Prediction of academic achievement in adolescents from teacher reports of inattention in childhood - a methodological pattern classification study

Astri J. Lundervold [1,2] *Tormod Bøe* [3] *Arvid Lundervold* [4,5]

[1]Department of Biological and Medical Psychology, University of Bergen, Jonas Lies vei 91, 5009 Bergen, Norway &

[2] K.G. Jebsen Center for Research on Neuropsychiatric Disorders, University of Bergen, Bergen, Norway

[3]Regional Centre for Child and Youth Mental Health and Child Welfare, Uni Health, Uni Research, Bergen, Norway

[4] Neuroinfromatics and Image Analysis Laboratory Department of Biomedicine, University of Bergen

[5] Department of Radiology, Haukeland University Hospital, Bergen

11-February-2016

Bergen

Aiming for *Psychological Methods*

`http://www.apa.org/pubs/journals/met`

Abstract

Background Primary school teachers commonly report behaviour in their pupils that are part of what we defined within the concept of depression, behavior that is associated with academic achievement obtained in highschool. The aim of the present study is to investigate this association by including statistical analyses dealing with studies where both predictors and the outcome variable are at a categorical level.

Methods Inattention in a sample 2397 individuals were rated by their primary school teachers when they participated in the first wave of the Bergen Child Study (BCS) (7 - 9 years old), and their academic achievements were available from an official school register when attending high-school (16 - 19 years old). Inattention was assessed by the nine items rated at a categorical leve, and the academic achievement scores were divided into three parts including a similar number of participants.

Results Boys obtained higher inattention scores and lower academic scores than girls. Inattention problems related to sustained attention and distractibility turned out to have the highest predictive value of academic achievement level across all selected statistical analyses, and the full model showed that inattention explained about 10% of the variance in high school scores about 10 years later. . A high odds-ration of being allocated to the lowest academic achievement category was shown by a multinominal regression analysis, while a pattern of problems related to sustained attention and distractibility was revealed by generating classification trees. By including recursive learning algorithms, the most successful classification was found between the inattention items and the highest level of achievement scores.

Summary The present study showed the importance of a pattern of early problems related to sustained attention and distractibility in predicting future academic results. By including different statistical classification models we showed that this pattern was fairly consistent. Furthermore, calculation of classification errors gave information about the uncertainty when predicting the outcome for individual children. Further studies should include a wider range of variables.

1 INTRODUCTION

Questionnaire data with few response categories are frequently used in psychological research. Using such data in a longitudinal design, we are typically interested in relating the patterns of individual response items to a given outcome. In this work we have explored a collection of pattern analysis and machine learning methods to study the task of predicting academic achievements in adolescents from teacher report of inattention in childhood.

Inattention in early childhood has been associated with a wide range of behavioral and social problems [Bellanti and Bierman, 2000, Connors et al., 2012], with a fairly good documentation of a close link to poor academic achievement [Polderman et al., 2010, Pingault et al., 2014, Garner et al., 2013, Holmberg and Bolte, 2014, Gray et al., 2014]. Early detection of inattention by teachers of primary school children is therefore of great importance. Their observation of classroom behaviour is not only known to give reliable information about problems related to inattention, but their report is also shown to be a valuable predictor of later academic achievement [Garner et al., 2013]. Assessment of inattention is, however, not a simple task. As a multidimensional concept, inattention includes a range of behavioural problems reflecting impairment of sustained and focused attention, impaired working memory, distractibility, forgetfulness, and/or impaired ability to organise and plan activities and tasks. These aspects of inattention have been described as independent at a biological level (see [Berry et al., 2014]), but may be extremely difficult to disentangle behaviourally. They rather tend to occur as patterns of behaviour. For example, most children may be distracted by external stimuli in a classroom situation [Rescorla et al., 2007], and these distractions will probably be especially hard to handle by a child having problems in maintaining attention and engagement in a task. Thus, specific behavioural patterns of inattention, where different weights or importance may be attached to its different components, may have a detrimental effect on the child's present and future function at school. Longitudinal studies searching for predictors of future academic achievement from primary school teachers' reports of inattention are thus called for.

In the present study we explored a set of methods to compute such predictions by using data from a Norwegian longitudinal study, the Bergen Child Study. Here, teachers of a group of 7 - 9 years old children have reported behaviour problems related to inattention by responding to nine items on a questionnaire. These reports, scored according to a Likert scale with three response alternatives, were together with age and gender used in a multivariate pattern analysis setting as predictors of academic achievement in high-school about 10 years later. The academic achievement, regarded as outcome, was obtained from official registry and discretised into three intervals (categories) constructed such that prior probability was non-informative, i.e. same number of participants in each outcome category.

The aim of the study was to introduce and investigate methods for answering: (**1**) how can early reported problems of inattention in a child be used to classify later academic achievement level at the time when he/she is a high-school student, and what is the success-rate of such prediction; (**2**) are there specific patterns of inattention problems that are associated with a given academic achievement outcome, and what are the most important predictors in such association, and finally, (**3**) what are the roles of gender and age of the high-school student in such predictions.

From previous reports we assume, however, that such predictions from inattention reports will be weak, given the long delay between the two time-points for data collection, but still substantial for some of the items [Refs ?].

