Predicting depression onset in young people based on clinical, cognitive, environmental and neurobiological data

Yara J. Toenders, MSc[1,2], Akhil Kottaram, Ph.D.[1,2], Richard Dinga, Ph.D.[3], Christopher G. Davey, MBBS(Hons), Ph.D.[4], Tobias Banaschewski, M.D., Ph.D.[5], Arun L.W. Bokde, Ph.D.[6], Erin Burke Quinlan, PhD[7], Sylvane Desrivières, Ph.D.[7], Herta Flor, Ph.D.[8,9], Antoine Grigis, Ph.D.[10], Hugh Garavan, Ph.D.[11], Penny Gowland, Ph.D.[12], Andreas Heinz, M.D., Ph.D.[13], Rüdiger Brühl, Ph.D.[14], Jean-Luc Martinot, M.D., Ph.D.[15], Marie-Laure Paillère Martinot, M.D., Ph.D.[16], Frauke Nees, Ph.D.[5,8,17], Dimitri Papadopoulos Orfanos, Ph.D.[10], Herve Lemaitre, Ph.D[10,18], Tomáš Paus, M.D., Ph.D.[19], Luise Poustka, M.D.[20], Sarah Hohmann, M.D.[5], Juliane H. Fröhner, MSc[21], Michael N. Smolka, M.D.[21], Henrik Walter, M.D., Ph.D.[13], Robert Whelan, Ph.D.[22], Argyris Stringaris, M.D., Ph.D.[23], Betteke van Noort, MSc[24], Jani Penttilä, M.D. Ph.D.[25], Yvonne Grimmer, M.D.[5], Corinna Insensee, Ph.D.[20], Andreas Becker, Ph.D.[20], Gunter Schumann, M.D.[7,26], Lianne Schmaal, Ph.D.[1,2]

Affiliations

[1] Orygen, Parkville, Australia

[2] Centre for Youth Mental Health, The University of Melbourne, Parkville, Australia

[3] Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands

[4] Department of Psychiatry, The University of Melbourne, Australia

[5] Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Square J5, 68159 Mannheim, Germany;

[6] Discipline of Psychiatry, School of Medicine and Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland;

[7] Centre for Population Neuroscience and Precision Medicine (PONS), Institute of Psychiatry, Psychology & Neuroscience, SGDP Centre, King's College London, United Kingdom;

[8] Institute of Cognitive and Clinical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Square J5, Mannheim, Germany;

[9] Department of Psychology, School of Social Sciences, University of Mannheim, 68131 Mannheim, Germany;

[10] NeuroSpin, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France;

[11] Departments of Psychiatry and Psychology, University of Vermont, 05405 Burlington, Vermont, USA;

[12] Sir Peter Mansfield Imaging Centre School of Physics and Astronomy, University of Nottingham, University Park, Nottingham, United Kingdom;

[13] Department of Psychiatry and Psychotherapy CCM, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany;

[14] Physikalisch-Technische Bundesanstalt (PTB), Braunschweig and Berlin, Germany;

[15] Institut National de la Santé et de la Recherche Médicale, INSERM U A10 "Trajectoires développementales en psychiatrie"; Université Paris-Saclay, Ecole Normale supérieure Paris-Saclay, CNRS, Centre Borelli; Gif-sur-Yvette, France;

[16] Institut National de la Santé et de la Recherche Médicale, INSERM U A10 "Trajectoires développementales en psychiatrie"; Université Paris-Saclay, Ecole Normale supérieure Paris-Saclay, CNRS, Centre Borelli; and AP-HP.Sorbonne Université, Department of Child and Adolescent Psychiatry, Pitié-Salpêtrière Hospital, Paris, France;

[17] Institute of Medical Psychology and Medical Sociology, University Medical Center Schleswig Holstein, Kiel University, Kiel, Germany.

[18] Institut des Maladies Neurodégénératives, UMR 5293, CNRS, CEA, Université de Bordeaux, 33076 Bordeaux, France;

[19] Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital and Departments of Psychology and Psychiatry, University of Toronto, Toronto, Ontario, M6A 2E1, Canada;

[20] Department of Child and Adolescent Psychiatry and Psychotherapy, University Medical Centre Göttingen, von-Siebold-Str. 5, 37075, Göttingen, Germany;

[21] 23 Department of Psychiatry and Neuroimaging Center, Technische Universität Dresden, Dresden, Germany;

[22] School of Psychology and Global Brain Health Institute, Trinity College Dublin, Ireland;

[23] National Institute of Mental Health / NIH, 15K North Drive, Bethesda MD, 20892, USA;

[24] MSB Medical School Berlin, Hochschule für Gesundheit und Medizin, Siemens Villa, Berlin;

[25] Department of Social and Health Care, Psychosocial Services Adolescent Outpatient Clinic Kauppakatu 14, Lahti, Finland;

[26] PONS Research Group, Dept of Psychiatry and Psychotherapy, Campus Charite Mitte, Humboldt University, Berlin and Leibniz Institute for Neurobiology, Magdeburg, Germany, and Institute for Science and Technology of Brain-inspired Intelligence (ISTBI), Fudan University, Shanghai, P.R. China;

Short/running title: Predicting youth depression onset using multimodal data

Keywords

Depression, adolescents, machine learning, prediction, penalized logistic regression, major depressive disorder

Corresponding author

Name: Yara Toenders

Affiliation: Orygen & Centre for Youth Mental Health, The University of Melbourne

Address: 35 Poplar Road, Parkville, VIC 3052, Australia

Telephone: 0061 476227813

Email: yara.toenders@orygen.org.au

ABSTRACT

Background: Adolescent onset of depression is associated with long-lasting negative consequences. Identifying adolescents at risk for developing depression would enable the monitoring of risk-factors and the development of early intervention strategies. Using machine learning to combine several risk factors from multiple modalities might allow prediction of depression onset at the individual level.

