ABSTRACT

AIM: Few personalized medicine investigations have been conducted for mental health. We aimed to generate and validate a risk tool that predicts adult Attention-Deficit/Hyperactivity Disorder (ADHD).

METHODS: A composite risk calculator was generated in a population cohort (ALSPAC – UK, 5113 participants, birth to age 17) using childhood clinical and sociodemographic data with internal validation. The risk tool was externally validated in the E-Risk cohort (UK, 2040 participants, birth to age 18), the 1993 Pelotas Birth Cohort (Brazil, 3911 participants, birth to age 18), and the MTA clinical sample (US, 476 children with ADHD and 241 controls followed for 16 years).

RESULTS: In the generating sample, the Area Under the Curve (AUC) for predicting adult ADHD was ·82 (95% confidence interval [CI], ·80 to ·83). In the UK birth cohort test sample, AUC was ·75 (95% CI, ·71 to ·78). In the Brazilian birth cohort test sample, AUC was significantly lower – ·57 (95% CI, ·54 to ·60). In the clinical trial test sample, AUC was ·76 (95% CI, ·73 to ·80). The model did not predict adult Anxiety or Major Depressive Disorder. An open-source on-line risk calculator was generated for clinical use.

CONCLUSIONS: The risk tool based on childhood characteristics specifically predicts adult ADHD in European and North-American population-based and clinical samples with comparable discrimination to commonly used clinical tools in internal medicine and higher than most previous attempts for mental and neurological disorders. However, its use in middle-income settings requires caution.

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is consistently associated with an increased risk of several adverse health and social outcomes, including poor education achievement, risky sexual behaviors and premature mortality (Cortese *et al.*, 2013, Chang *et al.*, 2014, Dalsgaard *et al.*, 2015, Faraone *et al.*, 2015). ADHD might begin in childhood and persist throughout adulthood, or it may remit spontaneously in around half of the cases (Caye *et al.*, 2016b). Recent evidence suggested that subthreshold symptoms can get worse over time, causing the emergence of a full-blown syndrome only in adulthood (Caye *et al.*, 2017), although the topic is still under debate in the literature (Cooper *et al.*, 2018, Manfro *et al.*, 2018). Although some risk factors for the persistence or emergence of adult ADHD are known (Caye *et al.*, 2016b, Caye *et al.*, 2016c), the attending psychiatrist is currently unable to correctly predict the course of the disorder based on clinical assessments of children or to propose a preventive intervention for those at risk.

One issue might be the inability to combine what is already known about risk factors. Although mental disorders arise from multiple risk factors, previous studies frequently define risk for targeted preventive interventions on the basis of a single risk factor, for instance, an affected first-degree relative or presence of subthreshold symptoms (Brent *et al.*, 2015, Taylor *et al.*, 2015, Buntrock *et al.*, 2016). Meanwhile, multivariable risk scores such as the Framingham risk score for cardiovascular disease have been one of the main frameworks for the study of preventive strategies in other areas of medicine.

Our aim was to develop and validate a multivariable risk calculator that estimates the individual risk of ADHD in late adolescence/young adulthood based on childhood characteristics. ADHD lends itself easily to the development of a risk calculator for the following reasons: First, its adverse health and social consequences are well established (Asherson *et al.*, 2016). Second, it is widely accepted that its roots are in early childhood, although some argue the full syndrome might develop later in some individuals (Moffitt *et al.*, 2015, Agnew-Blais *et al.*, 2016, Caye *et al.*, 2016a). Third, being a neurodevelopmental disorder, early intervention has the potential to change brain development and improve later clinical outcomes (Shaw *et al.*, 2006). Fourth, there is substantive evidence to support *a priori* hypotheses about specific childhood risk factors (Caye *et al.*, 2016b).

Method

Our methods follow well-established probability models in medicine and recommendations of the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement (Collins *et al.*, 2015). We developed the predictive model in one *a priori* selected sample and validated it independently in three external samples (TRIPOD analysis type 3). We selected the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort as the generating sample based on the following *a priori* defined criteria: population-based sample, largest sample.

Samples and participants

ALSPAC

The (ALSPAC) is a prospective birth cohort study in the UK. Pregnant women with expected delivery dates between April 1st, 1991 and December 31st, 1992, were invited to participate. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Further details on assessments can be found elsewhere (Boyd *et al.*, 2013). Please note that the study website contains details of all the data that is available through a fully searchable data dictionary (http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/). For the current study, we included 5113 subjects that were assessed for ADHD in childhood (age 7 or 10) and in the last available assessment (age 17).

E-Risk

The Environmental Risk (E-Risk) Longitudinal Twin Study is a prospective birth cohort study designed to represent the UK population. In 1999-2000, investigators enrolled 1116 families with same-sex 5-year-old twins (N=2232) born from January 1st, 1994 to December 4th,1995 (Moffitt and Team, 2002). The study was approved by the Joint South London and Maudsley and the Institute of Psychiatry Research Ethics Committee, and parents gave informed

written consent. Further details can be found elsewhere (Moffitt and Team, 2002). For the analyses, we included 2040 subjects with data on ADHD in childhood (ages 5, 7, 10 or 12) and in young adulthood (age 18).

Pelotas 1993

The 1993 Pelotas Birth Cohort is a prospective longitudinal birth cohort set in Brazil. In 1993, mothers of all children born in the city of Pelotas were contacted and 5249 children were enrolled. The study was approved by the institutional review board of the Federal University of Pelotas, and participants provided written informed consent. Further information on the cohort design can be found elsewhere (Goncalves *et al.*, 2014). For the current study, we included 4039 participants that had complete ADHD assessment at age 18 to 19 years old.

MTA

The Multimodal Treatment Study of Children with ADHD (MTA) is the largest clinical trial and observational follow up conducted with children with ADHD. In the first phase of the study, investigators enrolled 579 children aged 7 to 10 years old with ADHD and assigned them to 14 months of one of four groups of management. Two years after baseline, 515 consented to enter an observational follow-up and a local normative comparison group of 289 classmates (258 without ADHD) was added. Assessments were conducted at 12, 14, and 16 years after baseline. Informed consent (parental permission and child assent) was obtained for all participating families, using forms approved by both local institutional review boards and the NIH. Detailed design and methods have been presented in previous publications (1999). We included 717 subjects with any complete ADHD assessment in young adulthood (mean age 24).

Assessment and definition of the outcome variable

In each sample, the outcome was a dichotomous ADHD definition in late adolescence or young adulthood. In ALSPAC, participants' parents completed the hyperactive subscale of the Strengths and Difficulties Questionnaire (SDQ-HS) at 17 years of age. The scale showed excellent discrimination against a DSM-IV diagnosis derived from the Development and Well-Being Assessment (DAWBA) conducted in a subsample of 1673 participants (AUC = ·89, 95% CI ·81 to ·96). The best cut-off score to define diagnosis was at least 6 points on the SDQ-HS (sensitivity = 83·3%, and specificity = 83·3%). In the E-Risk, ADHD was ascertained at age 18 years using structured interviews based on full DSM-5 criteria (Agnew-Blais *et al.*, 2016). In the MTA sample, ADHD symptoms were derived from the parents´ Conners Adult ADHD Rating Scale (CAARS). At least five DSM-5 symptoms of inattention and/or hyperactivity were required for the symptom criteria. Impairment was evaluated with the Impairment Rating Scale (IRS), which has strong psychometrics and accurately identifies impairment in adults with ADHD (Sibley *et al.*, 2012). This diagnostic approach was chosen because it has better diagnostic accuracy than a semi-structured interview in this sample (Sibley *et al.*, 2017b). In the Pelotas cohort, trained psychologists interviewed the participants at 18 to 19 years old with a structured interview for ADHD based on DSM-5 criteria (Caye *et al.*, 2016a). A strict age-at-onset criterion was not required to define ADHD in young adulthood to take into account recent evidence suggesting a significant prevalence of late onset ADHD presentation (Moffitt *et al.*, 2015, Agnew-Blais *et al.*, 2016, Caye *et al.*, 2016a).

Assessment and definition of predictor variables

We selected the following predictor variables assessed in childhood: female sex, socioeconomic status (SES), mother's depression, intelligence quotient, maltreatment, ADHD symptoms, depressive symptoms, oppositional defiant behavior and conduct disorders, and single parent family. All predictors were collected before age 12, with the exception of intelligence in Pelotas, which was measured at age 18. Their selection was based on extensive review of previous reports in the literature and a meta-analysis conducted by our group (Moffitt *et al.*, 2015, Agnew-Blais *et al.*, 2016, Caye *et al.*, 2016a, Caye *et al.*, 2016b). We have included all variables that were available across the four samples with some level of comparability, without performing univariate analysis or stepwise techniques for variable selection. Definition of predictors was defined a priori according to relevant literature in the field. Further details are provided in on-line eTable 1.

Statistical analysis

When developing a predictive model in multiple samples, a recommended approach consists in selecting and tuning the best model in one *a priori* selected sample and assessing its performance fo in the remaining independent samples for external validity. Because the evaluation of internal performance within the same sample where the model was derived is affected by overfitting, internal validation optimism correction should be performed. Among the most accepted techniques for internal validation is bootstrap resampling.

