Original Article

*List of FondaMental Advanced Center of Expertise for Bipolar Disorders (FACE-BD) collaborators:

FACE-BD Clinical Coordinating Center (Fondation FondaMental): B. Etain, E. Olié, M. Leboyer, P.M. Llorca; FACE-BD Data Coordinating Center (Fondation FondaMental): V. Barteau, S. Bensalem, O. Godin, H. Laouamri, and K. Souryis; FACE-BD Clinical Sites and Principal Collaborators in France; AP-HP, DHU PePSY, Pôle de Psychiatrie et d'Addictologie des Hôpitaux Universitaires H Mondor, Créteil, France: S. Hotier, A. Pelletier; APHP, Groupe Hospitalo-universitaire AP-HP Nord, DMU Neurosciences, Département de Psychiatrie et Médecine Addictologie, Hôpital Fernand Widal, Paris, France: F. Bellivier, B. Etain, V. Hennion, E. Marlinge; Hôpital C. Perrens, Centre Expert Trouble Bipolaire, Service de Psychiatrie Adulte, Pôle 3-4-7, Bordeaux: B. Aouizerate, A. Desage, S. Gard; Département d'Urgence et Post Urgence Psychiatrique, CHRU Montpellier, Montpellier, France: P. Courtet, D. Ducasse, F. Molière, E. Olié; Pôle de Psychiatrie, addictologie et pédopsychiatrie, Hôpital Sainte Marguerite, Marseille, France: A. Lefrere, E.Moreau, J. Pastol; Pôle Hospitalo-Universitaire de Psychiatrie du Grand Nancy, Centre Psychothérapique de Nancy, Laxou, France: G. Gross, R. Schwan, T. Schwitzer; Service Universitaire de Psychiatrie, CHU de Grenoble et des Alpes, Grenoble, France: T. Bougerol, A. Pouchon and M. PolosanCentre Hospitalier de Versailles, Centre Hospitalier de Versailles, Service Universitaire de Psychiatrie d'Adultes et d'Addictologie, Le Chesnay: G. Bonny, L. Brehon, L. Durand, V. Feuga, A.M. Galliot, N. Kayser, C. Passerieux, and P. Roux; Service de Psychiatrie, Centre Hospitalier Princesse Grace, Monaco: V. Aubin, I. Cussac, J. Loftus; APHP, Groupe Hospitalo-universitaire AP-HP Nord, DMU ESPRIT, Service de psychiatrie et addictologie, Hôpital Louis Mourier, Colombes, France: C. Dubertret, N. Mazer. Service de Psychiatrie de l'adulte B, Centre Expert Trouble Bipolaire, CHU de Clermont-Ferrand, Clermont-Ferrand, France: P.M. Llorca, L. Samalin, O. Blanc. Service de Psychiatrie, Centre Expert Bipolaire, CHU de Besançon F-25030: E. Haffen, D. Bennabi.

Author for correspondence:
Raoul Belzeaux,
E-mail: raoul.belzeaux@univ-amu

CAMBRIDGE UNIVERSITY PRESS

Definition of early age at onset in bipolar disorder according to distinctive neurodevelopmental pathways: insights from the FACE-BD study

Filippo Corponi[1], Antoine Lefrere[2,3,4], Marion Leboyer[3,5,6], Frank Bellivier[3,7,8], Ophelia Godin[3,6], Josephine Loftus[3,9], Philippe Courtet[3,10,11], Caroline Dubertret[3,12,13], Emmanuel Haffen[3,14,15], Pierre Michel Llorca[3,16], Paul Roux[17,18], Mircea Polosan[3,19], Raymund Schwan[3,20], Ludovic Samalin[3,16], Emilie Olié[3,10,11], Bruno Etain[3,7,21], FACE-BD (FondaMental Academic Centers of Expertise for Bipolar Disorder) Groups, Peggy Seriès[1] and Raoul Belzeaux[22,*] (ID)

Abstract

Background. Converging evidence suggests that a subgroup of bipolar disorder (BD) with an early age at onset (AAO) may develop from aberrant neurodevelopment. However, the definition of early AAO remains unprecise. We thus tested which age cut-off for early AAO best corresponds to distinguishable neurodevelopmental pathways.

Methods. We analyzed data from the FondaMental Advanced Center of Expertise-Bipolar Disorder cohort, a naturalistic sample of 4421 patients. First, a supervised learning framework was applied in binary classification experiments using neurodevelopmental history to predict early AAO, defined either with Gaussian mixture models (GMM) clustering or with each of the different cut-offs in the range 14 to 25 years. Second, an unsupervised learning approach was used to find clusters based on neurodevelopmental factors and to examine the overlap between such data-driven groups and definitions of early AAO used for supervised learning.

Results. A young cut-off, i.e. 14 up to 16 years, induced higher separability [mean nested cross-validation test AUROC = 0.7327 (± 0.0169) for ⩽16 years]. Predictive performance deteriorated increasing the cut-off or setting early AAO with GMM. Similarly, defining early AAO below 17 years was associated with a higher degree of overlap with data-driven clusters (Normalized Mutual Information = 0.41 for ⩽17 years) relatively to other definitions.

Conclusions. Early AAO best captures distinctive neurodevelopmental patterns when defined as ⩽17 years. GMM-based definition of early AAO falls short of mapping to highly distinguishable neurodevelopmental pathways. These results should be used to improve patients' stratification in future studies of BD pathophysiology and biomarkers.

Introduction

Bipolar disorder (BD) is a severe, relapsing-remitting mental illness, taking its name from the severe fluctuations in mood, behaviors, and energy that patients experience. It is a complex and heterogenous condition, comprising multiple phenotypes, further to the type I/type II subcategories. Consequently, efforts have been made to further dissect BD into clinically meaningful subgroups underlying distinct pathophysiological pathways which would enable better stratification for diagnosis, prognosis and treatment (Coombes et al., 2020). Several clinically measurable neurodevelopmental liabilities likely contribute to the pathophysiology and etiology of a subset of patients suffering from early onset BD. However, in some instances the evidence is sparse and replication is needed (Kloiber et al., 2020). In particular, several perinatal factors, including obstetric complications, cesarean section, older paternal age, and exposure to substances and infections during pregnancy were associated with a higher risk of developing BD, supposedly through a disruption in neurodevelopment (Chudal et al., 2014; Martelon, Wilens, Anderson, Morrison, & Wozniak, 2012; Parboosing, Bao, Shen, Schaefer, & Brown, 2013), even though evidence from previous studies is inconsistent. Comorbidity with Autism spectrum disorder (ASD) and attention deficit hyperactivity disorder (ADHD) is further indication of a neurodevelopmental component to BD (Schiweck et al., 2021). Another neurodevelopmental marker is the familiar co-aggregation of BD and schizophrenia (SCZ) as well as shared genetic, and brain imaging markers (Valli, Fabbri, & Young, 2019). All this evidence suggests that a subset of patients with BD may have neurodevelopment impairment which would seem to clinically correlate with an early age at onset

(AAO) and/or the presence of psychotic symptoms. The presence of psychotic features specifically would indicate a pathophysiological proximity to SCZ and suggest a continuum between SCZ and a subset of patients with BD (Arango, Fraguas, & Parellada, 2014). A case in point is the over-representation of premorbid cognitive, motor and language impairment in SCZ as well as BD with psychotic symptoms (Murray et al., 2004). Furthermore, childhood traumatic events and cannabis misuse prior to BD onset, both well-characterized neurodevelopmental disruptors, were associated with an early onset in BD (Daruy-Filho, Brietzke, Lafer, & Grassi-Oliveira, 2011; Lagerberg et al., 2011).