Methods: A subsample of a multi-site longitudinal study in adolescents, the IMAGEN study, was used to predict future (subthreshold) major depressive disorder (MDD) onset in healthy adolescents. Based on 2-year and 5-year follow-up data, participants were grouped into: 1) developing an MDD diagnosis or subthreshold MDD and 2) healthy controls. Baseline measurements of 145 variables from different modalities (clinical, cognitive, environmental and structural magnetic resonance imaging [MRI]) at age 14 were used as input to penalized logistic regression (with different levels of penalization) to predict depression onset in a training dataset (N=407). The features contributing highest to the prediction were validated in an independent hold-out sample (3 independent IMAGEN sites; N=137).

Results: The area under the receiver operating characteristics curve (AUROC) for predicting depression onset ranged between 0.70-0.72 in the training dataset. Baseline severity of depressive symptoms, female sex, neuroticism, stressful life events and surface area of the supramarginal gyrus contributed most to the predictive model and predicted onset of depression with an AUROC between 0.68-0.72 in the independent validation sample.

Conclusions: This study showed that depression onset in adolescents can be predicted based on a combination multimodal data of clinical, life events, personality traits, brain structure variables.

INTRODUCTION

Major depressive disorder (MDD) usually has its onset in adolescence and young adulthood (1), which can have deleterious consequences for a young person's educational and occupational functioning, and personal and social life (2). Moreover, adolescent onset depression can have adverse economic consequences for society, since depression onset in adolescence is associated with poorer social and occupational functioning and recurrent or persistent mental illness in adulthood (4, 6). Predicting onset of depression at an early stage is of high clinical relevance, as it might guide the deployment of early interventions and preventions, thereby reducing the negative long-term consequences associated with adolescent onset depression.

Various studies have examined clinical, cognitive and environmental predictors of depression onset (7, 8). However, most of these studies examined cross-sectional associations and, hence, did not provide information on directionality (10, 11). Longitudinal studies are required to study the predictive value of these factors for the onset of depression, but only few studies exist that have investigated the longitudinal association between clinical, cognitive and environmental risk factors and subsequent onset of depression in young people. These studies have shown that risk factors such as anxiety symptoms, diagnosis of another psychiatric disorder, stressful life events and neuroticism precede the onset of depression (12–17). There are few studies that have examined the predictive characteristics of neuroimaging markers, and of those, most were conducted with small sample sizes (18–22). Our recent review showed that findings have been inconsistent, although there is some consistent preliminary evidence for blunted (ventral striatum) response to reward processing as a predictor for later depression (23).

Most of the longitudinal studies investigating clinical, environmental and neurobiological risk factors for the onset of depression in adolescence have examined these risk factors in isolation. It remains to be investigated whether a combination of risk factors may yield better

predictive performance, and which risk factors are most predictive. In addition, most of the studies have used a traditional group comparison approach. However, a statistically significant variable at group level will not necessarily be useful for individual prediction, due to low effect size or because of its redundancy with respect to other variables. Conversely, even seemingly insignificant variables may become important when combined with other variables. Some studies have however used a multimodal approach to predict depression, and have identified important predictors such as sex, neuroticism, rumination, negative affect, low self-esteem, childhood abuse and familial history of mood disorders among others (3, 5, 7, 9). Machine learning-based predictive models are also well suited for combining large amount of data and different data modalities into a single model. In addition, contrary to traditional multivariate prediction methods, they are optimized for evaluating the model's predictive value for previously unseen individuals ("new" individuals). Thus, they allow evaluation of the predictive model at the level of the individual.

A recent machine learning study in 15-year old adolescents using psychosocial variables as predictors showed that school failure, social isolation, involvement in physical fights, drug use, running away from home and maltreatment were predictive of MDD onset within 3 to 4 years after baseline, with a receiver operating characteristics curve (AUROC) between 0.76 and 0.79 (24). Importantly, the predictive model was externally validated in two separate datasets. With regard to neurobiological risk factors, Foland-Ross and colleagues showed that cortical thickness can predict onset of depression within 5 years after a baseline scan with 70% accuracy when 55% of the girls developed depression (25). Thickness of the right precentral and medial orbitofrontal cortex, left anterior cingulate cortex and insula were the most predictive features in their predictive model.

These machine learning studies are an important first step towards the development of a predictive model that enables identification of adolescents at risk for depression. A critical next step is to elucidate whether we can predict depression onset in adolescents using a

combination of risk factors found in these studies described above (neurobiological, clinical, cognitive and environmental). Therefore, in the current study we examined the predictive value of a multimodal data, using of clinical, cognitive, environmental and neurobiological variables, for the onset of MDD, including subthreshold MDD. We included subthreshold MDD as the DSM diagnostic criteria for adolescent MDD have low diagnostic validity and specificity, with unclear diagnostic boundaries (26, 27). In addition, earlier studies have shown that subthreshold MDD is associated with a higher risk for developing future MDD and other adverse effects that are associated with MDD (28), highlighting the clinical importance of considering subthreshold MDD when predicting onset of depression in adolescence. We employed a machine learning method (penalized logistic regression) as this machine learning algorithm is appropriate to identify, in combination with a feature selection approach, the optimal set of measures that prospectively predict onset of depression over 5 years in a subsample of 407 subjects from the IMAGEN study who were aged 14 at baseline(29). The predictive model was validated in an independent hold-out sample from the IMAGEN study (N=137), and specificity for depression onset was tested in a sample with risky alcohol use (N=268). To our knowledge, this is the first machine learning study in adolescents that combines a number of different modalities to predict depression onset.

METHODS

Participants

The IMAGEN cohort study is a multisite study, in which the baseline (BL) sample consisted of 2223 adolescents (around 14 years old) who were followed-up at age 16 (follow-up 1; FU1), 19 (FU2) and 22 (FU3, these data are still being collected) (29). The participants were recruited from schools, and their diversity in terms of academic performance, socioeconomic status and behavioral and emotional functioning was maximized. Exclusion criteria included: receiving treatment for schizophrenia or bipolar disorder, IQ<70, autism diagnosis, nutritional

or metabolic diseases, neurological conditions (e.g. brain tumor, epilepsy) and other medical diagnoses. The data were collected from 8 sites in Europe (France, Germany, Ireland and United Kingdom). Ethics was approved by local ethics committees. Participants' parents signed informed consent and participants gave written assent. Participants older than 18 gave informed consent at FU2. Detailed information about the study protocol can be found in prior literature (29).