In this context different types of pattern analysis and machine learning methods were selected according to the following criteria: (**i**) the methods must handle multiple predictors jointly each having a small set of response alternatives, e.g. binary, trinary, or quaternary predictor variables, and with a small set of outcome categories, e.g. low, intermediate and high level of academic achievement; (**ii**) the prediction methods should not produce results that are particularly adapted or tuned to the the sample being included, i.e. have good generalisation abilities without overfitting; (**iii**) the methods should be generic and of interest to other similar data analysis situations and prediction challenges occurring in the behavioural sciences.

Based on these criteria we explored the following methods: *multinomial logistic regression* (MLR), *classification and regression trees* (CART), *random forest ensemble learning* (RF), *support vector machines* (SVM), and *multi-layered feed-forward neural networks* (ANN), and compared qualitatively their performance.

The rest of the paper is organized as follows. In the Methods section we present the Bergen Child Study this work is part of, the sample being used, and the various data items, including the teachers's report (SNAP IV items) and academic achievement scores. We also provide description of the data according to simple explorative data analysis and visualization. In the Results section we present the aggregation of the numerical outcome for each of the classification and machine learning methods we have applied. Finally, in the Discussion section we conclude our findings according to the different methods being used and suggest interpretation in the psychological context of the overall study. We also comment on the strength and weaknesses of the different methods applied to our research questions and provide some future perspectives.

2 METHODS

The present study includes data from the Bergen Child Study (BCS) and the ung@hordaland-survey of adolescents in the county of Hordaland, Western Norway.

The first wave of BCS was launched in October 2002 and included the total population of 9,430 children attending second to fourth grade (7-9 years old, born in 1993, 1994 and 1995) in all public, private, and special schools in Bergen, and 222 children from a small municipality (Sund) outside Bergen. During the initial screening phase, parents and teachers were asked to complete a four-page BCS questionnaire (BCSQ), including, among other scales, a somewhat modified Swanson, Nolan, and Pelham Questionnaire-Fourth Edition (SNAP-IV; [Swanson et al., 2001] with all inattention and hyperactivity symptoms included in an DSM diagnosis of ADHD. Sample protocols of the first wave have been described in previous publications [Heiervang et al., 2007, Lundervold et al., 2011].

A subsample of the adolescents in the ung@hordaland study (born 1993 to 1995) participated when they were 7 to 9 years old. The adolescents received information via e-mail, and one school hour was allocated for them to complete the questionnaire. Adolescents not in school received information by postal mail to their home addresses. The questionnaire was web-based and covered a broad range of mental health- and demographic background variables, including the ADHD Symptom Rating Scale (ASRS) [Kessler et al., 2005]). The Regional Centre for Child and Youth Mental Health and Child Welfare, Uni Research, collaborated with Hordaland County Council to conduct the study. The studies were approved by the Regional Committee for Medical and Health Research Ethics in Western Norway.

2.1 The sample

The present study included participants with teacher reports on all selected SNAP-IV items from primary school when they were 7 to 9 years old (in grade 2, 3, or 4), and information about age, gender and academic achievement when they were about 16 years old - in total 2397 participants, 54.4% girls. Mean age when included in wave 4 was 16.95 years (SD $= .846$), with a nonsignificant age-difference between girls and boys ($p = .088$).

Teacher reports

Items representing *Inattention* were selected from the SNAP-IV [Swanson et al., 2001], assessing the nine items included in the inattention score of an ADHD diagnosis in the DSM-V (APA). The original SNAP-IV uses four levels to evaluate each item, whereas in our study, the parents and the teachers evaluated each answer on a 3-level item Likert-type scale ('not true', 'somewhat true', or 'certainly true') to follow the response pattern of the remaining scales included in the first stage of the BCS. Each answer was assigned a value 0, 1, or 2, with a sum score ranging from 0 to 18. Teacher reports on each item were included in the present study. These items are listed in Table 1.

Table 1: The SNAP-IV items

Inattention	
SNAP 1	Often fails to give close attention to details or makes careless mistakes in schoolwork, work, or other activities
SNAP 2	Often has difficulty sustaining attention in tasks or play activities
SNAP 3	Often does not seem to listen when spoken to directly
SNAP 4	Often does not follow through on instructions and fails to finish schoolwork, chores, or duties
SNAP 5	Often has difficulty organizing tasks and activities
SNAP 6	Often avoids, dislikes, or is reluctant to engage in tasks that require sustained mental effort (e.g., schoolwork or homework)
SNAP 7	Often loses things necessary for tasks or activities (e.g., toys, school assignments, pencils, books, or tools)
SNAP 8	Often is distracted by extraneous stimuli
SNAP 9	Often is forgetful in daily activities

Academic achievement

Academic achievement scores were provided by the official registers from the Hordaland County. In Norway, secondary schools use a scale spanning from 1 to 6, with '6' being the highest grade (outstanding competence), '2' the lowest passing grade (low level of competence), and '1' being a *fail*. The outcome scores that were available to our study was $average\_grade \in [0, 6]$, the mean value of the grades during high-school (not the final exam?), comprising all subjects except for physical education (gym).

2.2 Pattern analysis and machine learning methods

The data analysis can be divided into three parts: (**a**) data preparation, (**b**) explorative data analysis, and (**c**) pattern classification using multivariate statistical methods and machine learning techniques.

To perform these steps, explained in detail below, we used `R` (ver. 3.2.3) with selected packages in the `RStudio` (ver. 0.99.878) environment, except for Fig. 2 where `MATLAB` (R2015b) was used. Some of the methods were also scripted in `Julia` (version 0.4.2), as a `Jupyter` notebook. Most of these scripts are available upon request from the last author.