We have developed the predictive model in the ALSPAC cohort. We ran a logistic regression including outcome (ADHD at last assessment) as the dependent variable and all eligible predictor variables as covariates. We inspected linearity assumptions of continuous variables by plotting the predictor and the logit of the outcome, and trough Box-Tidwell regressions. We derived the model using linear splines of equal sample sizes (with knots at 25th, 50th, and 75th percentiles) in the ADHD symptoms variable, and this model had better fit indexes (AIC, BIC). Multiple imputation with chained equations (10 imputations) using the remaining predictors was used to deal with missing values in the predictor variables. We used a fixed number of 10 iterations and assessed convergence with trace plots. In the ALSPAC cohort, for each of the 1000 bootstrap resamples, we have performed pooled regression coefficient estimates and variance across imputations with the command *mi estimate* in Stata (Rubin, 1987). We evaluated the predictive discrimination of the probability model calculating the area under the receiver operating characteristic curve (c statistic) of the estimated probability against the actual outcome as an index of model performance. We have assessed optimism of internal validation with bootstrap inference using 1000 replications with the R package *rms* (Harrell et al., 1996). We have assessed internal and external model calibration with calibration curves, plotting predicted probabilities against observed frequencies. Extreme predictions at the right end of the distribution (highest risk) including less than 1% of the sample at risk were excluded of the calibration analyses to avoid instability of the estimates, and these ranges are not shown in each graph. Multiple imputation and model generation were conducted in Stata MP 13.0. Finally, we tested the predictive discrimination of the same predictors using Machine Learning approaches with the R package *caret* (see eMethods).

We performed several sensitivity analyses to assess the robustness of our findings. We analyzed the performance (measured by the c-statistic) of the model among individuals who endorsed a very low number of ADHD symptoms at baseline (operationalized as equal or below the median of each population) in ALSPAC, E-Risk and Pelotas samples. We had also analyzed the performance (measured by the c-statistic) of the model excluding one variable at each time. Finally, we present the variation of the predicted probability within fixed levels of ADHD symptoms to assess the contribution of the remaining variables to the model.

Results

The number of participants with a dichotomous definition of adult ADHD and the frequency of childhood predictors in each sample can be found in Table 1.

Performance of the predictive model in the generating sample

All variables entered in the probabilistic model were used for the calculation of the estimated risk of the individual (Table 2). Only ADHD symptoms were corrected with splines. The predictive model discriminated between adult ADHD vs. no adult ADHD with an AUC of ·82 (Bootstrap-corrected 95% CI, ·80 to ·83, $p < .001$), which indicates very good discrimination (Figure 1). Correction for optimism with bootstrapping yielded an AUC of ·81. The calibration plot showed that predicted probability and observed frequency of adult ADHD closely agreed throughout the entire range of risk (0 to around 50% - Figure 2). The bias-corrected calibration curve was nearly identical (eFigure 1). The AUC varied within a range of ·74 to ·82 in sensitivity analyses taking out one predictor at a time (eTable 2 in Supplemental material). Proposed probability cut-offs are presented with sensitivity, specificity, positive predictive value and negative predictive value in eTable 3 in Supplemental material.

Performance of the predictive model in a validating cohort sample in the same country

In the E-Risk study, the predictive model discriminated between adult ADHD vs. no adult ADHD with an AUC of ·75 (Bootstrap-corrected 95% CI, ·71 to ·78, $p < .001$), which indicates fair discrimination (Figure 1). The

calibration plot showed reasonable agreement between predicted and observed event frequencies, especially in the lower range of risk (Figure 2). The discrimination was the same when restricting the sample to randomly selected non-siblings (eTable 4 in Supplemental material).

Performance of the predictive model in a validating sample in a middle-income country

In the Pelotas cohort, the predictive model discriminated between adult ADHD vs. no adult ADHD with an AUC of ·57 (Bootstrap-corrected 95% CI, ·55 to ·60, p < .001), which indicates poor discrimination (Figure 1). There was low agreement between estimated probability and observed frequency of the outcome (Figure 2).

Performance of the predictive model in a validating clinical sample in a country with similar income

In the MTA, the predictive model discriminated between adult ADHD vs. no adult ADHD with an AUC of ·76 (Bootstrap-corrected 95% CI, ·73 to ·80, p < .001) (Figure 1). The calibration plot showed that predicted probability and observed frequency of adult ADHD closely agreed throughout the entire range of risk (0 to around 70% - Figure 2), although the model had underestimated event frequency consistently.

Performance of the predictive model within participants with very low endorsement of ADHD symptoms in childhood

We tested the performance of the model for predicting late-onset ADHD in population samples, among only participants that endorsed few ADHD symptoms in childhood – the median or lower number of symptoms in their respective populations. The model had fair discrimination in these subgroups, except for the Pelotas sample in which the model already had poor discrimination (Table 3).

Performance of the predictive model removing one predictor at a time

We tested the model taking out one predictor at a time (eTable 2). The most relevant individual predictor was the level of ADHD symptoms in childhood. However, the model still had fair performance in the model without ADHD symptoms in childhood, with an AUC of ·74 (95% CI, ·72 to ·76, p < .001).

Variation of the predicted probability within fixed levels of ADHD symptoms

We assessed the predicted probabilities of an adult ADHD diagnosis at any fixed level of ADHD symptoms, considering maximum variation of the remaining factors (see eFigure 2). The observed variance indicates that ADHD symptoms are not the only relevant predictive factor in the model. These findings analyzed together clearly indicate that this is not a model based on just one variable.

Specificity of the predictive model in predicting ADHD

Considering that E-risk is the population cohort with the most comprehensive assessment of comorbid mental disorders, we tested model's discrimination predicting adult Anxiety and Major Depressive Disorder. The performance was significantly lower than for ADHD, showing specificity for ADHD compared to other forms of adult psychopathology (eTable 5 in Supplemental material).

Risk calculator and robustness of findings

Predictive discrimination estimates using three different machine-learning approaches were almost the same (see eTable 6 in Supplemental material). In a secondary analysis, we also have developed one comprehensive predictive model with all samples at once, using site as one more predictor variable (see eTable 7; eFigure 3). A risk calculator can be found at http://www.ufrgs.br/prodah/adhd-calculator/.

Discussion

The widespread use of tools that predict clinical outcomes in medical practice has promoted development and testing of preventive interventions, but this approach has been rarely attempted for mental health (Bitton and Gaziano, 2010). We generated a probability model to predict adult ADHD in a large birth cohort in the UK, with very good discrimination – AUC of ·81 after optimism correction – and calibration. This performance compares to

the most used clinical tools in Medicine (Morrow *et al.*, 2000). Recent attempts for mental health reported risk scores with good calibration (Fusar-Poli *et al.*, 2017, Hafeman *et al.*, 2017). These studies lacked, however, a consistent external validation with completely independent samples.

Our next step was to validate the score in independent samples. First, we tested the score in another UK birth cohort, the E-Risk. Its performance for predicting adult ADHD was similar. This is an important finding because several risk models in mental health did not replicate well even in samples from similar settings (Kivipelto *et al.*, 2006, Anstey *et al.*, 2014). Since data generated in population samples frequently do not translate to clinical samples (Weissman *et al.*, 2011), we tested the performance of the score in the MTA study, the largest clinical trial ever conducted for ADHD. As for ALSPAC and E-risk, the score worked well with good discrimination and calibration.

We then tested the score in a third birth cohort from Brazil. We observed that the score was much less accurate with an AUC of ·57. This finding is not surprising, since previous evidence suggests that the predictive discrimination of risk tools is lower in diverse sociocultural and ethnic populations (Chia *et al.*, 2015). However, since predictor factors assessment in Pelotas was the most heterogeneous, observed low discrimination might have been an effect of measurement error.

Models that predict a diagnosis of chronic disorders often include premorbid signs and symptoms of the disease as predictive factors. For example, the factor that increased discrimination the most in the recently published calculator for psychosis was the index diagnosis when presenting to secondary care, where Psychotic disorders had the greatest weight compared to other disorders such as mood disorders (Fusar-Poli *et al.*, 2017). Although this is a valid approach, other variables must also add to prediction, otherwise models would be tautological. Therefore, we also validated the score in subjects with low endorsement of ADHD symptoms in childhood. The performance was good even in this sensitivity analysis. In addition, we assessed probabilities of an adult ADHD diagnosis at any fixed level of ADHD symptoms, allowing maximum variation of the remaining factors. Finally, we checked discrimination of the model removing each factor at once. Findings suggested that although ADHD symptoms are the most important overall predictor, the complete model works as a necessary refinement and a model without ADHD symptoms has good discrimination as well.

We also conducted other secondary analyses to assess robustness of our findings. We tested the impact of using other statistical methods on our results. We observed that the discrimination of the prediction models remained stable regardless of chosen statistical methods. Finally, we tested the hypothesis of whether the score was specific for the prediction of ADHD. This is an important proof-of-concept: personalized medicine has always been a challenge for the area of psychiatry, as it has been shown consistently that most identified biomarkers and risk factors associated with one mental disorder are also associated with several others (Cross-Disorder Group of the Psychiatric Genomics *et al.*, 2013). We observed that the score was specific for ADHD, not predicting Major Depressive Disorder or Anxiety Disorders.

Previous cohort investigations included in the present study did not find significant childhood DSM dichotomous ADHD diagnosis in the trajectory of late onset ADHD (Agnew-Blais *et al.*, 2016, Caye *et al.*, 2016a). Thus, it might seem surprising that childhood ADHD symptoms predict adult ADHD. The MTA report also highlighted the importance of child ADHD subthreshold symptoms in adult ADHD in cases where formal DSM diagnosis were not found in childhood (Sibley *et al.*, 2017a). Since this approach was not the main focus of previous cohort studies (ADHD subthreshold symptoms), this might explain why childhood ADHD symptoms predict adult ADHD even in cohorts where childhood dichotomous diagnosis was not relevant for adult ADHD.