At each timepoint, participants filled out a psychiatric symptom self-assessment using the Development and Well-Being Assessment (DAWBA) (30). We used the self-report version of the DAWBA instead of the clinical version in order to be consistent with previous reports. Three groups were created based on the DAWBA self-assessment: (1) healthy controls who did not meet criteria for any mental disorder or subthreshold MDD at any of the assessments (N=430), (2) those who developed subthreshold MDD at follow-up (FU1 and/or FU2; N=177) or full-threshold MDD at follow-up (N=71). We excluded participants who met criteria for a psychiatric diagnosis, or subthreshold MDD, at baseline. Full-threshold MDD and subthreshold MDD were defined based on earlier research in the IMAGEN sample (Supplementary Methods and Supplementary Figure 1) (31, 32). We will use the term 'depression' when referring to the combined group of subthreshold MDD and full-threshold MDD. We kept a subset of healthy controls (N=134) for a sensitivity analysis with regard to predicting onset of risky alcohol use (described below). Therefore, N=296 healthy controls were included for the main analysis (Supplementary Figure 2).

To investigate whether our model's performance was specific to the prediction of onset of depression or was broadly predictive of psychopathology, additional groups (non-overlapping) were defined based on onset of risky alcohol use at FU. A risky alcohol group (N=134) was defined by having a total score of 8 or above on the AUDIT at FU1 and/or FU2, while not meeting criteria for any other psychiatric disorder (including MDD and subthreshold MDD) at BL and FU. The healthy controls for this analysis were a randomly selected

subsample (to match the number of participants in the risky alcohol use group, N=134) of those participants that did not meet criteria for psychiatric disorders and had a score lower than 8 on the AUDIT at BL and FU.

Predictor variables

All measures were collected at multiple timepoints, however only baseline variables were included as predictors in this study. Demographic (n=2), clinical (n=7), cognitive (n=24), personality (n=9), environmental (n=22), substance use (n=4), developmental (n=1) and structural MRI (i.e. surface-based morphometry) (n=76) variables were used as predictors. In total, 145 predictors were included from these different modalities, described in the Supplementary Methods and Supplemental Table S2.

Statistical analysis

Splitting the sample into training and validation sets

The dataset for the main analysis was divided into a training dataset (N=407) and independent validation dataset (N=137) based on recruitment site. Data from three randomly selected recruitment sites (Dublin, Mannheim and Paris) were kept separately as the independent validation set (between-site split). The other five sites formed the training dataset. A between-site split instead of within-site split was chosen to examine if the model would generalize to completely new sites, which is especially relevant for neuroimaging, as machine learning models can be influenced by scanner effects. The age, sex and diagnosis distribution did not differ between training and validation set. The group labels we aimed to predict were 1) healthy controls versus 2) those who developed depression at follow-up.

Prediction of depression onset at follow-up in training dataset

Penalized (to prevent overfitting) logistic regression was performed on the training dataset including all predictors to predict depression onset at follow-up (Figure 1) (33). We tested

model performance across four different values for α (1 to 0.25, with 0.25 decreases) in the penalized logistic regression. When α is 1, it means that the Lasso penalty was applied, and when α decreased a combination of Lasso and Ridge penalties were applied. Lasso facilitates feature selection as it shrinks coefficients of features to zero, thereby removing these features from the model. Multiple values of α were used to examine which features were selected consistently. The hyperparameter λ value, the weight of the penalty, was determined by selecting the optimal λ associated with the minimum Brier score in an inner cross-validation loop. Using the R package 'glmnet', a sparse model that uses feature selection was created (34). We applied a 10-fold cross-validation (CV), which was repeated 10 times. For the CV, the training data was divided into 10 sets, and within each cross-validation fold, 9 sets formed the training set while the 10^{th} was held out for testing. We ensured that the distribution of scanning sites within each group was the same across all 10 CV folds in order to correct for possible site effects. In each CV fold, a random subsample of healthy controls was selected to match the number of participants in the depression group. All variables were scaled and centered in the fold and missing values were imputed in the training sets based on data of the 5 nearest neighbors (35). The parameters of the training set were used to impute the test set separately to prevent data leakage.

To identify features that contributed most to the prediction, the models (at different levels of α) were fitted 10 times in random subsamples within the 10 folds (75% of the training dataset in the fold). Features that were selected in 90% or more of the 100 repeats were identified (36).

Replication in independent validation dataset

The features that were identified as most predictive (i.e., selected at least 90% of the times in the random subsamples) were subsequently used to build a Ridge logistic regression (α = 0) model using the whole training dataset. The Ridge regression approach ensured that all

features were used in the model. This model was then applied to the independent validation dataset (3 recruitment sites as a separate hold-out sample) to evaluate the predictive value of this subset of features for onset of depression in participants from independent sites.

Performance measures

Performance of the models was examined using the AUROC, sensitivity, specificity and balanced accuracy (average of sensitivity and specificity). The AUROC represents the probability that a subject from the depression group is ranked lower than a randomly selected HC subject across all classification thresholds. An AUROC higher than 0.5 is performing better than chance level. Permutation testing was used to test if the models performed statistically better than chance level prediction (1000 permutations with randomly permuted group labels). A non-parametric significance level p-value was estimated as the proportion the randomly permuted groups that had a higher AUROC than the AUROC for the original groups.

Prediction of future risky alcohol use

To evaluate if the features that were selected in the training set were specific to predicting onset of depression or whether they predict onset of psychopathology more generally, we used the Ridge model with the selected feature to predict risky alcohol use at follow-up.