Data preparation [`code2/data_prep_20160205.R`;`data_prep_and_visualization_20160203.m`]

The original data was provided as an SPSS file and imported into the R environment. For the analysis we used the sample of $n = 2397$ children with complete data on 11 predictor variables and academic achievement as outcome according the data description shown in Fig. 1.

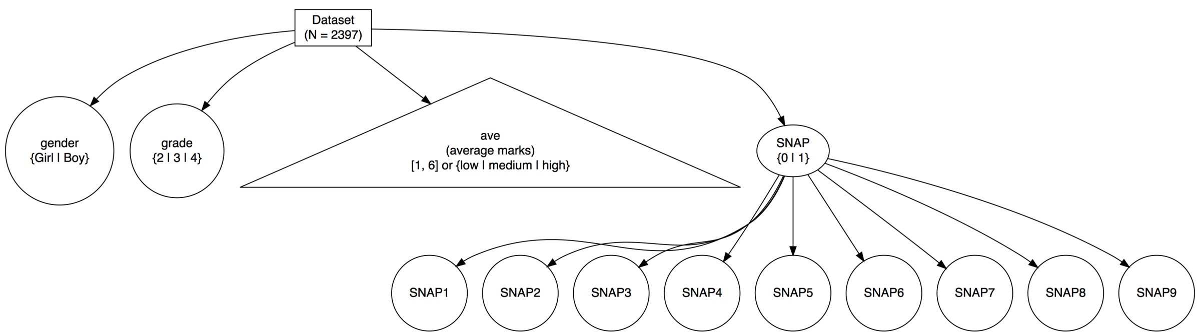

Figure 1: The data organization for the sample. *gender* is 0 (girls) or 1 (boys), *grade* is 2, 3, or 4. The three-level Likert items from Snap-IV: SNAP$1, \ldots,$ SNAP9 have each three possible values: 0 ("certainly true"), 1 ("not true"), and 2 ("somewhat true").

For classification purposes, the mean average academic achievement scores were discretised into three intervals (level of academic achievement) constructed such that prior probability was non-informative, i.e. about same number of participants in each outcome category, ignoring gender. These three level (`bins` = 3) `cutpoints` for the mean average academic achievement score `aver` was obtained as: `cutpoints<-quantile(aver,(0:bins)/bins)`.

By this we obtained the following levels and intervals: *low*: $[1.000, 3.750)$ ($n_{\text{low}} = 779$); *medium*: $[3.750, 4.429)$ ($n_{\text{medium}} = 818$); *high*: $[4.429, 5.900)$ ($n_{\text{high}} = 800$);

Depiction of the data values and the classification methods being used are given in Fig. 2.

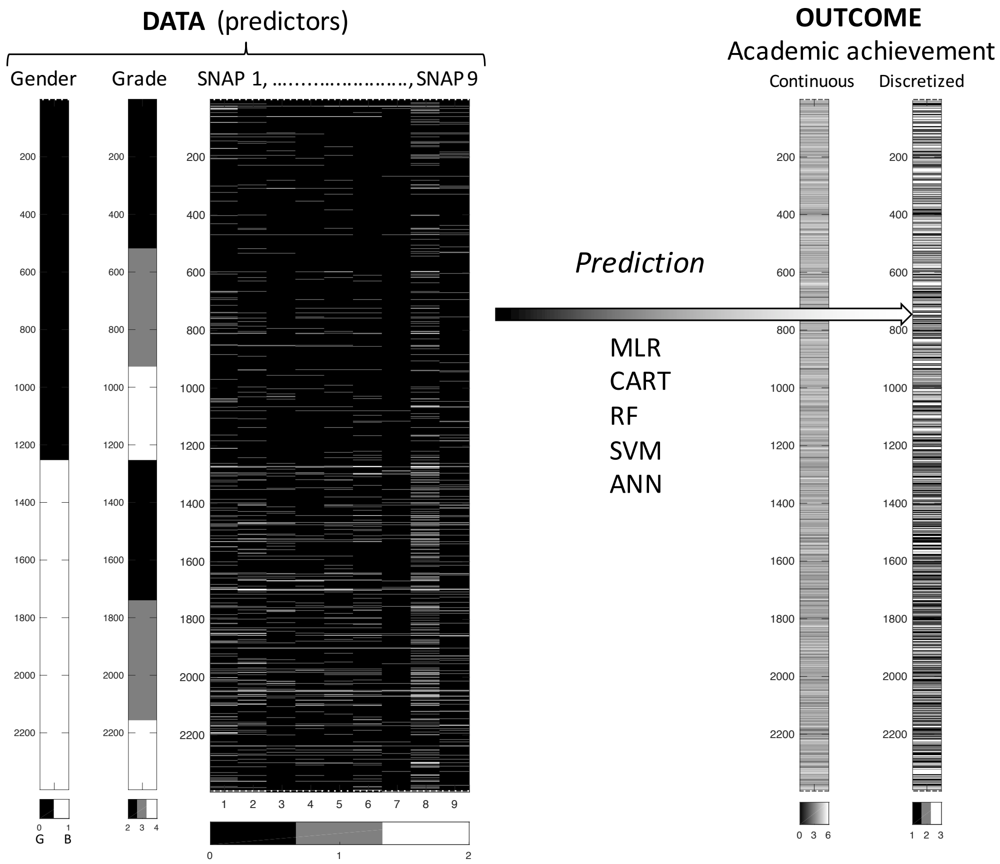

Figure 2: The predictor data (explanatory variables), the academic achievement outcome, and the types of classification analyses being performed. See Fig. 1 for explanation of variables. MLR = Multinomial logistic regression, CART = Classification and regression trees, RF = Random forest ensemble learning, SVM = Support vector machines, ANN = Artificial neural networks.