Our findings should be interpreted considering a set of limitations. First, the design and assessments of different samples were not uniform, limiting the discrimination of the score in the validating samples. Adult ADHD, for instance, was measured with a scale rather than with a structured interview in the generating sample, but not in the validating samples. It is possible, therefore, that the proposed estimated predictive discrimination in validating samples might actually be an underestimation. Further validating efforts with assessments that more closely resemble those of the generating sample might observe higher AUCs. However, this could also be seen as strength of the study, since observed discrimination indices are considered good, even with different methodologies implemented in individual studies. Second, there was attrition in the generating sample's assessments. Nevertheless, potential selection bias does not appear to affect the prediction of outcomes in this cohort, as shown in previous publications (Boyd *et al.*, 2013). Also, we have used multiple imputation techniques to deal with missing values. Third, the observed positive predictive value in selected cut-offs reaches a maximum of 61·8%, while the negative

predictive value is much higher throughout prediction. Although this might be considered insufficient, we ought to remember that the positive predictive value depends much on the prevalence of the studied condition, and we are working with population-based samples where the base rate of the condition is low. As a comparison, the Framingham risk score, that is also a tool developed in the general population, yields a positive predictive value of up to 30-40%. The risk score for Bipolar Disorder reports a positive predictive value of up to 32%, even among offspring of Bipolar patients (a high-risk sample). Fifth, it is important to note that other variables that are related to ADHD could have been part of the risk score like prematurity and ADHD in first degree relatives. However, they were not available for testing in the 4 data sets and our guide for risk factors was evidence-based guided by a previous meta-analysis (Caye *et al.*, 2016b). Accordingly, the predicted probability provided by the model should be considered an estimate probability obtained with a pre-specified set of variables.

What is the clinical utility of this score, provided that previous literature already has shown that most variables included in our model that are non-specific risk factors for mental disorders and ADHD symptoms in childhood, as expected, are key predicted risk factor for adult ADHD? No previous effort combined all these variables in a single risk calculator. Therefore, the only information that clinicians could offer was that some variables, like comorbidity with CD/ODD in childhood, increase the risk of persistence of ADHD. By using this calculator, attending clinicians can identify high-risk individuals to inform parents and guide decisions.

Thus, we propose a multivariable risk model to predict ADHD in young adulthood based on childhood factors that has good discrimination in both population and clinical settings. Clinicians can use the model to guide long-term decisions based on identification of children at high risk for future adult ADHD diagnosis. Also, it provides a framework for testing the effectiveness of preventive interventions focused on high-risk individuals. Furthermore, the score might be used to identify at-risk individuals for investigating neurobiological features including brain development. The lower discrimination observed in a middle-income country urges the discussion of how globally generalizable are the risk models that are currently being widely used in clinical practice. Indeed, even the well-established Framingham cardiovascular risk model is being subjected to criticism for its wide variation in performance across different populations. Therefore, future attempts to improve the current model should include setting-specific recalibration analyses that should then be translated to specific risk calculators to be used across different settings. Also, we suggest that cohorts use more standardized methods of collection of predictors and outcomes in Psychiatry for the study of risk factors, so that we can disentangle whether failure to replicate is due to heterogeneity of methods or population. Hence, our work adds to the need for validation of risk models in low and middle-income countries.

Acknowledgments

Funding sources: ALSPAC: The UK Medical Research Council and Wellcome (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. E-Risk: The Environmental Risk (E-Risk) Longitudinal Twin Study is funded by grant G1002190 from the United Kingdom Medical Research Council UKMRC. Additional support was provided by grant HD077482 from the National Institute of Child Health and Human Development and the Jacobs Foundation. Pelotas 1993: funded by the National Council for Scientific and Technological Development (CNPq, Brazil) and the Hospital de Clinicas de Porto Alegre (HCPA), Porto Alegre, Brazil. This article is based on data from the study "Pelotas Birth Cohort, 1993" conducted by Postgraduate Program in Epidemiology at Universidade Federal de Pelotas with the collaboration of the Brazilian Public Health Association (ABRASCO). MTA: The Multimodal Treatment Study of Children with ADHD (MTA) was an NIMH cooperative agreement randomized clinical trial, continued under an NIMH contract as a follow-up study and finally under a NIDA contract.

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses.

Contributions: All authors equally contributed for the conceptual design and planning the analyses of the current study. AC, TBM and ICP analyzed the data and the remaining authors interpreted and supervised the analyses. AC wrote the first draft, and the remaining authors revised until the final version of the manuscript was submitted. LAR coordinated the work and was the main supervisor of all the steps of this work.

Conflicts of Interest: James. M. Swanson acknowledges research support, advisory board membership, speaker´s bureau membership, and/or consulting for Alza, Richwood, Shire, Celgene, Novartis, Celltech, Gliatech, Cephalon, Watson, CIBA, UCB, Jansen-Cilag, McNeil and Eli-Lilly. Christian Kieling receives authorship royalties from ArtMed and Manole. Luis A. Rohde has received Honoraria, has been on the speakers' bureau/advisory board and/or has acted as a consultant for Eli-Lilly, Janssen-Cilag, Novartis and Shire in the last three years. He receives authorship royalties from Oxford Press and ArtMed. He also received travel awards for taking part of 2014 APA and 2015 WFADHD meetings from Shire. The ADHD and Juvenile Bipolar Disorder Outpatient Programs chaired by him received unrestricted educational and research support from the following pharmaceutical companies in the last three years: Eli-Lilly, Janssen-Cilag, Novartis, and Shire. Other authors report no conflict of interest.
2
Availability of data and material: Due to constraints on the data sharing permissions of the samples included in this study, we are not allowed to share the data for public use.

References

The MTA Cooperative Group (1999). A 14-month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder. The MTA Cooperative Group. Multimodal Treatment Study of Children with ADHD. *Arch Gen Psychiatry* **56**, 1073-86.

Agnew-Blais, J. C., Polanczyk, G. V., Danese, A., Wertz, J., Moffitt, T. E. & Arseneault, L. (2016). Evaluation of the Persistence, Remission, and Emergence of Attention-Deficit/Hyperactivity Disorder in Young Adulthood. *JAMA Psychiatry* **73**, 713-20.

Anstey, K. J., Cherbuin, N., Herath, P. M., Qiu, C., Kuller, L. H., Lopez, O. L., Wilson, R. S. & Fratiglioni, L. (2014). A self-report risk index to predict occurrence of dementia in three independent cohorts of older adults: the ANU-ADRI. *PLoS One* **9**, e86141.

Asherson, P., Buitelaar, J., Faraone, S. V. & Rohde, L. A. (2016). Adult attention-deficit hyperactivity disorder: key conceptual issues. *Lancet Psychiatry* **3**, 568-78.

Bitton, A. & Gaziano, T. A. (2010). The Framingham Heart Study's impact on global risk assessment. *Prog Cardiovasc Dis* **53**, 68-78.

Boyd, A., Golding, J., Macleod, J., Lawlor, D. A., Fraser, A., Henderson, J., Molloy, L., Ness, A., Ring, S. & Davey Smith, G. (2013). Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol* **42**, 111-27.

Brent, D. A., Brunwasser, S. M., Hollon, S. D., Weersing, V. R., Clarke, G. N., Dickerson, J. F., Beardslee, W. R., Gladstone, T. R., Porta, G., Lynch, F. L., Iyengar, S. & Garber, J. (2015). Effect of a Cognitive-Behavioral Prevention Program on Depression 6 Years After Implementation Among At-Risk Adolescents: A Randomized Clinical Trial. *JAMA Psychiatry* **72**, 1110-8.

Buntrock, C., Ebert, D. D., Lehr, D., Smit, F., Riper, H., Berking, M. & Cuijpers, P. (2016). Effect of a Web-Based Guided Self-help Intervention for Prevention of Major Depression in Adults With Subthreshold Depression: A Randomized Clinical Trial. *JAMA* **315**, 1854-63.

Caye, A., Rocha, T. B., Anselmi, L., Murray, J., Menezes, A. M., Barros, F. C., Goncalves, H., Wehrmeister, F., Jensen, C. M., Steinhausen, H. C., Swanson, J. M., Kieling, C. & Rohde, L. A. (2016a). Attention-Deficit/Hyperactivity Disorder Trajectories From Childhood to Young Adulthood: Evidence From a Birth Cohort Supporting a Late-Onset Syndrome. *JAMA Psychiatry* **73**, 705-12.

Caye, A., Sibley, M. H., Swanson, J. M. & Rohde, L. A. (2017). Late-Onset ADHD: Understanding the Evidence and Building Theoretical Frameworks. *Curr Psychiatry Rep* **19**, 106.

Caye, A., Spadini, A. V., Karam, R. G., Grevet, E. H., Rovaris, D. L., Bau, C. H., Rohde, L. A. & Kieling, C. (2016b). Predictors of persistence of ADHD into adulthood: a systematic review of the literature and meta-analysis. *Eur Child Adolesc Psychiatry*.

Caye, A., Swanson, J., Thapar, A., Sibley, M., Arseneault, L., Hechtman, L., Arnold, L. E., Niclasen, J., Moffitt, T. & Rohde, L. A. (2016c). Life Span Studies of ADHD-Conceptual Challenges and Predictors of Persistence and Outcome. *Curr Psychiatry Rep* **18**, 111.

Chang, Z., Lichtenstein, P., D'Onofrio, B. M., Sjolander, A. & Larsson, H. (2014). Serious transport accidents in adults with attention-deficit/hyperactivity disorder and the effect of medication: a population-based study. *JAMA Psychiatry* **71**, 319-25.

Chia, Y. C., Gray, S. Y., Ching, S. M., Lim, H. M. & Chinna, K. (2015). Validation of the Framingham general cardiovascular risk score in a multiethnic Asian population: a retrospective cohort study. *BMJ Open* **5**, e007324.

Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD). *Ann Intern Med* **162**, 735-6.