Prediction of MDD

To assess if we could predict onset of MDD, we did an exploratory penalized logistic regression in a CV predicting MDD in the whole IMAGEN dataset (8 sites) (see Supplemental Material), once excluding subthreshold depression (N = 349) and once with those with subthreshold depression included in the HC group (N=513).

[Figure 1]

RESULTS

Demographic and clinical characteristics of the healthy controls and participants who developed depression can be found in Table 1 and Supplemental Table S3.

[Table 1]

Prediction of depression onset

In the training dataset, depression onset (subthreshold and full-threshold MDD combined) could be predicted with an AUROC ranging between 0.70 and 0.72 across different levels of α (Table 2). This was significantly different from chance level (all p-values = 0.001).

[Table 2, Table 3]

Feature selection

With an α of 1, 4 features were selected in the feature selection procedure (Supplemental Table 3), as well as one recruitment site (Dresden). The features selected were depression score at baseline, sex and lifetime frequency of events in the family (sum score of the presence or absence of events such as parents divorced, abused alcohol, fought or argued, remarried or had money problems) and distress (seeing therapist, thought about suicide, face broke out in pimples, ran away, gained a lot of weight, got poor grades in school) categories. At α of 0.75 and 0.50, the same features were selected but with the addition of surface area of the supramarginal gyrus. Being bullied at school, neuroticism and verbal comprehension were additionally selected when α was 0.25.

Generalization to independent validation dataset

The features that were selected in the penalized logistic regression were used to predict depression onset in the independent validation dataset (3 independent IMAGEN sites), and an AUROC ranging between 0.68 and 0.72 was achieved (Table 4).

[Table 4]

Generalization to onset of risky alcohol use at follow-up

Demographic and clinical characteristics of the participants that had risky alcohol use at follow-up can be found in Supplemental Table S4. The model was able to discriminate between participants with risky alcohol use at follow-up and healthy controls with AUROC of 0.62 when using the features selected at different levels of α in the model predicting onset of depression (Table 3).

DISCUSSION

In a large longitudinal sample of young people, we were able to prospectively predict depression onset with an AUROC ranging between 0.70 and 0.72 using penalized logistic regression applied to a large set of clinical, cognitive, developmental, personality and neurobiological characteristics. Importantly, our prediction model was validated in an independent validation sample consisting of participants of the IMAGEN study assessed at independent sites (AUROC range 0.68-0.72), confirming the validity of the predictive model and its generalizability to independent recruitment sites.

Monitoring risk factors identified in this study could lead to early identification of those at risk for developing depression, which could help the development of risk-factor specific strategies for prevention of onset of depression. However, the question arises if an AUROC of 0.72 is high enough for a predictive model to be clinically relevant. Of note, the AUROC range is concordant with validated prognostic studies in psychosis (0.73-0.79), bipolar disorder (0.76) and cardiovascular disease (0.76-0.79) (37–39). The clinical utility of a

machine learning model should be assessed by considering the cost-effectiveness of monitoring risk factors for depression identified by the prediction model. Due to the high levels of disability that depression can cause, with consequences for not only the individual but for the broader community, monitoring low-cost risk factors such as clinical characteristics or life events that can predict depression onset in adolescents with an AUROC of 0.70 might be sufficient.

The relative contribution of the predictors should be interpreted with caution as the model performance is based on multivariate data, and features with small weights still contribute to the overall performance of the model. However, using only the subset of features that made substantial contributions to the prediction in the training set to predict depression onset in an independent validation dataset yielded similar AUROCs as in the training set (0.68-0.72). Higher depressive symptoms at baseline, being bullied at school, neuroticism, female sex and more negative life events were found to be among the largest contributors to depression onset, which is in line with previous research that examined these pre-existing risk factors in isolation, using multivariate non-machine learning methods or a machine-learning method (7, 14, 17, 24, 28). We found that a higher level of depressive symptoms was an important predictor for subsequent onset of depression, even though participants with subthreshold depression at baseline were excluded and thus the mean level of depressive symptoms at baseline was low (mean: 0.75, on a scale from 0-14). This may be due to shared method variance. The selection of negative life events seems to suggest that early life stress is an important predictor of depression onset, and that experiencing stressful life events could be a valid prospective risk factor to monitor. Additionally, the use of machine learning methods including internal and external validation in the current study strengthens the hypothesis that the predictive characteristics could be extrapolated to new individuals (40). However, the performance of the predictive model will likely have to be improved for it to be clinically useful. Future studies could focus on sex-specific predictors of depression, which might help improve the performance.

With regard to brain measures, we found that lower surface area of the supramarginal gyrus contributed to the model's predictive performance. Prior research has shown that cortical surface area alterations may play a particular role when depressive symptoms are experienced early in adolescence (41, 42). Given that cortical surface area, compared with cortical thickness, has a higher genetic heritability (43), is determined earlier in development, and is less strongly affected by later environmental influences (44), cortical surface area reductions may represent a pre-existing risk factor for depression, shaped by genetic factors and/or early life adversity (45). Of note, surface area of the supramarginal gyrus, involved in complex higher order cognitive processes, was not identified to be associated with MDD in adolescents in a large consortium study (N=505 adolescents) (41, 46, 47). Since the supramarginal gyrus was not selected at the highest α thus not affecting the AUROC, and has the supramarginal gyrus been identified as an important brain region in adolescent depression in previous literature, the predictive role of the surface area of the supramarginal gyrus is most likely marginal. This is in contrast to a previous study by Foland-Ross and colleagues who found a similar AUROC including only cortical thickness measures to predict depression onset in a relatively small sample (N=34) of young adolescent girls (25). An important difference between the Foland-Ross study and the current study is that we also included participants with subthreshold depression, and included multimodal predictors with other modalities that might be more informative than cortical thickness. Given that no other surface area regions, cortical thickness or subcortical volumes measures were identified in our feature selection approach and as it is costly to acquire structural neuroimaging measures, structural imaging might not be a useful predictor for depression onset in young people. However, this does not implicate that structural brain changes in young people with depression cannot provide information about the underlying mechanisms of depression.