Explorative data analysis [`code2/explorative_data_analysis_20160208.R`]

Sex differences in age, the total inattention score, and the total academic achievement score were compared using parametric tests and sex differences on the single inattention items by non-parametric tests.

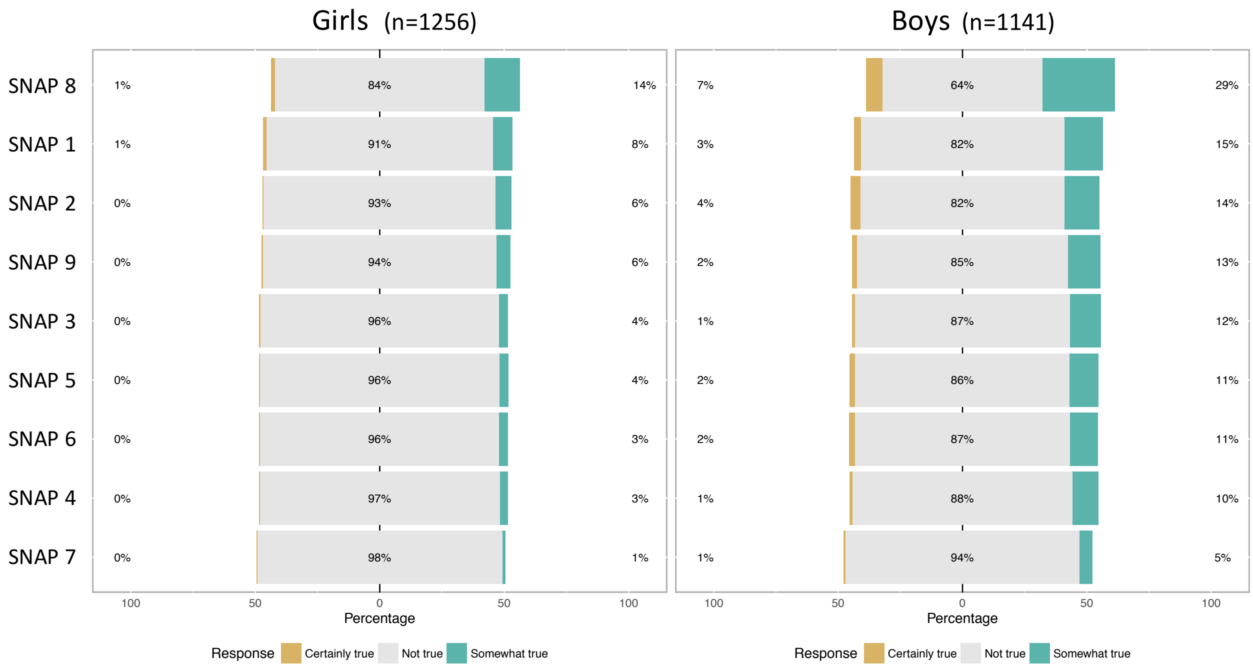

Figure 3: The gender-wise three-level Likert items SNAP 1, . . . , SNAP 9 sorted according to response pattern. See Fig. 1 for explanation of variables.

To complement Fig. 3, the numerical description of the inattention features are given in Tab. 2.2.

	Not true			*Somewhat true*			*Certainly true*		
	All (%)	Girls	Boys	All (%)	Girls	Boys	All (%)	Girls	Boys
SNAP 1	86.7	91.0	81.9**	11.3	7.6	15.4	2.0	1.3	2.7
SNAP 2	88.3	93.9	82.1**	9.6	5.6	14.0	2.1	.5	3.9
SNAP 3	91.8	96.6	86.6**	7.6	3.2	12.4	.6	.2	1.0
SNAP 4	92.5	84.4	87.4**	6.8	3.6	10.4	.7	.2	1.2
SNAP 5	91.4	95.9	86.3**	7.3	3.6	11.5	1.3	.5	2.2
SNAP 6	91.6	96.2	86.5**	7.1	3.4	11.5	1.3	.4	2.4
SNAP 7	96.5	98.5	94.2**	3.0	1.2	5.1	.5	.3	.7
SNAP 8	74.8	84.3	64.4**	21.3	14.3	29.0	3.9	1.4	6.6
SNAP 9	89.4	93.3	85.0**	9.5	6.3	13.1	1.1	.4	1.9

Table 2: Percentage of children obtaining a given response from their teachers on each inattention item (SNAP-IV). Total number of children = 2397, girls = 1256, boys = 1141. **: p value $<.001$ according to a chi-square test comparing a "not true" report in boys and girls. [Astri check this - in R?]

The distribution of average academic achievement scores are shown in Fig. 4, separately for girls and boys.

Figure 4: The gender-wise distribution of mean academic achievement scores. Vertical dashed lines represent the mean scores for boys ($= 3.91$) and girls ($= 4.12$), respectively.