Cooper, M., Hammerton, G., Collishaw, S., Langley, K., Thapar, A., Dalsgaard, S., Stergiakouli, E., Tilling, K., Davey Smith, G., Maughan, B., O'Donovan, M., Thapar, A. & Riglin, L. (2018). Investigating late-onset ADHD: a population cohort investigation. *J Child Psychol Psychiatry* **59**, 1105-1113.
Cortese, S., Faraone, S. V., Bernardi, S., Wang, S. & Blanco, C. (2013). Adult attention-deficit hyperactivity disorder and obesity: epidemiological study. *Br J Psychiatry* **203**, 24-34.
Cross-Disorder Group of the Psychiatric Genomics, C., Lee, S. H., Ripke, S., Neale, B. M., Faraone, S. V., Purcell, S. M., Perlis, R. H., Mowry, B. J., Thapar, A., Goddard, M. E., Witte, J. S., Absher, D., Agartz, I., Akil, H., Amin, F., Andreassen, O. A., Anjorin, A., Anney, R., Anttila, V., Arking, D. E., Asherson, P., Azevedo, M. H., Backlund, L., Badner, J. A., Bailey, A. J., Banaschewski, T., Barchas, J. D., Barnes, M. R., Barrett, T. B., Bass, N., Battaglia, A., Bauer, M., Bayes, M., Bellivier, F., Bergen, S. E., Berrettini, W., Betancur, C., Bettecken, T., Biederman, J., Binder, E. B., Black, D. W., Blackwood, D. H., Bloss, C. S., Boehnke, M., Boomsma, D. I., Breen, G., Breuer, R., Bruggeman, R., Cormican, P., Buccola, N. G., Buitelaar, J. K., Bunney, W. E., Buxbaum, J. D., Byerley, W. F., Byrne, E. M., Caesar, S., Cahn, W., Cantor, R. M., Casas, M., Chakravarti, A., Chambert, K., Choudhury, K., Cichon, S., Cloninger, C. R., Collier, D. A., Cook, E. H., Coon, H., Cormand, B., Corvin, A., Coryell, W. H., Craig, D. W., Craig, I. W., Crosbie, J., Cuccaro, M. L., Curtis, D., Czamara, D., Datta, S., Dawson, G., Day, R., De Geus, E. J., Degenhardt, F., Djurovic, S., Donohoe, G. J., Doyle, A. E., Duan, J., Dudbridge, F., Duketis, E., Ebstein, R. P., Edenberg, H. J., Elia, J., Ennis, S., Etain, B., Fanous, A., Farmer, A. E., Ferrier, I. N., Flickinger, M., Fombonne, E., Foroud, T., Frank, J., Franke, B., Fraser, C., Freedman, R., Freimer, N. B., Freitag, C. M., Friedl, M., Frisen, L., Gallagher, L., Gejman, P. V., Georgieva, L., Gershon, E. S., Geschwind, D. H., Giegling, I., Gill, M., Gordon, S. D., Gordon-Smith, K., Green, E. K., Greenwood, T. A., Grice, D. E., Gross, M., Grozeva, D., Guan, W., Gurling, H., De Haan, L., Haines, J. L., Hakonarson, H., Hallmayer, J., Hamilton, S. P., Hamshere, M. L., Hansen, T. F., Hartmann, A. M., Hautzinger, M., Heath, A. C., Henders, A. K., Herms, S., Hickie, I. B., Hipolito, M., Hoefels, S., Holmans, P. A., Holsboer, F., Hoogendijk, W. J., Hottenga, J. J., Hultman, C. M., Hus, V., Ingason, A., Ising, M., Jamain, S., Jones, E. G., Jones, I., Jones, L., Tzeng, J. Y., Kahler, A. K., Kahn, R. S., Kandaswamy, R., Keller, M. C., Kennedy, J. L., Kenny, E., Kent, L., Kim, Y., Kirov, G. K., Klauck, S. M., Klei, L., Knowles, J. A., Kohli, M. A., Koller, D. L., Konte, B., Korszun, A., Krabbendam, L., Krasucki, R., Kuntsi, J., Kwan, P., Landen, M., Langstrom, N., Lathrop, M., Lawrence, J., Lawson, W. B., Leboyer, M., Ledbetter, D. H., Lee, P. H., Lencz, T., Lesch, K. P., Levinson, D. F., Lewis, C. M., Li, J., Lichtenstein, P., Lieberman, J. A., Lin, D. Y., Linszen, D. H., Liu, C., Lohoff, F. W., Loo, S. K., Lord, C., Lowe, J. K., Lucae, S., MacIntyre, D. J., Madden, P. A., Maestrini, E., Magnusson, P. K., Mahon, P. B., Maier, W., Malhotra, A. K., Mane, S. M., Martin, C. L., Martin, N. G., Mattheisen, M., Matthews, K., Mattingsdal, M., McCarroll, S. A., McGhee, K. A., McGough, J. J., McGrath, P. J., McGuffin, P., McInnis, M. G., McIntosh, A., McKinney, R., McLean, A. W., McMahon, F. J., McMahon, W. M., McQuillin, A., Medeiros, H., Medland, S. E., Meier, S., Melle, I., Meng, F., Meyer, J., Middeldorp, C. M., Middleton, L., Milanova, V., Miranda, A., Monaco, A. P., Montgomery, G. W., Moran, J. L., Moreno-De-Luca, D., Morken, G., Morris, D. W., Morrow, E. M., Moskvina, V., Muglia, P., Muhleisen, T. W., Muir, W. J., Muller-Myhsok, B., Murtha, M., Myers, R. M., Myin-Germeys, I., Neale, M. C., Nelson, S. F., Nievergelt, C. M., Nikolov, I., Nimgaonkar, V., Nolen, W. A., Nothen, M. M., Nurnberger, J. I., Nwulia, E. A., Nyholt, D. R., O'Dushlaine, C., Oades, R. D., Olincy, A., Oliveira, G., Olsen, L., Ophoff, R. A., Osby, U., Owen, M. J., Palotie, A., Parr, J. R., Paterson, A. D., Pato, C. N., Pato, M. T., Penninx, B. W., Pergadia, M. L., Pericak-Vance, M. A., Pickard, B. S., Pimm, J., Piven, J., Posthuma, D., Potash, J. B., Poustka, F., Propping, P., Puri, V., Quested, D. J., Quinn, E. M., Ramos-Quiroga, J. A., Rasmussen, H. B., Raychaudhuri, S., Rehnstrom, K., Reif, A., Ribases, M., Rice, J. P., Rietschel, M., Roeder, K., Roeyers, H., Rossin, L., Rothenberger, A., Rouleau, G., Ruderfer, D., Rujescu, D., Sanders, A. R., Sanders, S. J., Santangelo, S. L., Sergeant, J. A., Schachar, R., Schalling, M., Schatzberg, A. F., Scheftner, W. A., Schellenberg, G. D., Scherer, S. W., Schork, N. J., Schulze, T. G., Schumacher, J., Schwarz, M., Scolnick, E., Scott, L. J., Shi, J., Shilling, P. D., Shyn, S. I., Silverman, J. M.,

Slager, S. L., Smalley, S. L., Smit, J. H., Smith, E. N., Sonuga-Barke, E. J., St Clair, D., State, M., Steffens, M., Steinhausen, H. C., Strauss, J. S., Strohmaier, J., Stroup, T. S., Sutcliffe, J. S., Szatmari, P., Szelinger, S., Thirumalai, S., Thompson, R. C., Todorov, A. A., Tozzi, F., Treutlein, J., Uhr, M., van den Oord, E. J., Van Grootheest, G., Van Os, J., Vicente, A. M., Vieland, V. J., Vincent, J. B., Visscher, P. M., Walsh, C. A., Wassink, T. H., Watson, S. J., Weissman, M. M., Werge, T., Wienker, T. F., Wijsman, E. M., Willemsen, G., Williams, N., Willsey, A. J., Witt, S. H., Xu, W., Young, A. H., Yu, T. W., Zammit, S., Zandi, P. P., Zhang, P., Zitman, F. G., Zollner, S., Devlin, B., Kelsoe, J. R., Sklar, P., Daly, M. J., O'Donovan, M. C., Craddock, N., Sullivan, P. F., Smoller, J. W., Kendler, K. S., Wray, N. R. & International Inflammatory Bowel Disease Genetics, C. (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* **45**, 984-94.

Dalsgaard, S., Ostergaard, S. D., Leckman, J. F., Mortensen, P. B. & Pedersen, M. G. (2015). Mortality in children, adolescents, and adults with attention deficit hyperactivity disorder: a nationwide cohort study. *Lancet*.

Faraone, S. V., Asherson, P., Banaschewski, T., Biederman, J., Buitelaar, J. K., Ramos-Quiroga, J. A., Rohde, L. A., Sonuga-Barke, E. J., Tannock, R. & Franke, B. (2015). Attention-deficit/hyperactivity disorder. *Nat Rev Dis Primers* **1**, 15020.

Fusar-Poli, P., Rutigliano, G., Stahl, D., Davies, C., Bonoldi, I., Reilly, T. & McGuire, P. (2017). Development and Validation of a Clinically Based Risk Calculator for the Transdiagnostic Prediction of Psychosis. *JAMA Psychiatry* **74**, 493-500.

Goncalves, H., Assuncao, M. C., Wehrmeister, F. C., Oliveira, I. O., Barros, F. C., Victora, C. G., Hallal, P. C. & Menezes, A. M. (2014). Cohort profile update: The 1993 Pelotas (Brazil) birth cohort follow-up visits in adolescence. *Int J Epidemiol* **43**, 1082-8.