While early AAO is often used as an easily measurable clinical criteria to identify a clinically homogenous BD subgroup, standing out in terms of pathogenetic pathways and disease outcomes, it remains unclear which AAO cut-off best identifies patients with BD in terms of neurodevelopmental load. Indeed, one of the reasons behind the limited replication of some of the studies investigating the link between early AAO and neurodevelopmental factors is probably the lack of consensus around the definition of early AAO. The very first studies used cut-offs for early AAO that were arbitrarily chosen, usually in the 18–25 years old (y.o.) range [e.g. (Post et al., 2010; Strober, 1992)]. A more principled and data-driven approach adopted to overcome this limitation was to find clusters of patients based on AAO by fitting the age of onset distribution using a mixture of Gaussians. A systematic review (Bolton, Warner, Harriss, Geddes, & Saunders, 2021) of studies adopting this latter approach found that the best fit was provided by three sub-groups: early [mean 17.3 years (S.D. = 1.19)], intermediate [mean 26.0 years (S.D. = 1.72)], and late [mean 41.9 years (S.D. = 6.16)] AAO group. Most BD cases, around 45%, belong to early AAO group, with 35 and 20% of patients with BD displaying intermediate and late AAO respectively. Such a distribution indicates that the early AAO group as defined above represents the majority group. Thus, some authors suggested that the term 'early onset', when used for designating an unusually low AAO (~3 S.D. from the mean of the corresponding Gaussian), should be limited to onsets under 14 years of age considering onset in the 14–21 age band is the norm in BD and therefore does not correspond to a minority subgroup of patients. Overall, the proportion of patients developing BD by the age of 14 is 5.1% based on a meta-analytical literature review (Solmi et al., 2022). However, while statistically principled and allegedly more robust than arbitrarily setting a cut-off, Gaussian mixture models (GMM) clustering only relies on the AAO distribution and, therefore, it is not guaranteed to converge to a solution where cluster membership is informative of any underlying pathophysiological process, such as the neurodevelopmental contribution to BD development. As a result, the clinical validity of an early AAO label assigned in this manner may be limited.

Existing studies investigating neurodevelopmental pathways in BD examined relatively small sample sizes, ranging from few dozens to few hundreds of patients at best (Kloiber et al., 2020); furthermore, neurodevelopmental disruptors and their correlates were mostly considered individually. In this study we capitalized on a large France-based national cohort of patients suffering from BD. Using a supervised learning approach, we aimed to determine which early AAO definition would maximize patients' separability in terms of neurodevelopmental load. Secondarily, we adopted an unsupervised learning approach and assuming, based on previous literature, that a subset of patients suffering from BD stand out from the rest in terms of neurodevelopmental load therefore constituting a separate cluster, we assessed how a data-driven two-clusters fit would stack up against each of the different early v. non-early AAO definitions.

Methods

Study population

The present study is based on the cross-sectional data set from Fonda-Mental Advanced Centres of Expertise for Bipolar Disorders (FACE-BD) cohort (Henry et al., 2015). This data set was collected from 2009 to 2020 through a French-based network of 12 BD Expert Centers (Besancon, Bordeaux, Clermont-Ferrand, Creteil, Colombes, Grenoble, Marseille, Monaco, Montpellier, Nancy, Paris and Versailles). This network was instituted by the French nonprofit foundation ' FondaMental Foundation'. Patients were referred by their psychiatrist or general practitioner for assistance in the diagnosis and management of BD. Outpatients aged 16 years or above and meeting a DSM-IV diagnosis of type I, type II, or not otherwise specified (NOS) BD met the FACE-BD cohort inclusion criteria. The same evaluation package was used by all centers and the entire assessment was performed by members of a specialized multidisciplinary team. The study was conducted in compliance with the ethical principles of medical research involving human (WMA, Declaration of Helsinki). The assessment protocol was approved by the relevant ethical review board (CPP-Ile de France IX, 18th January 2010). All data were collected anonymously. A more detailed description of the FACE-BD cohort is provided elsewhere (Henry et al., 2015).

Assessment

The Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders–Fourth Edition (DSM-IV) Axis I Disorders (SCID) (Kübler, 2013) was used to record AAO, i.e. age at first mood episode. Variables related to neurodevelopmental impairment recorded in the FACE-BD cohort and considered for this study included cannabis misuse (along with age of first cannabis misuse when applicable) and a family history of BD in first degree relatives, which were both identified as part of the SCID. Childhood trauma was recorded with the Childhood Trauma Questionnaire (CTQ) (Bernstein, Ahluvalia, Pogge, & Handelsman, 1997) and childhood symptomatology of ADHD using the Wender Utah Rating Scale (WURS) (Gift, Reimherr, Marchant, Steans, & Reimherr, 2021). Childhood history of dyslexia, dysorthographia, dyscalculia, dysphasia, dyspraxia, language delay, stuttering, gait delay, and febrile seizures was recorded by a psychiatrist and a neuropsychologist. Variables related to gestation and perinatality included new-born length, weight, and skull circumference, gestational age, cesarean section, maternal and parental age at birth, Apgar score at minute 1 and at minute 5 after birth (Apgar, 1966), and neonatal hospitalization were collected by psychiatrists based on medical interview and medical file. A complete account of all variables recorded as part of the FACE-BD cohort is presented elsewhere (Henry et al., 2015). All available variables from the FACE-BD cohort that have been previously associated with an aberrant neurodevelopment were included in this study, including variables with a known neurodevelopmental role but for which an association with BD has not been previously

investigated to the best of our knowledge [e.g. febrile seizures (Nilsson et al., 2019)]].

Machine learning strategies and data analysis

Early age at onset definitions

We set out to develop predictive models for early AAO, where the cut-off for binarizing early *v.* non-early AOO was set either by fitting a GMM to the AAO distribution or by systematically moving it between 14 and 25 y.o. We took a complete case approach, including only observations with no missing values under AAO.