Multinomial logistic regression model (MLR)

Multinomial logistic regression is used in the present study where we include a set of explanatory variables on a nominal level - the three levels of academic achievement scores - and the nine inattention items, gender and age as predictors. It is exploratory, in that we have no prior assumption regarding the weights of these items. In other words, the analysis investigates if a prediction can come from one or a limited set of items. To obtain this, the multinomial logistic regression model relates a set of explanatory variables $x_1, \ldots, x_p$ to a set of log-odds $\log(\pi_2/\pi_1), \ldots \log(\pi_J/\pi_1)$ according to

$$\log(\pi_j/\pi_1) = \beta_{j0} + \beta_{j1}x_1 + \cdots + \beta_{jp}x_p \tag{1}$$

for $j = 2, \ldots, J$. Here, $j = 1$ represents the base level category, $\pi_j = P(\text{academic achievement level} = j)$, $\pi_j/\pi_{j'}$ denotes the odds of category j relative to j', and $\sum_{j=1}^{J} \pi_j = 1$ (see e.g. [Bilder and Loughin, 2015] for details). In our case, we let the base level category $j = 1$ be the *high* (=3) mean academic achievement. For computations we used the `multinom()` function in the R package `nnet`.

Classification trees (CART)

The Snap-IV items were included together with age and gender as predictor variables in a conditional inference trees analysis (CART; Hothorn et al., 2006) with academic achievement level as the outcome variable by using the the R-package party: A Laboratory for Recursive Partytioning, including conditional inference trees (**ctree**) and conditional inference forests (**cforest**). trees (**mob**) [Strobl et al., 2009]. This particular kind of analysis takes into account the distributional properties of the measures. The conditional inference model states, that if the null hypothesis of there being independence between any of the covavariates and the response cannot be rejected, the variable in question is excluded from further exploration. However, when one variable distinguishes itself by having the strongest association with the response, a split is created with two disjoint sets of the variable in question. The Bonferroni adjusted p value of the split value is calculated. For each such node, the abovementioned procedure is repeated until none of the covariates can reject the null hypothesis. From these models we expect to find a cluster of items with a strong weight in predicting academic outcome, and that clusters of some of these variables will be generated from both methods and from the previous factor analysis.

Random forest ensemble learning (RF)

Random forest (RF) is an ensemble machine learning algorithm, which is best defined as a "combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest" [Breiman, 2001]. For classification problems, the forest prediction is the unweighted plurality of class votes, where the algorithm converges with a large enough number of trees.

For the classification of demographic (*Gender* and *Grade*) and the nine Snap-IV items, giving the feature vectors $x_1, x_2, \ldots, x_{11}$ defined above, into our set of academic achievement labels $\mathcal{C} = \{\textit{low}, \textit{medium}, \textit{high}\}$, we used the Breiman RF ensemble learning algorithm [Breiman, 2001] as implemented in the R package `randomForest`. In this setting we are given a set of training data $\mathcal{D} = \{(\mathbf{v}^i, c^i)\}_{i=1,\ldots,n}$ (to be defined below). The classification task is then to learn a general mapping from previously unseen test data to their corresponding class labels, i.e. $c : \mathbf{R}^d \to \mathcal{C}$, where $d = 11$ in our case. More specifically, adopting the notation in [Criminisi and Shotton, 2013], let $\mathbf{v} = (x_1, \ldots, x_d) \in \mathbf{R}^d$ denote the input data feature vector (predictor), and let $c \in \mathcal{C}$ denote the output academic achievement label (response). The RF algorithm incorporate a collection of binary classification trees indexed by $t = 1, \ldots, n_{\text{tree}}$. Each classification tree is characterised by its input root node, internal split nodes, and its leaf terminal nodes containing class labels.

In this setting, the RF algorithm can be briefly described as follows: (i) Draw n_{tree} samples from the original data $\mathcal{D}$, using random sampling with replacement; (ii) For each bootstrap sample, grow a classification tree such that for each node: randomly sample m_{try} of the predictor variables and chose the "best split" (Gini criterion to be defined below) from among those feature variables ($1 \leq m_{\text{try}} \ll d$). The largest tree possible is grown and is not pruned. Using only m_{try} of the predictor variables selected at random is in contrast to standard tree classification (CART), where each node is split using the best split among all d variables. (iii) The forest consists of n_{tree} trees. Each tree gives a classification for a given data point. Predict new data point $\mathbf{x}$ by putting $\mathbf{x}$ down each of the n_{tree} trees and make a majority vote for classification across the forest.

For a given tree, we let $\mathcal{S}_0$ denote the set of input predictor data vectors that is fed into the root node, $\mathcal{S}_j$ the subset of data points reaching node j in the binary splitted tree, and $\{\mathcal{S}_j^L, \mathcal{S}_j^R\}$ denote the subsets of data points that reaches the left and right child, respectively, of node j, where $\mathcal{S}_j^L \cup \mathcal{S}_j^R = \mathcal{S}_j$ and $\mathcal{S}_j^L \cap \mathcal{S}_j^R = \emptyset$. In the "off-line" tree training, each split node j is associated with a parameter vector $\boldsymbol{\theta}_j$ that is trained by optimizing an objective function I (defined below), i.e. $\boldsymbol{\theta}_j = \arg\max_{\boldsymbol{\theta} \in \mathcal{T}_j} I(\mathcal{S}_j, \boldsymbol{\theta})$.