Hafeman, D. M., Merranko, J., Goldstein, T. R., Axelson, D., Goldstein, B. I., Monk, K., Hickey, M. B., Sakolsky, D., Diler, R., Iyengar, S., Brent, D. A., Kupfer, D. J., Kattan, M. W. & Birmaher, B. (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.

Harrell, F. E., Jr., Lee, K. L. & Mark, D. B. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* **15**, 361-87.

Kivipelto, M., Ngandu, T., Laatikainen, T., Winblad, B., Soininen, H. & Tuomilehto, J. (2006). Risk score for the prediction of dementia risk in 20 years among middle aged people: a longitudinal, population-based study. *Lancet Neurol* **5**, 735-41.

Manfro, A. G., Santoro, M., Polanczyk, G. V., Gadelha, A., Pan, P. M., Bressan, R. A., Brietzke, E., Talarico, F., Belangero, S., Rohde, L. A. & Salum, G. A. (2018). Heterotypic trajectories of dimensional psychopathology across the lifespan: the case of youth-onset attention deficit/hyperactivity disorder. *J Child Psychol Psychiatry*.

Moffitt, T. E., Houts, R., Asherson, P., Belsky, D. W., Corcoran, D. L., Hammerle, M., Harrington, H., Hogan, S., Meier, M. H., Polanczyk, G. V., Poulton, R., Ramrakha, S., Sugden, K., Williams, B., Rohde, L. A. & Caspi, A. (2015). Is Adult ADHD a Childhood-Onset Neurodevelopmental Disorder? Evidence From a Four-Decade Longitudinal Cohort Study. *Am J Psychiatry* **172**, 967-77.

Moffitt, T. E. & Team, E. R. S. (2002). Teen-aged mothers in contemporary Britain. *J Child Psychol Psychiatry* **43**, 727-42.

Morrow, D. A., Antman, E. M., Charlesworth, A., Cairns, R., Murphy, S. A., de Lemos, J. A., Giugliano, R. P., McCabe, C. H. & Braunwald, E. (2000). TIMI risk score for ST-elevation myocardial infarction: A convenient, bedside, clinical score for risk assessment at presentation: An intravenous nPA for treatment of infarcting myocardium early II trial substudy. *Circulation* **102**, 2031-7.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys.* Wiley: New York.

Shaw, P., Lerch, J., Greenstein, D., Sharp, W., Clasen, L., Evans, A., Giedd, J., Castellanos, F. X. & Rapoport, J. (2006). Longitudinal mapping of cortical thickness and clinical outcome in children and adolescents with attention-deficit/hyperactivity disorder. *Arch Gen Psychiatry* **63**, 540-9.

Sibley, M. H., Pelham, W. E., Molina, B. S., Gnagy, E. M., Waxmonsky, J. G., Waschbusch, D. A., Derefinko, K. J., Wymbs, B. T., Garefino, A. C., Babinski, D. E. & Kuriyan, A. B. (2012). When diagnosing ADHD in young adults emphasize informant reports, DSM items, and impairment. *J Consult Clin Psychol* **80**, 1052-61.

Sibley, M. H., Rohde, L. A., Swanson, J. M., Hechtman, L. T., Molina, B. S. G., Mitchell, J. T., Arnold, L. E., Caye, A., Kennedy, T. M., Roy, A., Stehli, A. & Multimodal Treatment Study of Children with, A. C. G. (2017a). Late-Onset ADHD Reconsidered With Comprehensive Repeated Assessments Between Ages 10 and 25. *Am J Psychiatry*, appiajp201717030298.

Sibley, M. H., Swanson, J. M., Arnold, L. E., Hechtman, L. T., Owens, E. B., Stehli, A., Abikoff, H., Hinshaw, S. P., Molina, B. S. G., Mitchell, J. T., Jensen, P. S., Howard, A. L., Lakes, K. D., Pelham, W. E. & Group, M. T. A. C. (2017b). Defining ADHD symptom persistence in adulthood: optimizing sensitivity and specificity. *J Child Psychol Psychiatry* **58**, 655-662.

Taylor, J. A., Valentine, A. Z., Sellman, E., Bransby-Adams, K., Daley, D. & Sayal, K. (2015). A qualitative process evaluation of a randomised controlled trial of a parenting intervention in community (school) settings for children at risk of attention deficit hyperactivity disorder (ADHD). *BMC Psychiatry* **15**, 290.

Weissman, M. M., Brown, A. S. & Talati, A. (2011). Translational epidemiology in psychiatry: linking population to clinical and basic sciences. *Arch Gen Psychiatry* **68**, 600-8.

Table 1. Frequency of young adulthood ADHD and of childhood predictors across the four samples

	ALSPAC (n = 5113)	E-Risk (n = 2040)	MTA (n = 717)	Pelotas (n = 4039)
Adult ADHD	486 (9·5%)	166 (8·1%)	205 (28·6%)	492 (12·2%)
Female sex	2619 (51·2%)	1071 (52·5%)	153 (21·3%)	2061 (51·0%)
Socioeconomic status				
Upper	868 (18·6%)	401 (19·7%)	136 (18·9%)	763 (19·6%)
Middle	2172 (46·4%)	966 (47·5%)	356 (50·7%)	1775 (45·6%)
Lower	1637 (35·0%)	665 (32·7%)	210 (29·9%)	1358 (34·9%)
Single parent	519 (11·8%)	450 (22·6%)	190 (26·5%)	882 (22·7%)
ODD or CD	157 (3·4%)	602 (29·5%)	304 (43·6%)	275 (7·0%)
Maltreatment				
Not detected	2084 (41·0%)	1609 (78·9%)	384 (55·3%)	2475 (67·0%)
Probable	2568 (50·5%)	312 (15·3%)	279 (40·1%)	672 (18·3%)
Severe	430 (8·5%)	119 (5·8%)	32 (4·6%)	548 (14·8%)
Lifetime Depression of the mother[a]	1850 (36·3%)	990 (48·5%)	326 (48·2%)	1881 (48·4%)
	Mean (SD)	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**
IQ	106·9 (16·3)	98·9 (15·6)	103·1 (19·5)	96·5 (12·5)
	Median (IQR)	**Median (IQR)**	**Median (IQR)**	**Median (IQR)**
Depressive Symptoms[b]	0 (1)	1 (2.5)	5·4 (6·7)	4 (4)
Number of ADHD symptoms[c]	2 (6)	1·5 (3·3)	8·3 (9·6)	4 (5)

ADHD *Attention-deficit hyperactivity disorder* ODD *Oppositional Defiant Disorder* CD *Conduct disorder* SD *Standard deviation* IQR *Interquartile range* IQ *Intelligence quotient*

a. Definition of lifetime depression of the mother was designed to be very sensitive, either by multiple assessments and/or by applying a very low threshold (further details on Table S1 of Supplementary material).

b. ALSPAC: Number of DSM-IV depressive items endorsed. E-Risk, MTA: Children's Depressive Inventory (CDI) score. Pelotas: Emotional subscale score of the SDQ.

c. ALSPAC, E-Risk, MTA: number of DSM-IV ADHD items endorsed. Pelotas: Hyperactivity subscale score of the SDQ.

Note: reported values before multiple imputation. Because each factor may have missing values, we report total number of participants and a proportion where the denominator is the total number of valid subjects.

Table 2. The probability model in the generating sample (n = 5113)

Predictors	**OR (BC 95% CI)**	**BC p-value**
Female sex	·72 (·58 - ·89)	·003
Socioeconomic status	-	-
Upper social class	*reference*	-
Middle social class	1·58 (1·15 – 2·16)	·004
Lower social class	1·55 (1·11 – 2·15)	·010
Single parent family	1·19 (·90 – 1·58)	·215
ADHD symptoms – 0-25th	3·77 (2·09 – 6·79)	< ·001
ADHD symptoms – 25-50th	1·19 (1·02 – 1·40)	·031
ADHD symptoms – 50-75th	1·13 (1·05 – 1·22)	·001
ADHD symptoms – 75-100th	1·18 (1·12 – 1·25)	< ·001
ODD or CD	1·81 (1·21 – 2·71)	·004
Childhood maltreatment	-	-
No detected maltreatment	*reference*	-
Probable maltreatment	1·28 (1·01 – 1·64)	·045
Severe maltreatment	1·35 (·93 – 1·95)	·115
Depression of the mother	1·41 (1·13 – 1·75)	·002
Intelligence quotient[a]	·89 (·85 - ·95)	< ·001
Depressive symptoms (z-score)[b]	1·00 (·92 – 1·10)	·940

OR *Odds Ratio;* ODD *Oppositional Defiant Disorder;* CD *Conduct Disorder;* ADHD *Attention-deficit hyperactivity disorder*

BC *Bootstrap-corrected*

a. We report the OR for a 10-point change in the intelligence quotient scale.