The model was not specific to predicting depression onset at follow-up, but could also successfully predict risky alcohol use in an independent sample, with a slightly lower

AUROC (0.62). This may not be surprising, given the high comorbidity between alcohol use disorder and MDD, with an increase in comorbidity in young adulthood (46). In addition, a risky lifestyle, including risky alcohol use, in adolescence is predictive of depressive symptoms (47). Lastly, risky alcohol use occurred in the depression group, which might have contributed to the lack of specificity of the predictors. Beyond this, comorbidity of mental disorders is common; most people who experience mental illness will be diagnosed with more than one psychiatric disorder during their lifetime and an early age of onset of the first psychiatric disorder has been associated with having more comorbid psychiatric disorders during the lifetime (48). We anticipate that our model could be similarly predictive for the onset of mental disorders other than depression or alcohol abuse, in line with previous longitudinal studies showing that other psychiatric disorders are associated with similar risk factors as the risk factors identified in the current study such as bullying, neuroticism, depressive symptoms and stressful life events (28, 49–51). However, since the prevalence of other disorders such as bipolar disorder and psychosis were limited in the IMAGEN sample, the hypothesis about the non-specificity for depression of the model requires further investigation in other samples.

When the analysis was restricted to patients with MDD and those with subthreshold depression were excluded, the AUROC was higher than in the main analysis. Unfortunately, the sample size of the MDD group was too small to allow validation in an independent dataset. This increase in AUROC when excluding subthreshold depression could be due to the fact that adolescents who will develop MDD are more differentiated from healthy adolescents than adolescents developing subthreshold depressive symptoms. When those who developed subthreshold depression were treated as healthy controls, the AUROC decreased. These findings further support the postulation that depression based on a cut-off for a diagnosis of MDD is arbitrary, as young people with a full-threshold MDD diagnosis cannot reliably be distinguished from those with subthreshold depression as indicated by our findings.

The current study has major strengths, including its large sample size, longitudinal design, and integration of predictors across multiple modalities. However, an important limitation is that the diagnostic information was based on the self-report DAWBA, a measure that only captures a period of four weeks prior to each follow-up assessment. Since there was no information available on possible depressive episodes in the periods between the follow-ups, we may have missed depression in the healthy control group, which might have impacted the classification performance of our model (though likely in the direction of weakening it). It could also have led to underdiagnosing depression at baseline, potentially leading to a less healthy group at baseline. Additionally, the DAWBA is clinically reliable (30), although due to the use of a self-report measure, symptoms might have been underreported (52).

There are still challenges with translating these types of models into clinical practice, including that the rate of depression is high in the selected sample. Participants with a psychiatric diagnosis were removed from the healthy control group, which limits the clinical utility of the model as people in the general population might show non-depression psychiatric diagnoses. Therefore, future studies should test if a predictive model works in the general population that includes people who have already experienced episodes of mental ill health. Additionally, the depression group might include young people with comorbid diagnoses such as anxiety disorders with similar risk factors, which could increase the predictive power. However, in a sensitivity analysis, excluding those who developed comorbid anxiety disorder at follow-up, showed that the predictive performance measures were similar.

In conclusion, the current study showed that depression onset in adolescents can be predicted based on multimodal data, including clinical, cognitive, life events, personality traits and neurobiological variables. The variables contributing most to the predictive model were found to be depressive symptoms at baseline, neuroticism, cognition, supramarginal gyrus

surface area and stressful life events. Since the model was also predictive of onset of risky alcohol use, these risk factors may likely be predictive more generally of onset psychopathology during adolescence.

Acknowledgements

This work was supported by the MQ Brighter Futures Award MQBFC/2 (LS) and the National Institute of Mental Health of the National Institutes of Health under Award Number R01MH117601 (LS). LS is supported by a NHMRC Career Development Fellowship (1140764), and the Dame Kate Campbell Fellowship from the Faculty of Medicine, Dentistry and Health Sciences at The University of Melbourne. CGD is supported by an NHMRC Career Development Award (141738).

This work received support from the following sources: the European Union-funded FP6 Integrated Project IMAGEN (Reinforcement-related behaviour in normal brain function and psychopathology) (LSHM-CT- 2007-037286), the Horizon 2020 funded ERC Advanced Grant 'STRATIFY' (Brain network based stratification of reinforcement-related disorders) (695313), Human Brain Project (HBP SGA 2, 785907, and HBP SGA 3, 945539), the Medical Research Council Grant 'c-VEDA' (Consortium on Vulnerability to Externalizing Disorders and Addictions) (MR/N000390/1), the National Institute of Health (NIH) (R01DA049238, A decentralized macro and micro gene-by-environment interaction analysis of substance use behavior and its brain biomarkers), the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, the Bundesministeriumfür Bildung und Forschung (BMBF grants 01GS08152; 01EV0711; Forschungsnetz AERIAL 01EE1406A, 01EE1406B; Forschungsnetz IMAC-Mind 01GL1745B), the Deutsche Forschungsgemeinschaft (DFG grants SM 80/7-2, SFB 940, TRR 265, NE 1383/14-1), the Medical Research Foundation and Medical Research Council (grants MR/R00465X/1 and MR/S020306/1), the National Institutes of Health (NIH) funded ENIGMA (grants 5U54EB020403-05 and 1R56AG058854-01). Further support was provided by grants from: –

the ANR (ANR-12-SAMA-0004, AAPG2019 – GeBra), the Eranet Neuron (AF12-NEUR0008-01 – WM2NA; and ANR-18-NEUR00002-01 – ADORe), the Fondation de France (00081242), the Fondation pour la Recherche Médicale (DPA20140629802), the Mission Interministérielle de Lutte-contre-les-Drogues-et-les-Conduites-Addictives (MILDECA), the Assistance-Publique-Hôpitaux-de-Paris and INSERM (interface grant), Paris Sud University IDEX 2012, the Fondation de l'Avenir (grant AP-RM-17-013), the Fédération pour la Recherche sur le Cerveau; the National Institutes of Health, Science Foundation Ireland (16/ERCD/3797), U.S.A. (Axon, Testosterone and Mental Health during Adolescence; RO1 MH085772-01A1), and by NIH Consortium grant U54 EB020403, supported by a cross-NIH alliance that funds Big Data to Knowledge Centres of Excellence.