We used both Bayesian information criterion (BIC) and Akaike information criterion (AIK) to compare different GMMs and to identify the optimum number of clusters in the range 2 to 10 (Raftery & Dean, 2006). The best fitting model was then used to assign each observation to the Gaussian for which the posterior probability of belonging is highest. The level of uncertainty of cluster allocation for an observation can be derived by subtracting the maximal posterior probability under a Gaussian from one. This value can be interpreted as the posterior probability of not belonging to the Gaussian maximizing the posterior. To the best of our knowledge, only one previous study (Hamshere et al., 2009) filtered out individuals based on cluster allocation uncertainty, using a threshold of 0.4. As the optimal cluster allocation uncertainty threshold cannot be determined *a priori*, we considered the effect of either enforcing no uncertainty threshold on GMM clustering or adopting progressively more stringent thresholds, namely >0.40, >0.20, and >0.10. The latter approach results in removing patients not meeting the required cluster allocation uncertainty level from further analyses, thereby expectedly creating purer and more separate clusters at the expense of decreasing the sample size.

As regards the interval 14 to 25 y.o., within which possible cut-offs were systematically explored, 14 was chosen as the lower bound since it has been previously suggested to single out 'earlier than normally expected AAO' (Bolton et al., 2021). Furthermore only 5.1% of patients with BD have an earlier onset (Solmi et al., 2022). The upper bound was informed by evidence showing that neurodevelopment ends at 25 years (Arain et al., 2013).

Supervised learning experiments

For each different early AAO definition, we trained classifiers to predict whether patients had either an 'early' or a 'non-early' AAO based on their neurodevelopmental variables. Two different classification algorithms were trained and evaluated: Elastic Net (ENET) and Extreme Gradient Boosting (XGBoost). Stratified nested cross-validation was used as re-sampling strategy. This was preferred over a single train/test data split to obtain more robust and less unbiased estimates of classification performance (Cawley & Talbot, 2010). Specifically, in the outer resampling loop, ten pairs of training/test tests were produced. In each of these outer training sets the optimum configuration of hyperparameters of the classification algorithms were selected through a random search while optimizing the area under the Receiver Operator Curve (AUROC) in a 5-fold cross-validation. The so-tuned classification algorithms were then fitted on each outer training set and their performance was evaluated on the outer test sets. AUROC was selected as performance metric due to class imbalance and to favor comparisons in future investigations since it is the typical choice in studies applying machine learning to BD. Lastly, severe class imbalance, i.e. 1:99 ratio classes ratio, in which case AUROC could be overly optimistic with respect to other metrics prioritizing identification of the minority class,

e.g. precision and recall curves (Cook & Ramadas, 2020), was excluded during exploratory data analysis (EDA).

An illustration of the supervised learning experiments' aim is reported in Fig. 1. A sketch of the analysis workflow and a fuller description of the supervised learning experiments pipeline are provided in online Supplementary Fig. S1 and Appendix.

Unsupervised learning

We tested the overlap between a data-driven grouping of the patients based on neurodevelopmental factors and the different early AAO definitions used in our supervised learning experiments. To this end, we adopted Deep Embeddings Clustering (DEC) (Xie, Girshick, & Farhadi, 2015). DEC is a deep neural network (DNN) powered approach to cluster analysis, capitalizing on DNNs' ability to learn a more clustering-friendly representation of the original data matrix due to their inherent universal function approximation properties and iteratively optimizing a clustering objective in the learned lower-dimensional feature space until a stopping criterion is met. DEC was shown to outperform vanilla k-means as well as extensions of k-means designed to operate in kernel feature space or directly on the representations learned by a DNN (Aljalbout, Golkov, Siddiqui, Strobel, & Cremers, 2018). Based on previous literature, we postulated the existence of two BD clusters ($k = 2$) with respect to neurodevelopmental load. Once cluster membership was assigned with DEC, data-driven grouping was matched against the different early AAO definitions using normalized mutual information (NMI) (McDaid, Greene, & Hurley, 2013). NMI can take values between 0 (no mutual information) and 1 (perfect correlation). In our case, when comparing data-driven clustering obtained from DEC to clinical annotations, if the two groupings substantially overlap, higher values (closer to 1) should be expected. On the flip side, lower values (closer to 0) would be indicative of little overlap between the two groupings.

Results

Study population

4421 patients have been recruited in the FACE-BD cohort since the inception of expert centers in France. 423 cases were missing values for AAO and were subsequently removed from analyses, reducing the sample size for this study to 3998. Patients had a mean age of 40.45 (s.d. = 13.06). Females constituted 60.05% (n = 2401) of the sample. Type I BD was diagnosed in 45.37% of patients (n = 1814), type II BD in 43.27% (n = 1730), while 11.35% (n = 454) had a diagnosis of BD NOS. Variables from the FACE-BD cohort considered for this study are shown in online Supplementary Table S1.

Supervised learning experiments

In order to define clinical subgroups based on AAO (a numerical variable), two different approaches were adopted: GMM clustering and dichotomization varying the cut-off in the range 14 to 25 y.o.

GMM based definition

Based on AIC and BIC model section, we identified a three-cluster solution as the best fitting admixture model (Fig. 2). In line with previous reports (Joslyn, Hawes, Hunt, & Mitchell, 2016), patients assigned to cluster 1, that is the Gaussian with the lowest mean value of AAO, were considered as 'early AAO' while patients from either cluster 2 or 3 were both labeled as 'non-early AAO'. Results from GMM clustering
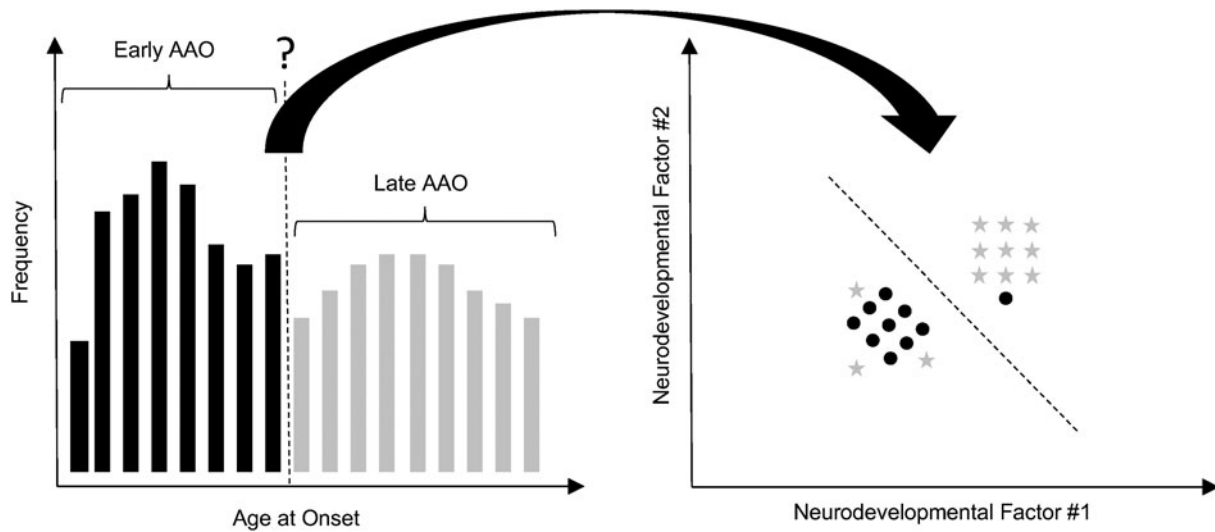
Fig. 1. Illustration of the supervised learning experiments' aim. The histogram on the left-hand side shows a sketch distribution of age at onset in the study population. It is convenient in the clinical practice to binarize patients based on a cut-off value for age at onset since accumulating evidence suggests that low age at onset is an indicator of distinctive neurodevelopmental pathways in bipolar disorder. However, the optimal cut-off value of age at onset towards mapping patients to distinctive neurodevelopmental pathways is unknown to this day. Thus, we were looking to determine the age at onset cut-off value which, upon accordingly assigning labels to patients, would maximize a classifier's ability to differentiate between the two classes, i.e. early v. non-early age at onset, using neurodevelopmental factors as predictors. For illustration purposes only two neurodevelopmental factors are depicted in the figure on the right-hand side. A complete account of neurodevelopmental variables included in this study as predictors is reported in Table 1.