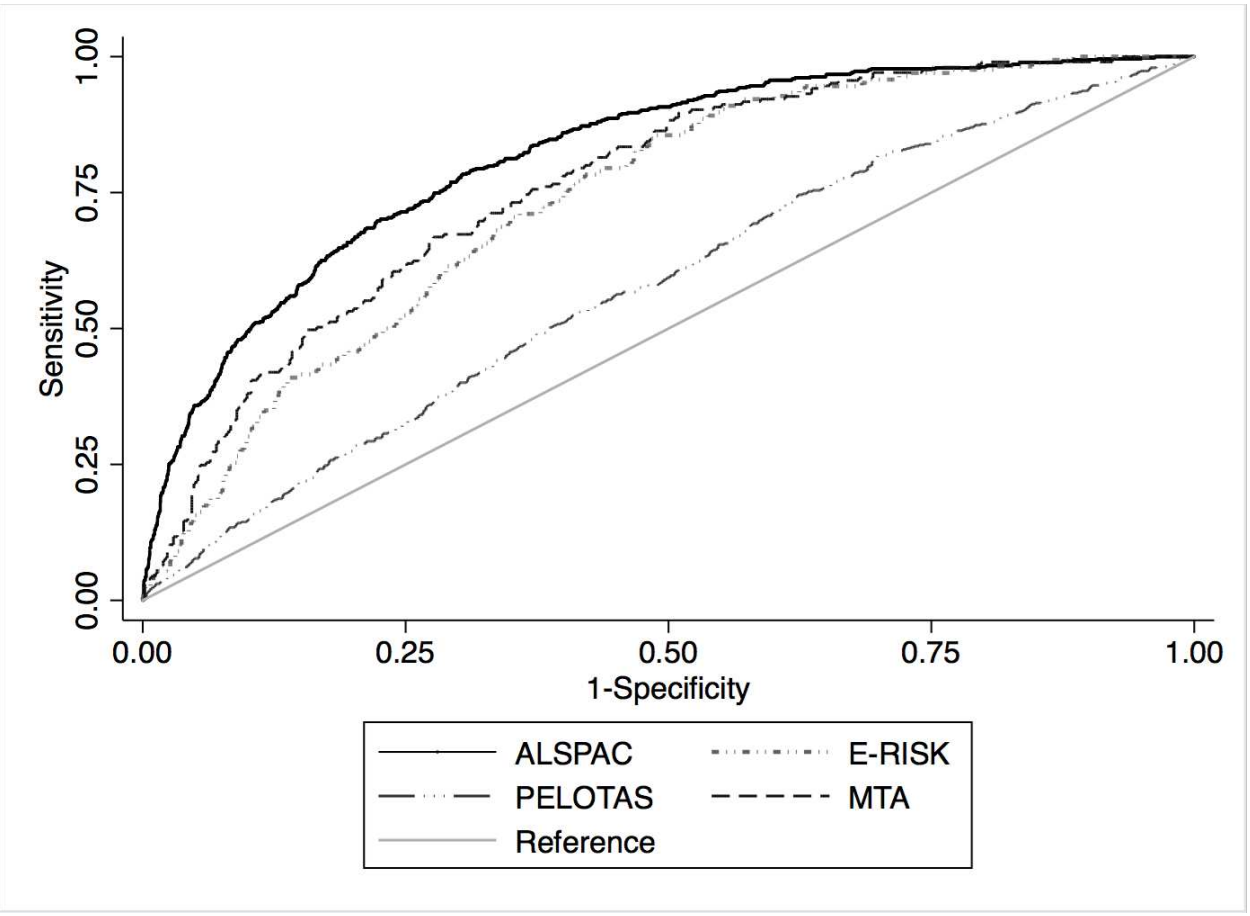
b. Due to the OR of 1·00 for depressive symptoms, we have omitted this variable from the on-line calculator.

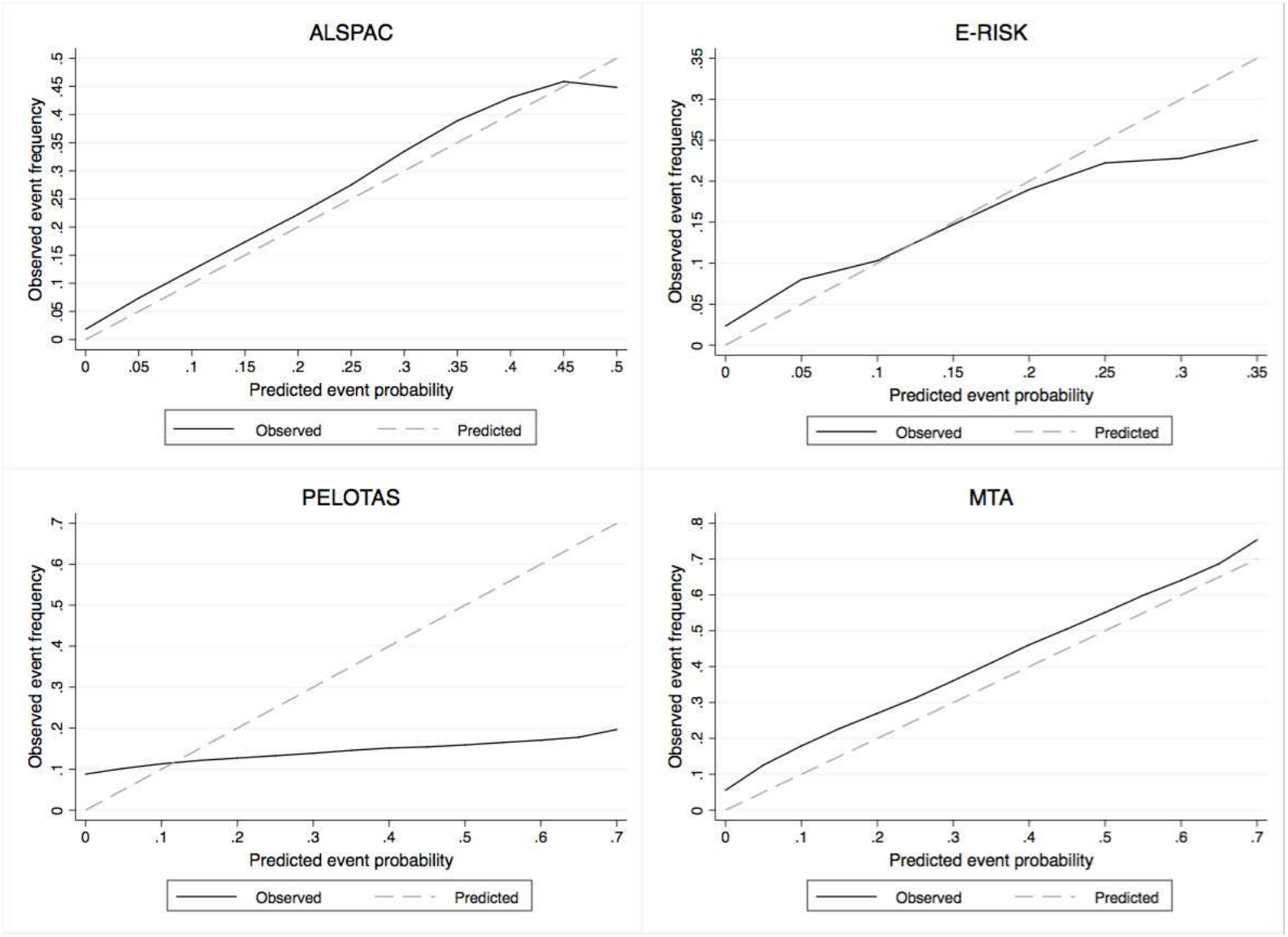
Table 3. Performance of the score for individuals with very low ADHD childhood symptoms.

	AUC	BC 95% CI	BC p-value
ALSPAC (n = 2688)	·77	·72 – ·82	< ·001
E-Risk (n = 1099)	·78	·71 - ·86	< ·001
Pelotas (n = 2135)	·56	·52 - ·60	< ·001

BC Bootstrap-corrected

ROC analyses were done only in participants with low endorsement of ADHD symptoms in childhood. Low endorsement was defined as median number of symptoms or below the median of their respective population (ALSPAC: 2 or less ADHD symptoms; E-Risk: 1 or 0 ADHD symptoms; Pelotas: the median or less than median (4) in the hyperactivity subscale of the SDQ).


Sensitivity
0.00
0.25
0.50
0.75
1.00
1-Specificity
0.00
0.25
0.50
0.75
1.00
ALSPAC
E-RISK
PELOTAS
MTA
Reference


ALSPAC
Observed event frequency
0 .05 .1 .15 .2 .25 .3 .35 .4 .45 .5
Predicted event probability
0 .05 .1 .15 .2 .25 .3 .35 .4 .45 .5
Observed
Predicted
E-RISK
Observed event frequency
0 .05 .1 .15 .2 .25 .3 .35
Predicted event probability
0 .05 .1 .15 .2 .25 .3 .35
Observed
Predicted
PELOTAS
Observed event frequency
0 .1 .2 .3 .4 .5 .6 .7
Predicted event probability
0 .1 .2 .3 .4 .5 .6 .7
Observed
Predicted
MTA
Observed event frequency
0 .1 .2 .3 .4 .5 .6 .7 .8
Predicted event probability
0 .1 .2 .3 .4 .5 .6 .7
Observed
Predicted

Supplemental material of "A risk calculator to predict adult Attention-deficit/Hyperactivity disorder: generation and external validation in three birth cohorts and one clinical sample".

Contents:

Arthur Caye, MD; Jessica Agnew-Blais, PhD; Louise Arseneault, PhD; Helen Gonçalves, PhD; Christian Kieling, MD PhD; Kate Langley, PhD; Ana M. B. Menezes, PhD; Terrie E. Moffitt, PhD; Ives Cavalcante-Passos, MD PhD; Thiago Botter-Maio Rocha, MD MSc; Margaret Sibley, PhD; James M. Swanson, MD, PhD; Anita Thapar, MD PhD; Fernando Wehrmeister, PhD; Luis Augusto Rohde, MD PhD

eTable 1. Assessment of predictor variables.