Disclosures

Dr Banaschewski served in an advisory or consultancy role for Lundbeck, Medice, Neurim Pharmaceuticals, Oberberg GmbH, Shire. He received conference support or speaker's fee by Lilly, Medice, Novartis and Shire. He has been involved in clinical trials conducted by Shire & Viforpharma. He received royalties from Hogrefe, Kohlhammer, CIP Medien, Oxford University Press. The present work is unrelated to the above grants and relationships. Dr Barker has received honoraria from General Electric Healthcare for teaching on scanner programming courses. Dr Poustka served in an advisory or consultancy role for Roche and Viforpharm and received speaker's fee by Shire. She received royalties from Hogrefe, Kohlhammer and Schattauer. The present work is unrelated to the above grants and relationships. The other authors report no biomedical financial interests or potential conflicts of interest.

References

1. Weissman MM, Bland RC, Canino GJ, Faravelli C, Greenwald S, Hwu H, *et al.* (1996): Cross-National Epidemiology of Major Depression and Bipolar Disorder. *JAMA Psychiatry*. 276: 293–299.

2. Copeland WE, Alaie I, Jonsson U, Shanahan L (2020): Associations of Childhood and Adolescent Depression With Adult Psychiatric and Functional Outcomes. *J Am Acad Child Adolesc Psychiatry*. . doi: 10.1016/j.jaac.2020.07.895.

3. Kendler KS, Gardner CO, Prescott CA (2002): Toward a comprehensive developmental model for major depression in women. *Am J Psychiatry*. 159: 1133–1145.

4. Fergusson DM, Woodward LJ (2002): Mental Health, Educational, and Social Role Outcomes of Adolescents With Depression. *Arch Gen Psychiatry*. 59: 225.

5. Kendler KS, Gardner CO, Prescott CA (2006): Toward a comprehensive developmental model for major depression in men. *Am J Psychiatry*. 163: 115–124.

6. McLeod GFH, Horwood LJ, Fergusson DM (2016): Adolescent depression, adult mental health and psychosocial outcomes at 30 and 35 years. *Psychol Med*. 46: 1401–1412.

7. Klein D, Glenn C, Kosty D, Seeley J, Rohde P, Lewinsohn PM (2013): Predictors of First Lifetime Onset of Major Depressive Disorder in Young Adulthood. *J Abnorm Psychol*. 122: 1–6.

8. Eaton W, Shao H, Nestadt G, Lee B, Bienvenu O, Zandi P (2008): Population-Based Study of First Onset and Chronicity in Major Depressive Disorder. *Arch Gen Psychiatry*. 65: 513–520.

9. Cohen JR, Thakur H, Young JF, Hankin BL (2020): The development and validation of an algorithm to predict future depression onset in unselected youth. *Psychol Med*. 50: 2548–2556.

10. Warner V, Weissman MM, Mufson L, Wickramaratne PJ (1999): Grandparents, Parents, and Grandchildren at High Risk for Depression: A Three-Generation Study. *J Am Acad Child Adolesc Psychiatry*. 38: 289–296.

11. Pine DS, Cohen P, Brook J (2001): Adolescent fears as predictors of depression. *Biol Psychiatry*. 50: 721–724.

12. Keenan K, Feng X, Hipwell A, Klostermann S (2009): Depression begets depression: Comparing the predictive utility of depression and anxiety symptoms to later depression. *J Child Psychol Psychiatry*. 50: 1167–1175.

13. Ebert DD, Buntrock C, Mortier P, Auerbach R, Weisel KK, Kessler RC, *et al.* (2018): Prediction of major depressive disorder onset in college students. *Depress Anxiety*. 36: 294–304.

14. Cole MG, Dendukuri N (2003): Risk factors for depression among elderly community subjects: A systematic review and meta-analysis. *Am J Psychiatry*. 160: 1147–1156.

15. Hovens JGFM, Giltay EJ, Spinhoven P, Van Hemert AM, Penninx BWJH (2015): Impact of Childhood Life Events and Childhood Trauma on the Onset and Recurrence of Depressive and Anxiety Disorders. *J Clin Psychiatry*. 76: 931–938.

16. Brown J, Cohen P, Johnson JG, Smailes EM (1999): Childhood Abuse and Neglect: Specificity of Effects on Adolescent and Young Adult Depression and Suicidality. *J Am Acad Child Adolesc Psychiatry*. 38: 1490–1496.

17. Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ (2015): A Longitudinal Twin Study of Personality and Major Depression in Women. *Arch Gen Psychiatry*. 1993: 853–862.

18. Whittle S, Lichter R, Dennison M, Vijayakumar N, Schwartz O, Byrne ML, *et al.* (2014): Structural brain development and depression onset during adolescence: A prospective longitudinal study. *Am J Psychiatry*. 171: 564–571.

19. Shapero BG, Chai XJ, Vangel M, Biederman J, Hoover CS, Whitfield-Gabrieli S, *et al.* (2019): Neural markers of depression risk predict the onset of depression. *Psychiatry Res - Neuroimaging*. 285: 31–39.

20. Barch DM, Harms MP, Tillman R, Hawkey E, Luby JL (2019): Early childhood depression, emotion regulation, episodic memory, and hippocampal development. *J Abnorm Psychol*. 128: 81–95.

21. Pagliaccio D, Luby JL, Luking KR, Belden AC, Barch DM (2014): Brain–behavior relationships in the experience and regulation of negative emotion in healthy children: Implications for risk for childhood depression. *Dev Psychopathol*. 26: 1289–1303.