and subsequently filtering cases on cluster allocation uncertainty at different thresholds are reported in Table 1. Following this approach, GMM clustering reduces to dichotomizing patients into early and non-early AAO where the cut-off between such two groups is set in a data-driven fashion based on the fit of an admixture model. As a more stringent threshold is used for filtering cases based on cluster allocation uncertainty, early and non-early AAO groups are progressively pulled apart and patients close to the boundaries are progressively dropped.

Nested cross-validation performance across learners is reported in Fig. 3, illustrating to what degree classifiers could

differentiate early v. non-early AAO based on the patients' neurodevelopmental variables. The highest average test AUROC resulted from adopting the most conservative uncertainty threshold ($\geqslant 0.1$): XGBoost achieved an average 0.6970 (S.D. = 0.0195) test AUROC with ENET close behind at 0.6935 (S.D. = 0.0151). Enforcing this threshold was equivalent to setting the maximum AAO value in the early AAO group to 20 y.o. and the minimum AAO value in the non-early AAO group to 29 y.o. Using less conservative uncertainty thresholds up to the point of no filtering drew the two AAO groups closer together increasing the maximum value allowed for defining early
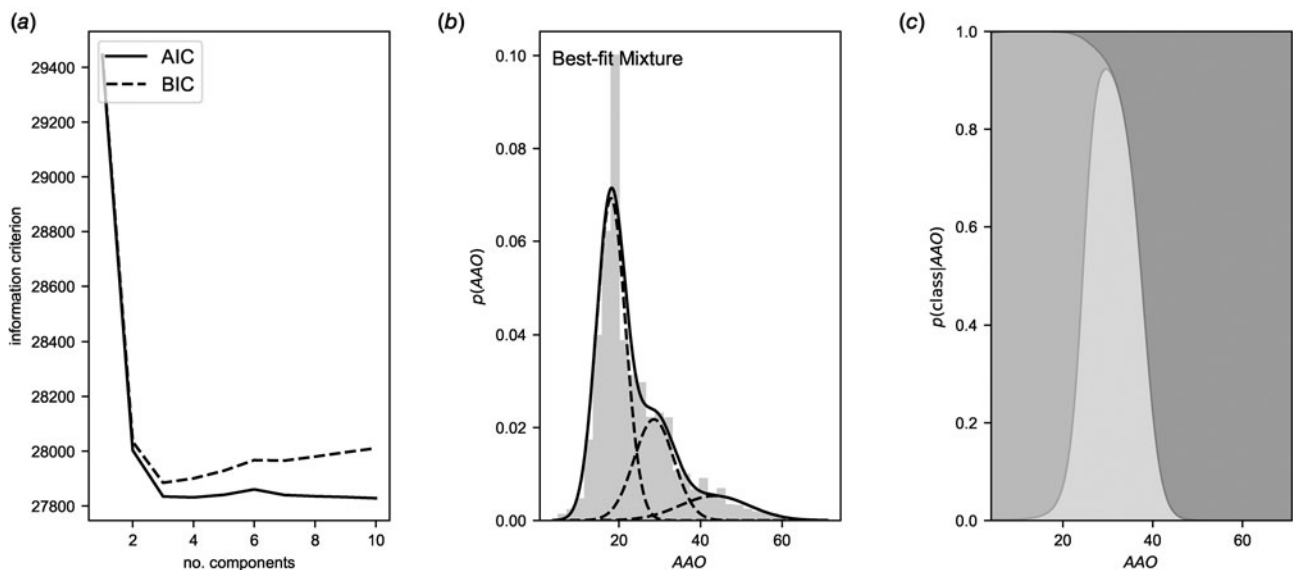

Fig. 2. Gaussian Mixture Models Clustering of age at onset. (a) Akaike information criterion and Bayesian information criterion model selection; (b) best fit; (c) posteriors under each Gaussian. AAO, age at onset.

AAO. In this manner, as the maximum AAO value in the early AAO group increased as a result of less stringent filtering, the average test AUROC across both XGBoost and ENET deteriorated.

Cut-offs in the range 14 to 25 years old

As an alternative to GMM clustering, we investigated how defining early *v.* non-early AAO patients by dichotomization with different cut-offs in the range 14 to 25 y.o. would affect separability in neurodevelopmental feature space. The proportions of early *v.* non-early AAO cases in our cohort across different age cut-offs ranged from 9.05% to 90.5% for 14 y.o. cut-off to 68.63% to 31.37% for 25 y.o. Average AUROC across outer test sets and one standard deviation are illustrated in Fig. 4. The highest average test AUROC was attained by XGBoost for the 16 y.o. cut-off (0.7327, S.D. = 0.0169). Overall, both classification algorithms reached higher (>0.70) average test AUROC values as the cut-off was kept in the range 14 to16 y.o., while for age >16 the performance showed a downward trend.

Unsupervised learning experiments

The use of an unsupervised learning approach was motivated by the lack of a ground-truth in terms of which early AAO definition best picks out distinctive neurodevelopmental pathways. Assuming the existence of two clusters in the neurodevelopmental feature space, DEC separated our cohort into 1314 (29.72%) and 3107 (70.28%) patients. NMI between DEC cluster assignments and the early AAO definitions explored in the supervised learning experiments was highest for early AAO as defined with a 17 y.o. cut-off (NMI = 0.41). NMI values for the different early *v.* non-early AAO definitions are reported in Fig. 4. Overall, NMI values across different definitions mirrored results from supervised learning experiments, showing that the use of an AAO cut-off up to 17 y.o. led to a greater overlap with the data-driven grouping by DEC. Similarly, a more conservative uncertainty threshold in GMM clustering, hence a lower maximum AAO value in the early AAO group, was associated with greater overlap with DEC clustering.