In this notation, a binary-valued test function $h(\mathbf{v}, \boldsymbol{\theta}_j) : R^d \times \mathcal{T} \to \{0, 1\}$ is applied at each split node j. Here, 0 and 1 denote "true" and "false", respectively, and the data point $\mathbf{v}$ arriving at split node j is sent to its left (0) or right (1) child node, accordingly. $\mathcal{T}$ is the set of all possible split function parameters, and $\mathcal{T}_j \subseteq \mathcal{T}$ is the subset of parameters available at node j. We thus have $\mathcal{S}_j^L(\mathcal{S}_j, \boldsymbol{\theta}) = \{(\mathbf{v}, c) \in \mathcal{S}_j \mid h(\mathbf{v}, \boldsymbol{\theta}) = 0\}$ and $\mathcal{S}_j^R(\mathcal{S}_j, \boldsymbol{\theta}) = \{(\mathbf{v}, c) \in \mathcal{S}_j \mid h(\mathbf{v}, \boldsymbol{\theta}) = 1\}$.

The objective function being used was the Gini index, i.e. $I = i(\tau) = 1 - \sum_{c \in \mathcal{C}} p_c^2$, measuring the likelihood that a data point would be incorrectly labeled if it was randomly classified according to the distribution of class labels within the node. The optimal binary split is then the one that maximises the improvement in the Gini index.

Proximity, $P(i,j)$ in the RF model is the proportion how often two data points (rows), i and j end in the same leaf node for different trees. *Outliers* are defined as cases having small proximities to all other cases. Since the data in some classes is more spread out than others, outlyingness is defined only with respect to other data in the same class as the given case. More specifically, for a given case i let $out(i) = 1/A_i$, where

$$A_i = \sum_{k \text{ in same class as } i} P(i,k)^2$$

For all i in the same class, let the mean absolute deviation $MAD = \frac{1}{n}\sum_{i=1}^{n} |out(i) - med|$, where med is the median of $out(i)$ for all i in the same class. The outlyingness of a case i is then the MAD-normalized measure, $outlyingness(i) = \max\{(out(i) - med)/MAD, 0\}$.

Support vector machine (SVM)

Support vector machines (SVMs) are supervised models with an aim to learn structures in the data, here by including the tree categories of academic achievement level as outcome variable, and learning algorithms to learn how these categories can be explained from information about teacher reports on the nine inattention items, age and sex. A training set is selected where the participants are marked for belonging to one of the three categories, a model is built by the SVM training algorithm to assign new participants into one of the categories. The SVM model is mapped as points in space, where the three categories are divided as clearly as possible. New examples are then predictied to belong to one of these categories depending on the side if the gap they fall on. In the present study we use a non-linear classification using the kernel trick, where the input is mapped into high-dimensional feature spaces.

Artificial neural networks (ANN)

Artificial neural networks (ANN) method is another machine learning technique inspired by biological neural networks. Like in biology, the method present the results as interconnected ”neurons” that may exchange information between each other. They may both have different weighs and are adaptive and therefore capable of learning. In the present study, the method is appropriate, because a category belonging can be dependent on several variables. In this model, these variables are processing elements or ”neurons” that are connected together to form a network. In the present study the neural network model is thus used to classify patterns of inattention and the sequence of their weights, and by this probably detect novel associations.

3 RESULTS

Multinomial logistic regression model (MLR)

For presenting MLR results using `~/Dropbox/Arvid_Inattention/code2/MLR_20160211.R`, see `~/Dropbox/Arvid_Inattention/code2/MLR_20160211.html`

and Andy Field et al. *Discovering Statistics Using R*. Sage Publishing 2012.

`http://studysites.uk.sagepub.com/dsur/main.htm`

Classification trees (CART)

A CART analysis, including the three categories of academic achievement scores as outcome variable, and the nine SNAP items (dichotomized) , gender and grade as predictors, generated five outcome clusters (Figure 2). The first strongly significant split ($p < .001$) was dependent on wether the teachers reported that a child *Often avoids, dislikes, or is reluctant to engage in tasks that require sustained mental effort (e.g., schoolwork or homework, item 6).* If teachers reported that this was somewhat true or true, 202 children were allocated to a cluster where more than 60% obtained academic achievement scores at the lowest level. A similar distribution was found if teachers reported that it was somewhat true or true that the the teacher reported that the child *Often is distracted by extraneous stimuli* AND *often has difficulty sustaining attention in tasks or play activities* ($p = .001$). Gender and grade were not found to influence these two cluster allocations, but the primary school class level was somwhat of importance if the teacher only reported that the distractibility was true, with a higher percentage of children allocated to the medium academic achievement group when in the third grade. Very few of these children obtained academic achievement scores at the highest level. If the child was neither reported to be distractible nor characterized by poor sustained attention, most girls obtained results in the highest or second highest level (70%). For boys, poor sustained attention interacted with grade levels. Boys evaluated in the fourth grade showed a weak trend towards a higher academic score than when evaluated in the third or second class level (p = .045).

Random forest ensemble learning (RF)

We ran the RF classification using the call:

```
randomForest(formula=averBinned ~.,
   data=D,importance=TRUE,proximity=TRUE,keep.forest=TRUE,ntree=500)
```

resulting in a out-of-bag (OOB) estimate of error rate: 56.95% and confusion matrix given by Table 3.