22. Papmeyer M, Sussmann JE, Stewart T, Giles S, Centola JG, Zannias V, *et al.* (2016): Prospective longitudinal study of subcortical brain volumes in individuals at high familial

583 risk of mood disorders with or without subsequent onset of depression. *Psychiatry Res*
584 *Neuroimaging*. 248: 119–125.
585 23. Toenders YJ, van Velzen LS, Heideman IZ, Harrison BJ, Davey CG, Schmaal L (2019):
586 Neuroimaging predictors of onset and course of depression in childhood and
587 adolescence: A systematic review of longitudinal studies. *Dev Cogn Neurosci*. 39:
588 100700.
589 24. Rocha TB, Fisher HL, Caye A, Anselmi L, Arseneault L, Barros FC, *et al.* (2019):
590 Identifying Adolescents at Risk for Depression: A Prediction Score Performance in
591 Cohorts Based in Three Different Continents. *J Am Acad Child Adolesc Psychiatry*. .
592 doi: 10.1016/j.jaac.2019.12.004.
593 25. Foland-Ross L, Sacchet MD, Prasad G, Gilbert B, Thompson PM, Gotlib IH (2015):
594 Cortical Thickness Predicts the First Onset of Major Depression in Adolescence. *Int J*
595 *Dev Neurosci*. 46: 125–131.
596 26. Blom EH, Forsman M, Yang TT, Serlachius E, Larsson J-O (2014): Latent Classes of
597 Symptoms related to Clinically Depressed Mood in Adolescents. *Scand J Child Adolesc*
598 *Psychiatry Psychol*. 2: 19–28.
599 27. Costello EJ, Mustillo S, Erkanli A, Keeler G, Angold A (2003): Prevalence and
600 development of psychiatric disorders in childhood and adolescence. *Arch Gen*
601 *Psychiatry*. 60: 837–844.
602 28. Fergusson DM, Horwood LJ, Ridder EM, Beautrais AL (2005): Subthreshold depression
603 in adolescence and mental health outcomes in adulthood. *Arch Gen Psychiatry*. 62: 66–
604 72.
605 29. Schumann G, Loth E, Banaschewski T, Barbot A, Barker G, Büchel C, *et al.* (2010): The
606 IMAGEN study: reinforcement-related behaviour in normal brain function and
607 psychopathology. *Mol Psychiatry*. 15: 1128–1139.
608 30. Goodman R, Ford T, Richards H, Gatward R, Meltzer H (2000): The Development and
609 Well-Being Assessment: Description and Initial Validation of an Integrated Assessment
610 of Child and Adolescent Psychopathology. *J Child Psychol Psychiatry*. 41: 645–655.

31. Vulser H, Lemaitre H, Artiges E, Miranda R, Penttilä J, Struve M, *et al.* (2015): Subthreshold depression and regional brain volumes in young community adolescents. *J Am Acad Child Adolesc Psychiatry*. 54: 832–840.

32. Stringaris A, Belil PVR, Artiges E, Lemaitre H, Gollier-Briant F, Wolke S, *et al.* (2015): The brain's response to reward anticipation and depression in adolescence: Dimensionality, specificity, and longitudinal predictions in a community-based sample. *Am J Psychiatry*. 172: 1215–1223.

33. Dinga R, Marquand AF, Veltman DJ, Beekman ATF, Schoevers RA, van Hemert AM, *et al.* (2018): Predicting the naturalistic course of depression from a wide range of clinical, psychological, and biological data: a machine learning approach. *Transl Psychiatry*. 8: 241.

34. Friedman J, Hastie T, Tibshirani R (2010): Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 33: 1–22.

35. Kuhn M (2008): Building Predictive Models in R Using the caret Package. *J Stat Softw*. 28: 1–26.

36. Ing A, Sämann PG, Chu C, Tay N, Biondo F, Robert G, *et al.* (2019): Identification of neurobehavioural symptom groups based on shared brain mechanisms. *Nat Hum Behav*. 3: 1306–1318.

37. Hafeman DM, Merranko J, Goldstein TR, Axelson D, Goldstein BI, Monk K, *et al.* (2017): Assessment of a person-level risk calculator to predict new-onset bipolar spectrum disorder in youth at familial risk. *JAMA Psychiatry*. 74: 841–847.

38. Fusar-Poli P, Werbeloff N, Rutigliano G, Oliver D, Davies C, Stahl D, *et al.* (2019): Transdiagnostic risk calculator for the automatic detection of individuals at risk and the prediction of psychosis: Second replication in an independent national health service trust. *Schizophr Bull*. 45: 562–570.

39. D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, Kannel WB (2008): General cardiovascular risk profile for use in primary care: The Framingham heart study. *Circulation*. 117: 743–753.

40. Bzdok D, Meyer-Lindenberg A (2017): Machine learning for precision psychiatry. 1–16.
41. Schmaal L, Hibar DP, Sämann PG, Hall GB, Baune BT, Jahanshad N, *et al.* (2017): Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. *Mol Psychiatry*. 22: 900–909.
42. Schmaal L, Yücel M, Ellis R, Vijayakumar N, Simmons JG, Allen NB, Whittle S (2017): Brain Structural Signatures of Adolescent Depressive Symptom Trajectories: A Longitudinal Magnetic Resonance Imaging Study. *J Am Acad Child Adolesc Psychiatry*. 56: 593-601.e9.
43. Eyler LT, Chen C, Panizzon MS, Fennema-Notestine C, Neale MC, Jak A, *et al.* (2012): A Comparison of Heritability Maps of Cortical Surface Area and Thickness and the Influence of Adjustment for Whole Brain Measures: A Magnetic Resonance Imaging Twin Study. *Twin Res Hum Genet*. 15: 304–314.
44. Wierenga LM, Langen M, Oranje B, Durston S (2014): Unique developmental trajectories of cortical thickness and surface area. *Neuroimage*. 87: 120–126.
45. Schmaal L (2019): Cortical surface area: a potential biological marker for depression onset and poor clinical outcomes? *The Lancet Psychiatry*. 6: 277–279.
46. Grant BF, Harford TC (1995): Comorbidity between DSM-IV alcohol use disorders and major depression: results of a national survey. *Drug Alcohol Depend*. 39: 197–206.
47. Mewton L, Champion K, Kay-Lambkin F, Sunderland M, Thornton L, Teesson M (2019): Lifestyle risk indices in adolescence and their relationships to adolescent disease burden: Findings from an Australian national survey. *BMC Public Health*. 19: 1–10.
48. Caspi A, Houts RM, Ambler A, Danese A, Elliott ML, Hariri A, *et al.* (2020): Longitudinal Assessment of Mental Health Disorders and Comorbidities Across 4 Decades Among Participants in the Dunedin Birth Cohort Study. *JAMA Netw open*. 3: e203221.
49. Tessner KD, Mittal V, Walker EF (2011): Longitudinal study of stressful life events and daily stressors among adolescents at high risk for psychotic disorders. *Schizophr Bull*. 37: 432–441.