Discussion

BD may be considered as a clinical construct comprising manifold phenotypes probably underlying heterogeneous pathophysiological pathways. In this respect breaking it down into more homogeneous and biologically-grounded entities would not only improve our understanding of the disorder but also help identify clinical subgroups that might benefit from bespoke treatment and management. Based on the existing literature, one such subgroup of patients with BD is characterized by a distinctive neurodevelopmental load, relatively to other patients, and would have an early AAO as clinical trademark (Kloiber et al., 2020). In the present study, we addressed the question of which AAO cut-off would be more appropriate for differentiating BD patients based on their neurodevelopmental load using both a supervised and an unsupervised learning approach.

In the supervised learning experiments, to test the usefulness of different cut-offs towards revealing distinguishable neurodevelopmental loads, we binarized patients into early *v.* non-early AAO groups by systematically moving the cut-off in the range 14 to 25 y.o. Our results demonstrated that setting the cut-off at 16 y.o. led to the highest average AUROC value (0.7327, S.D. = 0.0169), suggesting that such

Table 1. Results from Gaussian Mixture Models Clustering

Uncertainty Threshold	Cluster 1			Cluster 2			Cluster 3		
	N (%)	AAO mean (S.D.)	AAO min–max	N (%)	AAO mean (S.D.)	AAO min–max	N (%)	AAO mean (S.D.)	AAO min–max
No filtering	2585 (64.66)	18.05 (3.38)	5–24	1024 (25.61)	29.63 (3.59)	25–37	389 (9.73)	45.79 (6.76)	38–70
>0.4	2467 (67.57)	17.76 (3.19)	5–23	826 (22.62)	30.17 (3.00)	26–36	358 (9.81)	46.46 (6.63)	39–70
>0.2	2314 (71.84)	17.42 (2.98)	5–22	632 (19.62)	30.02 (2.17)	27–34	275 (8.54)	48.51 (6.25)	41–70
>0.1	1978 (79.34)	16.71(2.64)	5–20	263 (10.55%)	29.94 (0.73)	29–31	252 (10.11)	49.19 (6.08)	42–70

Cases assigned to cluster 1 were considered as early age at onset (AAO), whereas cases assigned to either cluster 2 or 3 were considered non-early AAO. Filtering cases on cluster allocation uncertainty using progressively more stringent thresholds results in increasingly pulling apart early from non-early AAO patients at the expenses of more cases being removed from further analyses. This can be seen from the maximum AAO of Cluster 1 decreasing and the minimum AAO of Cluster 2 increasing as the uncertainty threshold for filtering becomes more conservative.
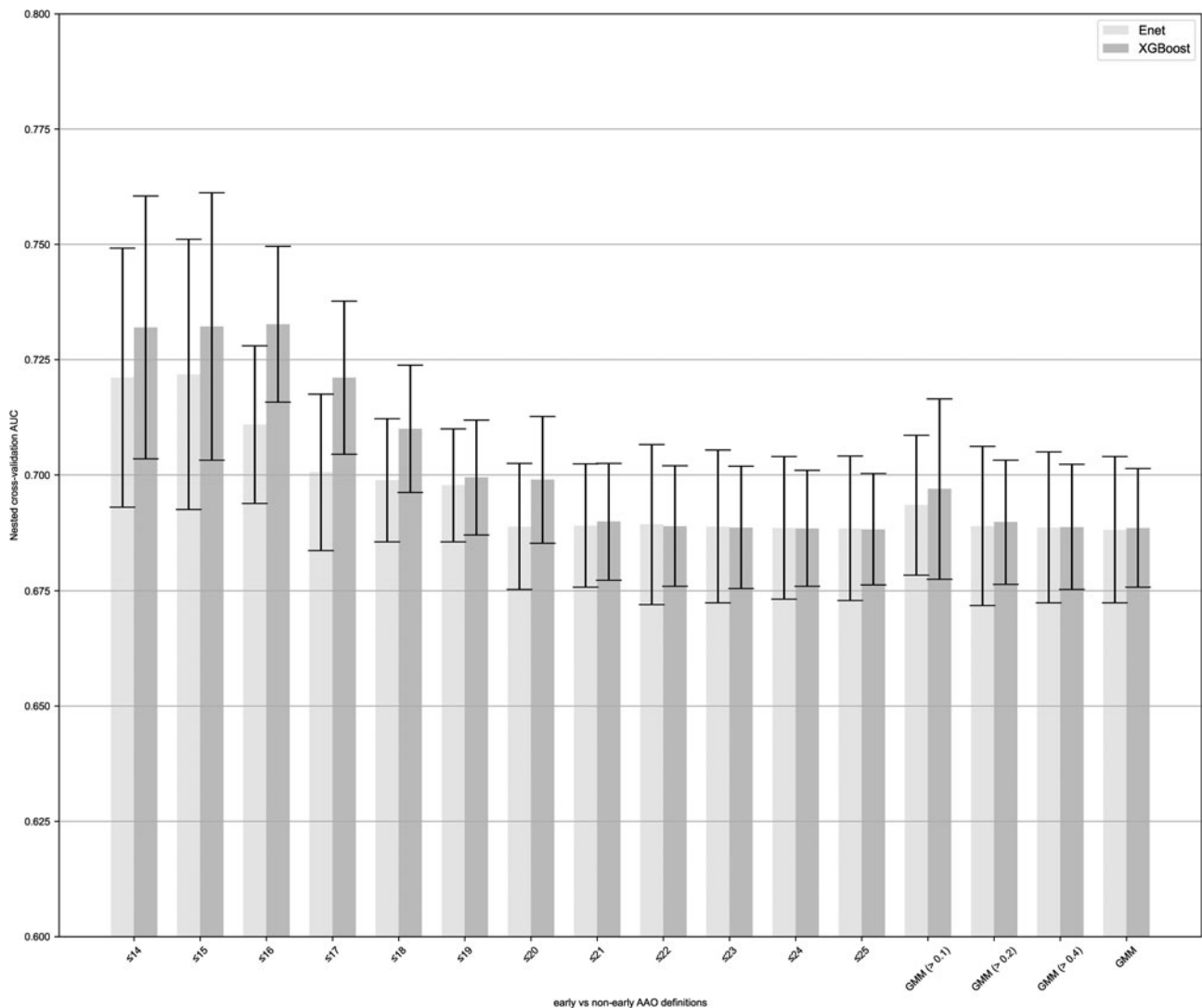
Fig. 3. Results from the supervised learning experiments. Test AUROC mean (and one standard deviation) across different early *v.* non-early age at onset (AAO) definitions is shown for the two classifiers, ENET and XGBoost. GMM, Gaussian Mixture Models Clustering, i.e. early age at onset was defined as membership to the Gaussian with lowest mean, conversely non-early age at onset was defined as membership to any of the other Gaussians. In GMM the number appearing within parentheses indicates the cluster allocation uncertainty threshold used for filtering observations.