Table 3: RF Confusion Matrix

	low	medium	high	class.error
low	245	243	291	0.685
medium	143	265	410	0.676
high	64	214	522	0.348

The variable importance got estimated according to Table 4, being depicted in sorted order of importance in Figure 5.

Table 4: RF variable importance

	low	medium	high	MeanDecreaseAccuracy	MeanDecreaseGini
Gender	5.120	-11.950	28.730	21.120	20.670
Grade	-6.140	14.830	17.840	17.030	30.170
SNAP1	0.070	-16.920	26.920	12.130	23.460
SNAP2	-0.510	-0.750	33.840	31.030	34.900
SNAP3	-22.720	2.900	21.800	6.310	17.050
SNAP4	-9.720	-9.200	20.450	5.650	16.550
SNAP5	-10.130	-2.050	20.170	11.630	19.370
SNAP6	-10.860	-3.590	28.680	15.570	22.610
SNAP7	-3.070	-5.620	12.120	3.910	11.080
SNAP8	1.440	-21.490	34.490	18.880	34.660
SNAP9	-7.540	-1.740	11.130	2.510	19.980

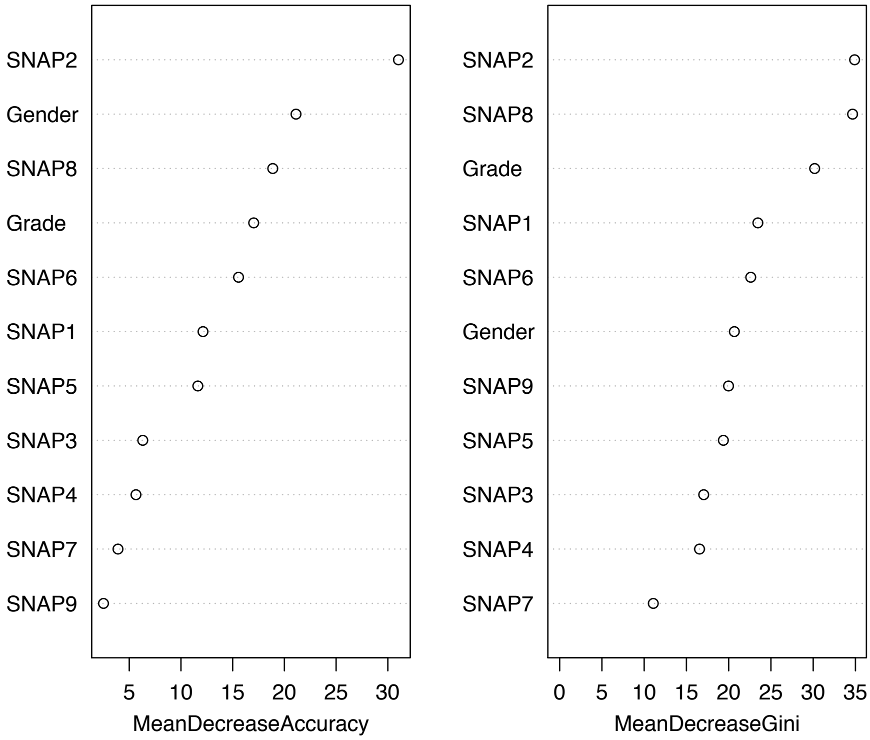

Figure 5: RF variable importance computed by the `importance()` function in the R package `randomForest`

Computing a measure of outlyingness for each case, based on a proximity matrix (explained in the Methods section), resulted in a numeric vector as depicted in Figure 6. These values are color-coded for academic achievement outcome in each case. According to Breiman (2007), an outlyingness value above 10 is reason to suspect the case of being outlying.

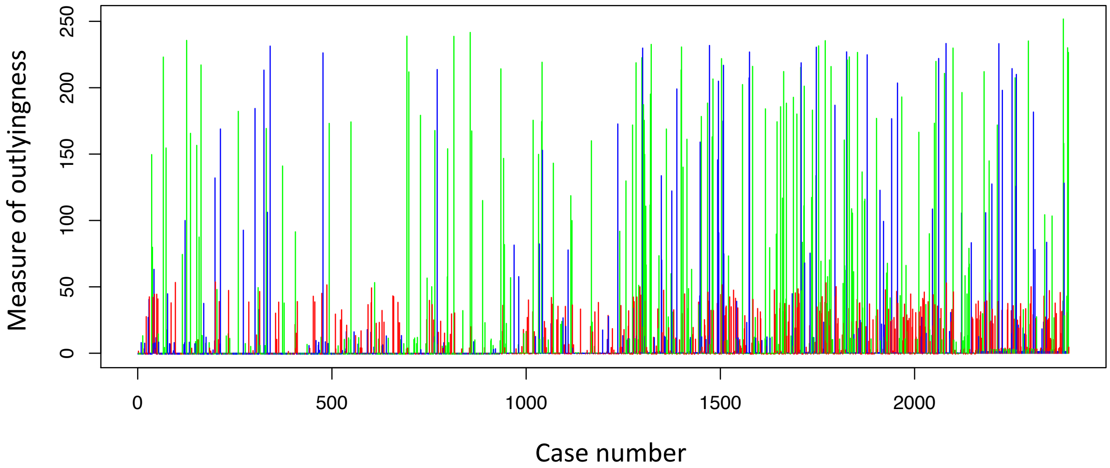

Figure 6: RF outlier measures computed by the `outlier()` function in the R package `randomForest`. Colorcoding: red = *low*, green = *medium*, blue = *high* academic achievement.