667 50. Mackie CJ, O'Leary-Barrett M, Al-Khudhairy N, Castellanos-Ryan N, Struve M, Topper L,
668 Conrod P (2013): Adolescent bullying, cannabis use and emerging psychotic
669 experiences: A longitudinal general population study. *Psychol Med*. 43: 1033–1044.
670 51. Kotov R, Gamez W, Schmidt F, Watson D (2010): Linking "Big" personality traits to
671 anxiety, depressive, and substance use disorders: A meta-analysis. *Psychol Bull*. 136:
672 768–821.
673 52. Eaton WW, Neufeld K, Chen L-S, Cai G (2000): A Comparison of Self-report and Clinical
674 Diagnostic Interviews for Depression. *Arch Gen Psychiatry*. 57: 217.
675
676

677 **Table 1. Demographics and clinical characteristics of the groups in the training and**

678 **validation datasets.**

	Training depression (n=180)	Training control (n=227)	Validation depression (n=68)	Validation control (n=69)
Age				
Mean (SD)	14.5 (0.54)	14.4 (0.44)	14.4 (0.59)	14.4 (0.61)
Sex				
Female	121 (67%)	104 (46%)	46 (68%)	29 (42%)
Male	59 (33%)	123 (54%)	22 (32%)	40 (58%)
Site				
Berlin	34 (19%)	17 (8%)	NA	NA
Dresden	17 (9%)	62 (27%)	NA	NA
Hamburg	35 (19%)	45 (20%)	NA	NA
London	47 (26%)	53 (23%)	NA	NA
Nottingham	47 (26%)	50 (22%)	NA	NA
Dublin	NA	NA	21 (31%)	11 (16%)
Mannheim	NA	NA	20 (29%)	27 (39%)
Paris	NA	NA	27 (40%)	31 (45%)
Depression score at BL (DAWBA)*				
Mean (SD)	1.07 (1.23)	0.59 (0.80)	0.82 (0.88)	0.36 (0.57)

DAWBA: development and well-being assessment, MDD: major depressive disorder, N: sample size, SD: standard deviation. * score based on number of depressive symptoms present according to youth self-report DAWBA, ranges between 0-14.

Table 2. Performance measures in penalized logistic regression for four different α (Ridge towards Lasso penalty) to predict depression onset in the training set.

α	*AUROC*	*SD AUROC*	*Sensitivity*	*Specificity*	*Accuracy*
0.25	0.70	0.10	0.66	0.66	0.66
0.5	0.70	0.08	0.66	0.65	0.65
0.75	0.72	0.08	0.67	0.66	0.66
1	0.71	0.07	0.65	0.66	0.66

SD: standard deviation across folds

Table 3. Selected features in penalized logistic regression for prediction of depression onset at different α levels in training dataset.

Predictor category	**Parameter threshold**			
	α = *0.25*	**α = *0.5***	**α = *0.75***	**α = *1***
Clinical	DAWBA depression	DAWBA depression	DAWBA depression	DAWBA depression
Life events	LEQ family lifetime	LEQ family lifetime	LEQ family lifetime	LEQ family lifetime
	LEQ distress lifetime	LEQ distress lifetime	LEQ distress lifetime	LEQ distress lifetime
	Bullied at school			
Personality	Neuroticism			

Cognitive	WISC-IV Similarities			
Biological	Sex	Sex	Sex	Sex
	Supramarginal gyrus surface area	Supramarginal gyrus surface area	Supramarginal gyrus surface area	

690 DAWBA: development and well-being assessment, LEQ: life events questionnaire, WISC-IV:
691 Wechsler intelligence scale for children – fourth edition.

692

693 **Table 4.** Performance measures of Ridge logistic regression with the features that were
694 selected in the training dataset across different levels of α to predict depression at follow-up
695 in the independent validation dataset.

696

Number of features (selected at which α in training set)	AUROC	Sensitivity	Specificity	Accuracy
Predicting depression in independent validation dataset				
8 *(α=0.25)*	0.72	0.51	0.83	0.67
5 *(α=0.50 and 0.75)*	0.68	0.49	0.77	0.63
4 *(α=1)*	0.71	0.50	0.81	0.66
Predicting risky alcohol use in independent dataset				
8 *(α=0.25)*	0.62	0.41	0.74	0.57
5 *(α=0.50 and 0.75)*	0.62	0.43	0.79	0.61
4 *(α=1)*	0.62	0.42	0.78	0.60

697 SD: standard deviation

698

Figure 1 Statistical procedure for penalized logistic regression. 1. Baseline predictors from different domains were used to predict subthreshold MDD or MDD onset at follow-up. 2. Penalized logistic regression with 10-fold cross-validation was applied to the training dataset (5 sites) and repeated 10 times with 4 different levels of α. Permutation testing was used to test the statistical significance of the model. 3. Features that were selected in 90% or more in 100 random subsamples of the training data were selected to be tested in the independent validation set. 4. The selected features from step 3 were used as input to Ridge logistic regression in the whole training set that was then used to predict depression onset in the validation set (3 independent sites), to test the generalizability of the model in the three sites that were left out from the training set. 5. The same Ridge model was used to evaluate its predictive value for onset of risky alcohol use in unseen individuals.