definition of early AAO maximizes patients' separability from neurodevelopmental features. Interestingly, on the one hand, 16 y.o. was lower than any of cut-off derived with GMM analysis, even after accounting for cluster allocation uncertainty. On the other hand, this cut-off is higher than the one (i.e. 14 y.o.) previously posited as 'earlier than expected' AAO by Bolton et al. (Bolton et al. 2021). Moreover, we demonstrated that the early AAO cut-off induced by GMM resulted in lower separability based on neurodevelopmental factors. The best GMM fit in our cohort was in line with those from previous reports (Joslyn et al., 2016), placing most patients under the Gaussian with the lowest mean, therefore considering early AAO as the most common condition. Admixture analysis on AAO ultimately amounts to setting a cut-off between early *v.* non-early AAO in a data-driven fashion (where the cut-offs depend solely on the AAO distribution). The shortcomings in admixture analysis ability to find biologically meaningful, in terms of neurodevelopmental pathways, AAO cut-offs may be explained with two arguments. First, our results showed that a stronger signal, in terms of early *v.* non-early AAO cases separability from neurodevelopmental history, emerged as a result of a more

stringent filtering on cluster allocation uncertainty. Indeed, a low cluster allocation uncertainty, i.e. >0.10, lowered the maximum AAO value to 20 y.o. in the early AAO group, which corresponded to the highest average test AUROC. However, the price for such filtering was a sharp drop in sample size (by 37.64%), which limits the application of this approach in clinical research. Of notice, to the best of our knowledge, cluster allocation uncertainty filtering was implemented in only one previous study, by Bolton et (Bolton et al., 2021). Second, as argued by Montlahuc et al. (Montlahuc, Curis, Jonas, Bellivier, & Chevret, 2017), since episodes of BD are difficult to recognize in early adolescence, especially during retrospective interviews of adult patients, the AAO distribution gets left-truncated, which may potentially have a major impact on the number of detected clusters and explain the absence of an cluster in the left tail of the AAO distribution.

The need to set up multiple supervised learning experiments and explore different AAO cut-off definitions was motivated by the lack of knowledge around the best definition of 'early' AAO, conceptualized as a criterion that could point towards distinctive
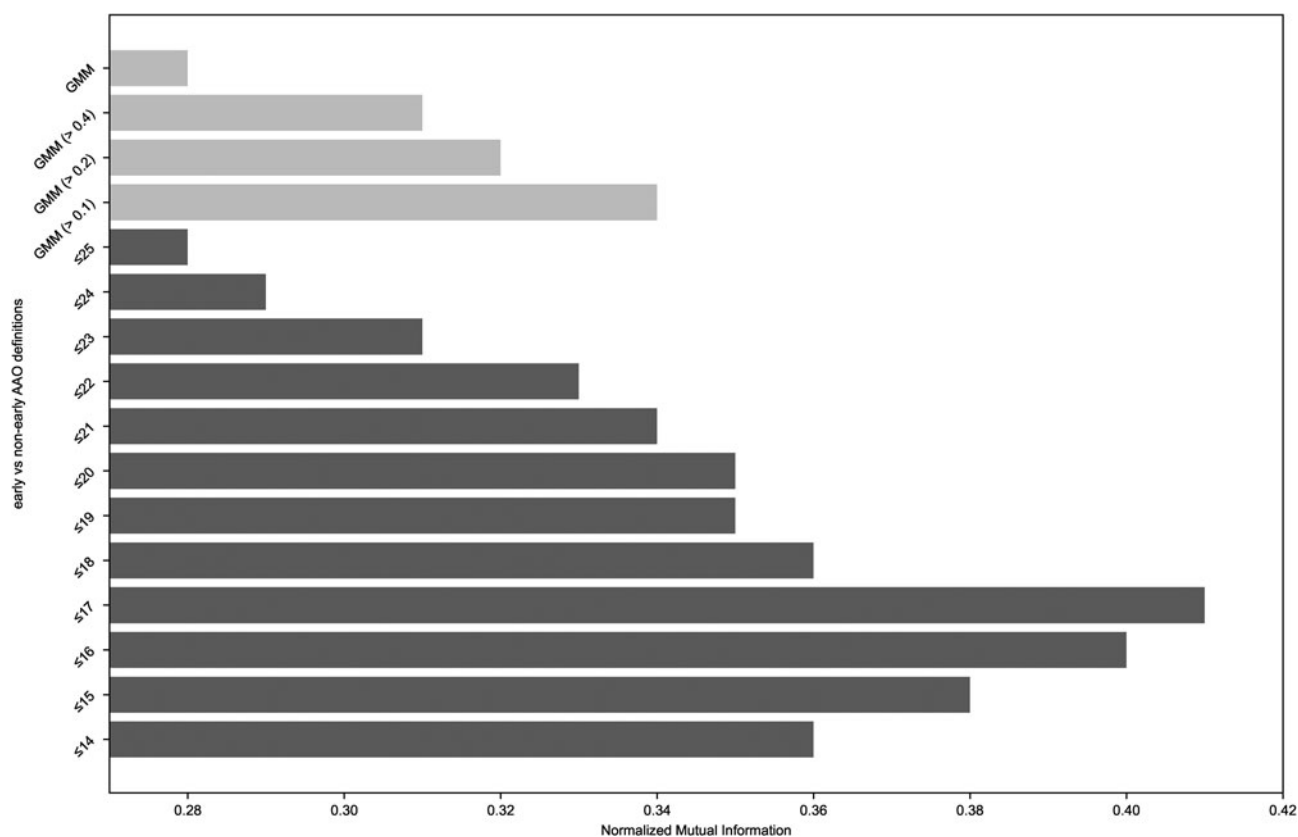
Fig. 4. Results from the unsupervised learning experiments. The degree of overlap with data-driven labels from Deep Embedding Clustering for the different early *v.* non-early age at onset definitions is shown as measured by the Normalized Mutual Information. GMM, Gaussian Mixture Models Clustering, i.e. early age at onset was defined as membership to the Gaussian with lowest mean, conversely non-early age at onset was defined as membership to any of the other Gaussians. In GMM the number appearing within parentheses indicates the cluster allocation uncertainty threshold used for filtering observations.

neurodevelopmental antecedents. As a complementary strategy, we adopted an unsupervised learning approach to examine how data-driven labels would compare against the various early AAO definitions used in the supervised learning experiments. The highest degree of overlap between data-driven clusters and the grouping induced by the different early *v.* non-early AAO definitions was recorded when lower cut-offs were used to define early AAO. The highest NMI was 0.41, achieved when early AAO was set to be ⩽17 y. o., coherently with the results from our supervised learning experiments. In other words, the unsupervised learning approach lent further support to the notion, emerging from our supervised learning experiments, that an AAO below 17 y.o. better corresponds to distinctive neurodevelopmental pathways than any of the grouping retrieved with GMM. Taken together, results from supervised and unsupervised learning experiments convergently show that the clinical label of early AAO best captures distinctive neurodevelopmental patterns when defined with a cut-off ⩽17 y.o. This result should be taken into consideration to provide better patients' stratification in future investigations of BD.