	ALSPAC	E-Risk	MTA	Pelotas
Intelligence quotient	Weschler Intelligence Scale for Children version III, age 8	Weschler Intelligence Scale for Children III-R, Age 12	Weschler Intelligence Scale for Children version III, ages 7 to 10	Weschler Adult Intelligence Scale, age 18
ODD or CD	DSM-IV criteria, age 10	DSM-IV criteria, ages 5 to 12 (or rule)	DSM-IV criteria, ages 7 to 12 (or rule)	SDQ-C >= 7, age 11.[a]
Depressive symptoms	DSM-IV (DAWBA), z-score of symptoms, age 10	CDI, z-score, age 12	CDI, z-score of mean reported symptoms, ages 7 to 12[a]	SDQ-E rated by parents, z-score, age 11
ADHD symptoms	DSM-IV (DAWBA) rated by parents, number of symptoms, age 10	DSM-IV rated by parents, z-score of number of reported symptoms, age 12	DSM-IV (DISC-IV), ages 7 to 12, z-score of mean reported symptoms[a], ages 7 to 12	SDQ-H rated by parents, z-score, age 11
Childhood maltreatment	Physical, emotional or sexual abuse and maladaptive parenting according to previous definitions (Lereya *et al.*, 2015). None, probable or severe if neither, one or both were present (ages 18 months to 7 years)	None, probable or severe according to previous definitions (Caspi *et al.*, 2003), children's ages 5 to 12.	Parent-Child Relationship Scale answered by parents ages 7 to 12, grouped into none, probable or severe.	None, probable or severe according to previous definitions (Caspi *et al.*, 2003). Asked retrospectively at age 15.
Depression of the mother	Positive if any of the following true: self-reported having had severe depression (age 11); self-reported having taken pills for depression in the last three years (age 9); EDPS of at least 10 (ages 8 months, 18 months).	DSM-IV, children's ages 5 to 12.	Positive if biological mom retrospectively reported having the blues at or after delivery, asked at baseline (children's age 7 to 10).	At least 7 points in the SRQ-20, as previously suggested (WHO, 1994), age 11.
Social class[c]	Registrar's General Classification, asked at birth.	Acorn classification, ages 5 to 12.	Gross household income in US$, age 7 to 10, ages 7 to 10.	ABEP criteria, age 11

CDI *Children's Depression Inventory* | SDQ-C *Strengths and Difficulties Questionnaire, conduct subscale* | SDQ-E *Strengths and Difficulties Questionnaire, emotional subscale* | SDQ-H *Strengths and Difficulties Questionnaire, hyperactivity subscale* | EDPS - *Edinburgh Postnatal Depression Scale* | SRQ-20 *Self-reporting Questionnaire 20* | ABEP *Associação Brasileira de Empresas de Pesquisa (the Brazilian Association of Research Companies)*

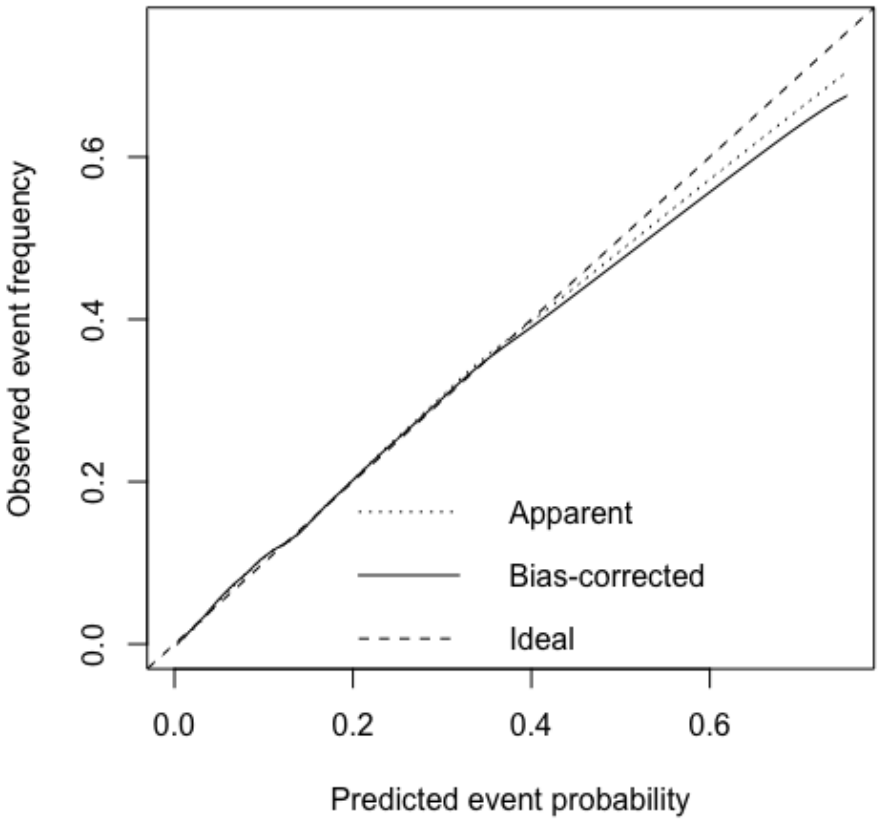
a. We ran a ROC curve analyses of the SDQ-C rated by parents against DAWBA in a subsample of children (n = 290). Discrimination was fair (0.77). The best selected cut-off was at least 7 points.
b. Since categories of risk were heterogeneous across studies, we have decided to group categories aiming to achieve similar percentages of the population included in each group. Therefore, the observed effect reflects rather the *relative* social class (how the individual family compares to the population) than the *absolute* concept (how much does the family actually earn or possess).

eMethods: Machine Learning approaches

We compared the logistic regression analysis to some well-established machine learning algorithms in order to assess the consistency of our findings. We used the package caret (Version 6.0-73) from R software (https://www.R-project.org/). We selected the caret package due to its automated tuning methods for machine learning algorithms, which enable the selection of the best fit for each model. We assumed a diagnostic classification problem with the abovementioned predictors used as input data. The main objective was to train a set of machine learning algorithms to estimate the probability of a subject belonging to either ADHD or healthy control groups given previously unseen subjects' data. In the present analysis, besides logistic regression, Random Forest, Artificial Neural Network, and Stochastic Gradient Boosting were used because 1) they are capable of modeling more complex patterns than nearly any algorithm; 2) they can handle categorical or continuous features; 3) they can be used on data with extremely large number of observations; 4) they can be used to classification prediction problems. Fforest (or decision tree forests) is an ensemble-based method that focuses only on ensembles of decision trees. This method was developed by Leo Breiman and Adele Cutler, and combines the base principles of "bagging" with random feature selection to add additional diversity to the decision tree models. An Artificial Neural Network (ANN) models the relationship between a set of input signals and an output signal using a model derived from our understanding of how a biological brain responds to stimuli from sensory inputs. Stochastic gradient boosting is another "bagging" procedure.[5]Machine learning approach was conducted in two phases: 1) training and validation phase and 2) test phase. In the first phase, we used the ALSPAC dataset to train, to validate, and to identify the best fit (parameter tuning) for each model. The parameters to be adjusted were 1) size and decay for ANN, 2) mtry (an optional integer specifying the number of features to randomly select at each split) for random forest, and 3) n.trees, interactions.depth, shrinkage, and n.minobsinnode for Stochastic Gradient Boosting. We used optimism bootstrapping (n=1000) as the resampling method and AUC to select the best fit for each model. In the second phase, we tested the selected models in E-Risk, MTA, and Pelotas datasets.

The parameters selected during the first phase were 1) mtry=3 for random forest, 2) size=1 and decay=0.01 for ANN, 3) n.trees=150, interactions.depth=1, shirinkage=0.1, n.minobsinnode=10 for Stochastic Gradient Boosting. eTable6 shows the AUC for each model in all test datasets.

eFigure 1. Bias-corrected calibration plot for internal validation in the ALSPAC cohort.



eTable 2. Predictive discrimination of the score leaving out predictors one at a time.

	AUC	95% CI	p-value
Full model	·82	·79 - ·84	< ·001
- ADHD symptoms	·74	·72 - ·76	< ·001
- Gender	·81	·80 - ·83	< ·001
- Social class	·82	·80 - ·83	< ·001
- Single parent	·82	·80 - ·83	< ·001
- ODD/CD	·81	·80 - ·83	<. ·001
- Childhood maltreatment	·82	·80 - ·83	< ·001
- Depressive symptoms	·82	·79 - ·83	< ·001
- Mother's depression	·81	·79 - ·84	< ·001
- IQ	·81	·79 - ·83	< ·001

AUC *Area under the Curve* | CI *Confidence Interval* | ADHD *Attention-deficit/Hyperactivity Disorder* | ODD *Oppositional Defiant Disorder* | CD *Conduct Disorder*

eTable 3. Sensitivity, specificity, positive and negative predictive values at selected risk cut-offs.

