				
AOO 18years				1.09
IV. Externalizing disorders[e]				
Intermittent explosive disorder				
Total	1.08			
Oppositional-defiant disorder				
Total				0.83
Substance dependence				
Total			1.34	
(n)	(2,869)	(3,958)	(6,465)	(8,261)

HR=Hazard ratio; OR=Odds-ratio; IDR=Incidence density ratio

[a]Predicted values based on MDD episode characteristics were used as an offset in the models. The following variables were previously found to be important and used to compute the offset scores: Incident episode symptoms of severe dysphoria, anhedonia, weight loss, weight gain, insomnia, hypersomnia, psychomotor agitation, psychomotor retardation, suicidality, inability to cope, panic, irritability, racing thoughts, high energy, endogenous onset, parental history of depression; interactions involving AOO <19 and suicidality', AOO <19 and anxiety symptoms (panic, worry, or irritability), AOO <35 and suicidality and anxiety.

[b] Coefficients based on 10-fold cross-validation with a lasso penalty. The link functions were Poisson for persistence and chronicity', Cox proportional hazards for hospitalization, and logistic for disability. No Confidence intervals are reported because standard errors of such simulated models are biased. Retained coefficients were selected based on evidence in cross validation that the predictors improved overall model fit rather than that coefficients fell outside of a pre-specified confidence interval.

[c] The initial predictor set also included separation anxiety disorder (both total and with AOO 18years), specific phobia with AOO 18years, total social phobia, and agoraphobia with AOO 18years, but none of these predictors was retained in the lasso penalized regression model for any of the outcomes.

[d] The initial predictor set also included total obsessive-compulsive disorder , but that predictor was not retained in the lasso penalized regression model for any of the outcomes.

[e] The initial predictor set also included attention-deficit/hyperactivity disorder, intermittent explosive disorder with AOO 18years, oppositional-defiant disorder with AOO 18years, conduct disorder (both total and with AOO 18years), substance abuse (both total and with AOO 18years), and substance dependence with AOO 18years, but none of these predictors was retained in the lasso penalized regression model for any of the outcomes.

Table 2

Operating characteristics of the dichotomized distinction between cases in the high-risk cluster versus other cases in predicting course of MDD in the sub-sample of cases having all outcomes (n=2,435)

	Sensitivity[a]		Positive predictive value[b]		Likelihood-ratio positive[c]	
	%	(se)	%	(se)	Est.	(95% CI)
Persistence	60.6	3.3	18.8	1.4	3.1	(2.4 – 4.0)
Chronicity	56.6	3.1	17.1	1.3	2.6	(2.1 – 3.3)
Hospitalization	61.8	3.4	18.6	1.4	3.3	(2.5 – 4.3)
Disability	72.9	6.2	4.6	0.7	5.4	(2.9 – 10.1)
Any of the above	55.8	2.2	40.7	1.9	2.6	(2.2 – 3.0)
Two or more	70.5	3.8	13.8	1.2	4.8	(3.4 – 6.9)

[a] Sensitivity is the percent of observed adverse outcomes that occurred in the high-risk cluster.

[b] Positive predictive value is the percent of respondents in the high-risk cluster that experienced the adverse outcome.

[c] Likelihood-ratio positive is the relative proportions of cases in the high-risk cluster versus the remainder of the sample that experienced the adverse outcome.