The results in this report should be balanced against several limitations. To the best of our knowledge, this is the first study using a machine learning framework to predict early AAO from neurodevelopmental antecedents. Therefore, we could not compare our results against other reports pursuing the same (or a comparable) aim. Given the post-hoc nature of this analysis, several neurodevelopment features reported in previous works as related to BD [for example perinatal infections, minor physical

anomalies (MPAs), dermatoglyphic anomalies, soft neurological signs, ASD comorbidity, and family history of SCZ (Kloiber et al., 2020)] were not available in this study. Furthermore, the very high rate (>80%) of missing values under some neurodevelopmental variables (newborn length, weight, and skull circumference, Apgar score at minute 1 and at minute 5 after birth) forced us to exclude these from the predictors. While our sample size was large relatively to most previous studies, we lacked an independent cohort for estimating out-of-sample generalization. Another important limitation we should acknowledge is the lack of biological data, such as genomics and neuroimaging. Our results indeed relied only on clinical variables collected during retrospective interviews of adult patients; these can be affected by recall bias. Furthermore, as neurodevelopmental abnormalities in BD were also described in terms of genomics [e.g. (Cao et al., 2014)] and neuroimaging [e.g. (Sarrazin et al., 2018)] findings, these modalities too should be studied for their predictive role. On the other hand, our work may help to stratify patients with more accuracy in future studies.

ANR-10-COHO-10-01.

[1]School of Informatics, University of Edinburgh, Edinburgh, UK; [2]Assistance Publique Hôpitaux de Marseille, Pôle de Psychiatrie, Marseille, France; [3]Fondation FondaMental, Créteil, France; [4]Institut de neurosciences de la Timone UMR 7289, Aix-Marseille Université & CNRS, Marseille, France; [5]Assistance publique des hôpitaux de Paris, Hôpitaux Universitaires Henri Mondor, Département Médico-Universitaire de Psychiatrie et d'Addictologie (DMU IMPACT), Fédération Hospitalo-Universitaire de Médecine de Précision en Psychiatrie (FHU ADAPT), Créteil, France; [6]Translational NeuroPsychiatry Laboratory, Université Paris Est Créteil (UPEC), INSERM U955, IMRB, Paris, France; [7]Université de Paris, INSERM UMR-S 1144, Optimisation Thérapeutique en Neuropsychopharmacologie OTeN, Paris, France; [8]Assistance publique des Hôpitaux de Paris, Groupe Hospitalo-universitaire AP-HP Nord, Hôpital Lariboisière, Département de Psychiatrie et de Médecine Addictologique, Paris, France; [9]Pôle de Psychiatrie, Centre Hospitalier Princesse Grace, France, Monaco; [10]IGF, Univ. Montpellier France, CNRS, INSERM, Montpellier, France; [11]Department of Emergency Psychiatry and Acute Care, Lapeyronie Hospital, CHU Montpellier, Montpellier, France; [12]Assistance publique des hôpitaux de Paris, Groupe Hospitalo-universitaire AP-HP Nord, DMU ESPRIT, Service de Psychiatrie et Addictologie, Hôpital Louis Mourier, Colombes, France; [13]Université de Paris, Inserm UMR1266, Sorbonne Paris Cité, Faculté de Médecine, Paris, France; [14]Service de Psychiatrie, CIC-1431 INSERM, CHU de Besançon, France; [15]UR481 Neurosciences, UFC, Besançon, France; [16]CHU Clermont-Ferrand, Department of Psychiatry, University of Clermont Auvergne, UMR 6602 Institut Pascal (IP), Clermont-Ferrand, France; [17]Centre Hospitalier de Versailles, Service Universitaire de Psychiatrie d'adulte et d'addictologie, Le Chesnay, France; [18]DisAP-DevPsy-CESP, INSERM UMR1018, Université de Versailles Saint-Quentin-En-Yvelines, Université Paris-Saclay, Villejuif, France; [19]Grenoble Alpes University, Inserm U1216 Grenoble Institute of Neuroscience, CHU Grenoble Alpes, Grenoble, France; [20]Université de Lorraine, Centre Psychothérapique de Nancy, Inserm U1254, Nancy, France; [21]Assistance publique des Hôpitaux de Paris, Groupe Hospitalo-universitaire AP-HP Nord, DMU Neurosciences, Hôpital Fernand Widal, Département de Psychiatrie et de Médecine Addictologique, Paris, France and [22]Pôle Universitaire de Psychiatrie, CHU de Montpellier, France

References

Aljalbout, E., Golkov, V., Siddiqui, Y., Strobel, M., & Cremers, D. (2018) Clustering with Deep Learning: Taxonomy and New Methods. *arXiv*:1801.07648 [cs, stat].

Apgar, V. (1966). The newborn (Apgar) scoring system. Reflections and advice. *Pediatric clinics of North America*, 13, 645–650.

Arain, M., Haque, M., Johal, L., Mathur, P., Nel, W., Rais, A., … Sharma, S. (2013). Maturation of the adolescent brain. *Neuropsychiatric Disease and Treatment*, 9, 449–461. https://doi.org/10.2147/NDT.S39776.

Arango, C., Fraguas, D., & Parellada, M. (2014). Differential neurodevelopmental trajectories in patients with early-onset bipolar and schizophrenia disorders. *Schizophrenia Bulletin*, 40(Suppl 2), S138–S146.

Bernstein, D. P., Ahluvalia, T., Pogge, D., & Handelsman, L. (1997). Validity of the childhood trauma questionnaire in an adolescent psychiatric population. *Journal of the American Academy of Child & Adolescent Psychiatry*, 36, 340–348.

Bolton, S., Warner, J., Harriss, E., Geddes, J., & Saunders, K. E. A. (2021). Bipolar disorder: Trimodal age-at-onset distribution. *Bipolar Disorders*, 23(4), 341–356. https://doi.org/10.1111/bdi.13016.

Cao, L., Deng, W., Guan, L., Yang, Z., Lin, Y., Ma, X., … Hu, X. (2014). Association of the 3′ region of the neuregulin 1 gene with bipolar I disorder in the Chinese Han population. *Journal of Affective Disorders*, 162, 81–88. https://doi.org/10.1016/j.jad.2014.03.037.

Cawley, G. C., & Talbot, N. L. C. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, 11, 2079–2107.

Chudal, R., Sourander, A., Polo-Kantola, P., Hinkka-Yli-Salomäki, S., Lehti, V., Sucksdorff, D., … Brown, A. S. (2014). Perinatal factors and the risk of bipolar disorder in Finland. *Journal of Affective Disorders*, 155, 75–80. https://doi.org/10.1016/j.jad.2013.10.026.