	Sensitivity	Specificity	PPV	NPV
Probability >= 10%	72·4%	74·3%	22·9%	96·3%
Probability >= 20%	45·1%	91·0%	34·4%	94·0%
Probability >= 30%	30·3%	96·0%	44·5%	92·9%
Probability >= 40%	20·4%	98·1%	52·4%	92·1%
Probability >= 50%	11·1%	99·0%	54·5%	91·4%
Probability >= 60%	6·0%	99·6%	61·8%	91·0%

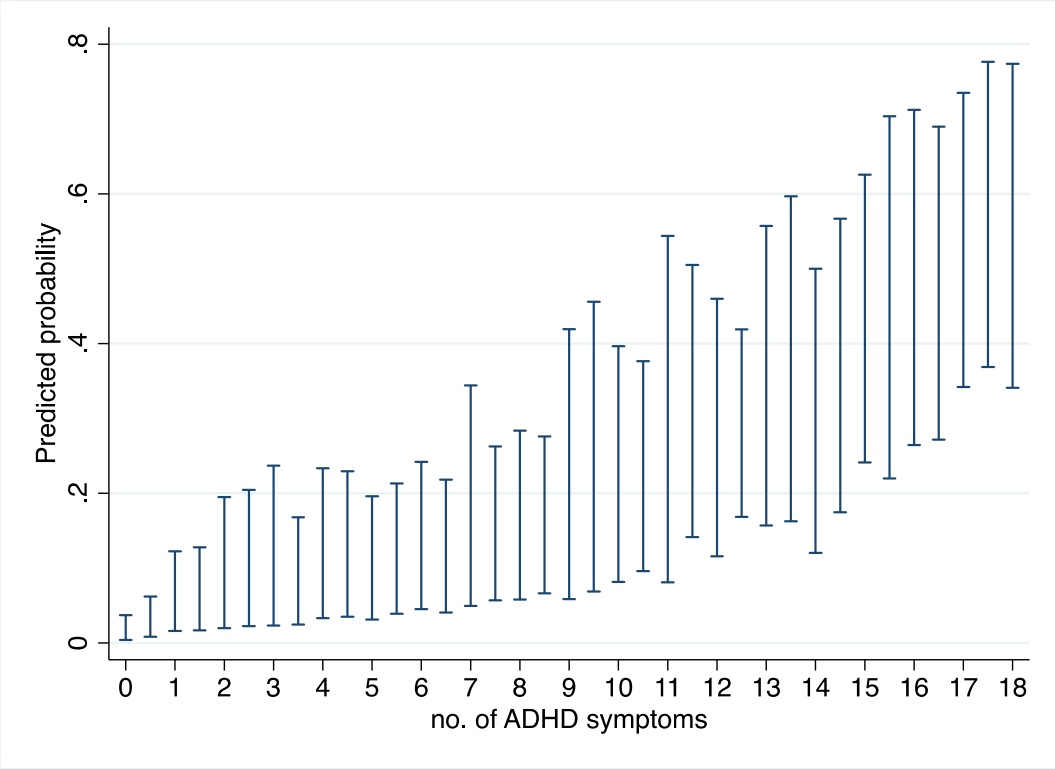
PPV+ *Positive Predictive Value* NPV *Negative Predictive Value*

eTable 4. The assessment of the confounding effect of twin pairs in the E-Risk.

	AUC	95% CI	p-value
Random non-siblings 1 (n = 1020)	·75	·70 - ·80	< 0·001
Random non-siblings 2 (n = 1020)	·75	·70 - ·80	< 0·001

AUC *Area under the Curve* | CI *Confidence Interval* Note: We tested the risk score in the E-Risk sample in subgroups of randomly selected non-sibling participants.

eFigure 2: Variation of predicted probabilities within fixed levels of ADHD symptoms



eTable 5. Performance of the score for Major Depression Disorder or Anxiety Disorders in young adulthood[1].

	AUC	95% CI	p-value	vs. ADHD (p-value)[2]
Anxiety Disorders	·52	·47 - ·59	·72	< 0·001
Major Depressive Disorder	·56	·52 - ·59	·001	< 0·001
Alcohol Use Disorder	.58	.54 - .62	< .001	< .001
Marijuana Use Disorder	.67	.60 - .73	< .001	.03

AUC *Area under the Curve* | CI *Confidence Interval* | ADHD *Attention-deficit/Hyperactivity Disorder*

1. Tested in the E-Risk sample.
2. Tested against the performance of the score for predicting ADHD in the E-Risk cohort

eTable 6. Performance of the predictive model using Machine Learning approaches.

	Area Under the Curve (95% Confidence Interval)			
	ALSPAC	E-Risk	MTA	Pelotas
Logistic Regression (original)	·82 (·80 - ·83)	·75 (·71 - ·78)	·76 (·73 - ·80)	·57 (·54 - ·60)
Random Forest	·80 (·74 - ·87)	·70 (·67 - ·74)	·72 (·68 – ·76)	·56 (·53 - ·59)
Stochastic Gradient Boosting	·81 (·78 - ·83)	·74 (·71 – ·77)	·76 (·72 - ·79)	·57 (·55 – ·60)
Artificial Neural Network	·81 (·77 - ·85)	·74 (·70 - ·77)	·76 (·72 - ·80)	·58 (·55 - ·61)

eTable 7. A comprehensive model including all samples

We have developed an alternative approach using all data at once, including site as a tenth predictor.

Predictor	OR (BC 95% CI)	BC p-value
Female sex	1·06 (·94 – 1·21)	·349
Social class	-	-
Higher	*reference*	
Middle	1·04 (·87– 1·24)	·639
Lower	1·02 (·84 – 1·23)	·850
Single parent family	1·09 (·94 – 1·26)	·262
Childhood maltreatment	-	-
None	*reference*	-
Probable	1·39 (1·21 – 1·59)	< ·001
Severe	1·98 (1·65 – 2·37)	< ·001
Oppositional Defiant Disorder or Conduct Disorder	1·14 (·97 – 1·34)	·104
Mother's depression	1·12 (·99 – 1·28)	·083
ADHD symptoms – 0-25th	2·19 (1·55 – 3·11)	< ·001
ADHD symptoms – 25-50th	1·21 (1·09 – 1·34)	< .001
ADHD symptoms – 50-75th	1·11 (1·06 – 1.16)	< .001
ADHD symptoms – 75-100th	1.07 (1.04 – 1·10)	< ·001
Intelligence quotient[a]	·91 (·88 – 1·10)	< ·001
Depressive symptoms	1·04 (·98 – 1·10)	·180
Site		
ALSPAC	*reference*	-
E-Risk	·99 (·80 – 1·22)	·944
Pelotas	·71 (·59 - ·85)	< ·001
MTA	2·35 (1·89 – 2·92)	< ·001

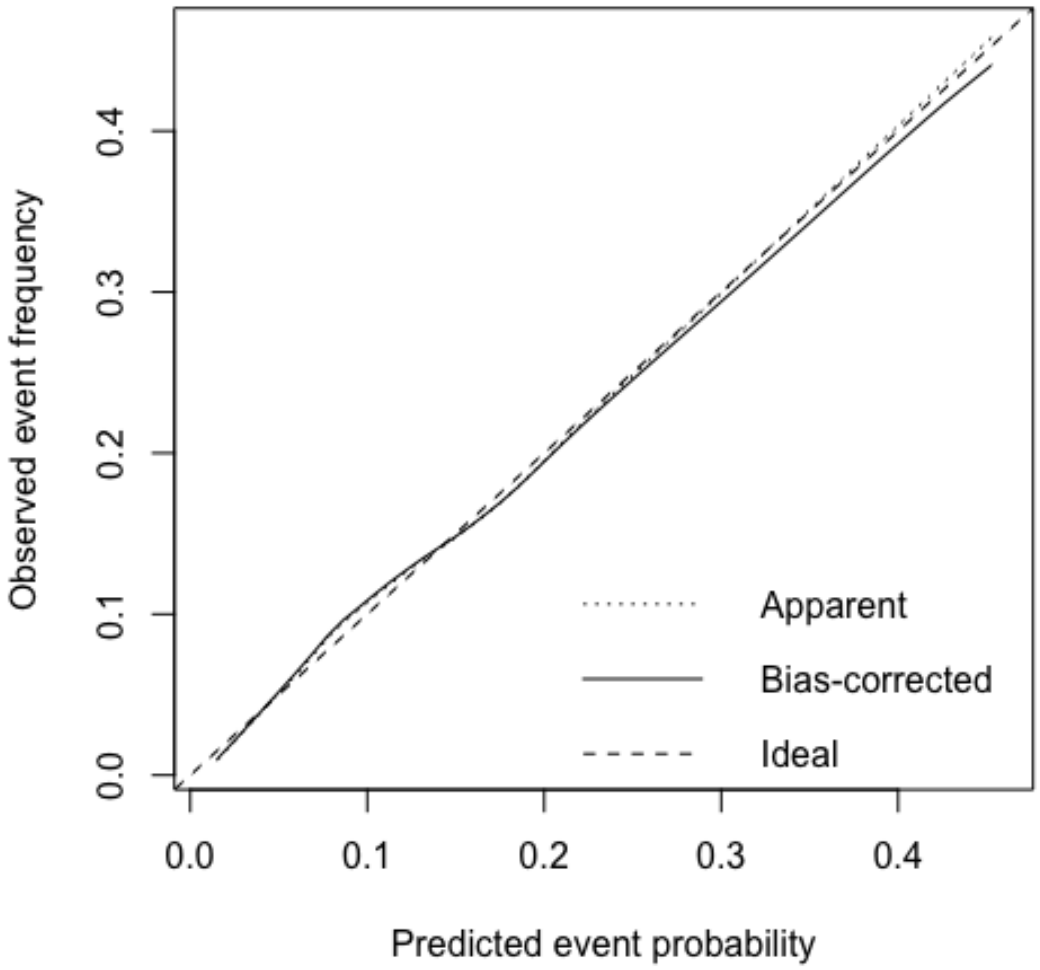
BC Bootstrap corrected

a. We report the OR for a 10-point change in the intelligence quotient scale.

Overall Area Under the Curve: ·74 (·73 - ·76), p < ·001 (Bootstrap optimism-corrected: ·73)

We also ran a comprehensive model including all two-way interactions between site and predictor variables, with an AUC of ·78 (available upon request).

eFigure 3. Bias-corrected calibration plot for internal validation in the comprehensive model.



Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., McClay, J., Mill, J., Martin, J., Braithwaite, A. & Poulton, R. (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science* **301**, 386-9.
Lereya, S. T., Copeland, W. E., Costello, E. J. & Wolke, D. (2015). Adult mental health consequences of peer bullying and maltreatment in childhood: two cohorts in two countries. *Lancet Psychiatry* **2**, 524-31.
WHO (1994). A user's guide to self-reporting questionnaires. . WHO: Division of mental health: Geneva.