Cook, J., & Ramadas, V. (2020). When to consult precision-recall curves. *The Stata Journal*, 20, 131–148.

Coombes, B. J., Markota, M., Mann, J. J., Colby, C., Stahl, E., Talati, A., … Biernacka, J. M. (2020). Dissecting clinical heterogeneity of bipolar disorder using multiple polygenic risk scores. *Translational Psychiatry*, 10(1), 314. https://doi.org/10.1038/s41398-020-00996-y.

Daruy-Filho, L., Brietzke, E., Lafer, B., & Grassi-Oliveira, R. (2011). Childhood maltreatment and clinical outcomes of bipolar disorder. *Acta Psychiatrica Scandinavica*, 124, 427–434.

Gift, T. E., Reimherr, M. L., Marchant, B. K., Steans, T. A., & Reimherr, F. W. (2021). Wender Utah rating scale: Psychometrics, clinical utility and implications regarding the elements of ADHD. *Journal of Psychiatric Research*, 135, 181–188.

Hamshere, M. L., Gordon-Smith, K., Forty, L., Jones, L., Caesar, S., Fraser, C., … Smith, D. J. (2009). Age-at-onset in bipolar-I disorder: Mixture analysis of 1369 cases identifies three distinct clinical sub-groups. *Journal of Affective Disorders*, 116(1–2), 23–29. https://doi.org/10.1016/j.jad.2008.10.021.

Henry, C., Etain, B., Godin, O., Dargel, A. A., Azorin, J.-M., Gard, S., … FACE-BD group. (2015). Bipolar patients referred to specialized services of care: Not resistant but impaired by sub-syndromal symptoms. Results from the FACE-BD cohort. *The Australian and New Zealand Journal of Psychiatry*, 49(10), 898–905. https://doi.org/10.1177/0004867415585582.

Joslyn, C., Hawes, D. J., Hunt, C., & Mitchell, P. B. (2016). Is age of onset associated with severity, prognosis, and clinical features in bipolar disorder? A meta-analytic review. *Bipolar Disorders*, 18, 389–403.

Kloiber, S., Rosenblat, J. D., Husain, M. I., Ortiz, A., Berk, M., Quevedo, J., … Carvalho, A. F. (2020). Neurodevelopmental pathways in bipolar disorder. *Neuroscience and Biobehavioral Reviews*, 112, 213–226. https://doi.org/10.1016/j.neubiorev.2020.02.005.

Kübler, U. (2013). Structured Clinical Interview for DSM-IV (SCID). In M. D. Gellman & J. R. Turner (Eds.), Encyclopedia of Behavioral Medicine (pp. 1919–1920. New York, NY: Springer. https://doi.org/10.1007/978-1-4419-1005-9_66.

Lagerberg, T. V., Sundet, K., Aminoff, S. R., Berg, A. O., Ringen, P. A., Andreassen, O. A., & Melle, I. (2011). Excessive cannabis use is associated with earlier age at onset in bipolar disorder. *European Archives of Psychiatry and Clinical Neuroscience*, 261, 397–405.

Martelon, M., Wilens, T. E., Anderson, J. P., Morrison, N. R., & Wozniak, J. (2012). Are obstetrical, perinatal, and infantile difficulties associated with pediatric bipolar disorder? *Bipolar Disorders*, 14, 507–514.

McDaid, A. F., Greene, D., & Hurley, N. (2011). *Normalized Mutual Information to evaluate overlapping community finding algorithms*. https://arxiv.org/abs/1110.2515.

Montlahuc, C., Curis, E., Jonas, S. F., Bellivier, F., & Chevret, S. (2017). Age-at-onset subsets of bipolar I disorders: A critical insight into admixture analyses. *International Journal of Methods in Psychiatric Research*, 26, e1536.

Murray, R. M., Sham, P., Van Os, J., Zanelli, J., Cannon, M., & McDonald, C. (2004). A developmental model for similarities and dissimilarities between schizophrenia and bipolar disorder. *Schizophrenia Research*, 71, 405–416.

Nilsson, G., Westerlund, J., Fernell, E., Billstedt, E., Miniscalco, C., Arvidsson, T., … Gillberg, C. (2019). Neurodevelopmental problems should be considered in children with febrile seizures. *Acta Paediatrica* (Oslo, Norway: 1992), 108(8), 1507–1514. https://doi.org/10.1111/apa.14716.

Parboosing, R., Bao, Y., Shen, L., Schaefer, C. A., & Brown, A. S. (2013). Gestational influenza and bipolar disorder in adult offspring. *JAMA Psychiatry*, *70*, 677–685.

Post, R. M., Leverich, G. S., Kupka, R. W., Keck, P. E., McElroy, S. L., Altshuler, L. L., … Nolen, W. A. (2010). Early-onset bipolar disorder and treatment delay are risk factors for poor outcome in adulthood. *The Journal of Clinical Psychiatry*, *71*(7), 864–872. https://doi.org/10.4088/JCP.08m04994yel.

Raftery, A. E., & Dean, N. (2006). Variable selection for model-based clustering. *Journal of the American Statistical Association*, *101*(473), 168–178. https://doi.org/10.1198/016214506000000113.

Sarrazin, S., Cachia, A., Hozer, F., McDonald, C., Emsell, L., Cannon, D. M., … Houenou, J. (2018). Neurodevelopmental subtypes of bipolar disorder are related to cortical folding patterns: An international multicenter study. *Bipolar Disorders*, *20*(8), 721–732. https://doi.org/10.1111/bdi.12664.

Schiweck, C., Arteaga-Henriquez, G., Aichholzer, M., Edwin Thanarajah, S., Vargas-Cáceres, S., Matura, S., … Reif, A. (2021). Comorbidity of ADHD and adult bipolar disorder: A systematic review and meta-analysis. *Neuroscience and Biobehavioral Reviews*, *124*, 100–123. https://doi.org/10.1016/j.neubiorev.2021.01.017.

Solmi, M., Radua, J., Olivola, M., Croce, E., Soardo, L., Salazar de Pablo, G., … Fusar-Poli, P. (2022). Age at onset of mental disorders worldwide: Large-scale meta-analysis of 192 epidemiological studies. *Molecular Psychiatry*, *27*(1), 281–295. https://doi.org/10.1038/s41380-021-01161-7.

Strober, M. (1992). Relevance of early age-of-onset in genetic studies of bipolar affective disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *31*, 606–610.

Valli, I., Fabbri, C., & Young, A. H. (2019). Uncovering neurodevelopmental features in bipolar affective disorder. *The British Journal of Psychiatry: The Journal of Mental Science*, *215*, 383–385.

Xie, J., Girshick, R., & Farhadi, A. (2015). Unsupervised Deep Embedding for Clustering Analysis. International Machine Learning Society (IMLS) 33*rd* International Conference on Machine Learning, ICML 2016 1, 740–749.