A random forest regression including the three categories of achievement scores generated 500 trees was computed. The mean squared residuals was .483, and the explained 9.3% of the variance in the academic achievement score in high school. The analysis also indicates the ”importance” of variables (higher value means higher ”importance”). The results (Table 4) select the item where teachers reported that the child *often has difficulty sustaining attention in tasks or play activities* as the most important, followed by class grade in primary school and the item reflecting that the teachers find that the child *Often is distracted by extraneous stimuli*. However, the confusion matrix showed that the classification errors were high both for the medium and low academic achievement scores, while it is fairly good for the highest level. This shows the importance of running analyses including a high number of repetitions. Doubling the number of repetitions (n = 1000) did not change the ”importance” of the single variables, but reduced the classification errors

Support vector machine (SVM)

Artificial neural networks (ANN)

4 DISCUSSION

Summary of results

The present study showed that inattention symptoms in early school years were not only frequently reported by teachers, but they did also contribute substantially to predict achievement in high-school nine years later.Inclusion of nine items reflecting different aspects of inattention in a multinominal regression analysis showed about a doubling of odds ratio for a low compared to high academic score for each extra score on items reflecting behaviour reflecting problems with sustained attention (items 1, 2, 6) and distractibility (item 8). Inclusion of the nine items, gender and sex in the CART analysis revealed some interesting behavioural patterns and weights of the single variables. The first strongly significant item was related to how the child avoids, dislikes or is reluctant to take part in tasks that require sustained mental effort. As it was reported by primary school teachers, this was probably first of all related to school- and homework. Reports of distractibility turned out to be as important, both alone, but also when appearing together with problems related to sustained attention. A random forest regression analysis generating 500 decision trees confirmed sustained attention and distractibility to be the most important items in the decision tree. Although the model only explained 10% of the variance and turned out to be the best predictor of the highest academic achievement level, the results should be carefully considered by teachers when detecting inattention problems in primary school children.

Some of our results clearly confirmed findings in a study by Holmberg et al. [Holmberg and Bolte, 2014], showing that teacher reports of failure to finish a task was one of the main factors in explaining academic outcome. In the present study, a more school- and homework related item was selected to have the strongest importance, but both variables are clearly related to the ability to sustain attention. Our study add to Holmberg and collegueas' findings by revealing the importance of a behaviour pattern of inattention including distractibility. Its relation to sustained attention in childhood is obvious, as a child with poor vigilance is expected to be disturbed by habits and disturbing cues in the environment. Taken together, these symptoms seem to represent the working memory and response inhibition problems described as hallmarks in several neuropsychiatric disorders, including ADHD [Sonuga-Barke, 2005], and which are known to lead to a casscade of other problems [Gillberg, 2010, Sonuga-Barke and Halperin, 2010]. When assessing sustained attention, one should therefore consider its association with distraction, as recently done in a study by Cassuto and collaborators [Cassuto et al., 2013] including external distractors as part of a standard continuous performance test (CPT). But even children handling distractions in a standardized test situation may have problems in the classroom. Here, a child is expected to stay focused in a more disorganized situations with several external and internal distractors. When having difficulties in maintaining attention on a task, these distractors may be especially tempting. From experients in cognitive psychology we know that this is a challenge that is strongest in situations with high cognitive load [**?**]. Problems related to inattention are thus expected to be more hard to handle when a higher load is put on the curriculum in higher grades.

Girls were reported by their primary school teachers to have less inattention symptoms than boys, and they obtained significantly higher academic achievement scores in high school. The importance of gender was, however, not as clear as expected when the relation was investigated by the selected statistical procedures. These results suggests that the relation between the two variables was more gender balanced than measures of each of the two separately.

In the present study we included a statistical method that handle situations where predictor variables with a few number of categories may hide complex interactions. We thus found this method appropriate with our aim to investigate if specific problem areas are of importance to predict future academic outcome. Inattention is a complex concept, where the weights of different problem areas and their interactions may give information of importance to understand how to identify essential problems in a child. An alternative approach to cope with large numbers of predictor variables would be to first apply dimension reduction techniques, such as principle components or factor analysis, and then use standard regression methods on the reduced data set. However, this approach has the disadvantage that the original input variables are projected into a reduced set of components, so that their individual effect is no longer identifiable. As opposed to that, random forests can process large numbers of predictor variables simultaneously and provide individual measures of variable importance. proves to be more stable than stepwise variable selection approaches available for logistic regression, which are known to be affected by order effects (see, e.g., Austin, Tu, 2004; Derksen & Keselman, 1992; Freedman, 1983). By this and results from the present study, we argue for the appropriateness of this cluster of statistical methods when dealing with a multidimensional dataset and a high number of participants.

dicotomi when these children were 16 to 19 years old high-school students. The were cademic achievement is here obtained from an official registry wheHere we investigate if academic achievement in a population-based sample of Norwegian 16 - 19 years old high-school students can be classified from reports of inattention reports given by their primary teachers about 10 years earlier. As both age [West et al., 2014] and gender [Gershon, 2002, Sørensen et al., 2008, Biederman et al., 2005] should be controlled for in studies focusing on inattention, they will both be added as predictors in the statistical analysis.

Strengths and Limitations

The large population-based sample of high-school students followed from childhood and inclusion of a standardized questionnaire assessing inattention and hyperactivity are two main strengths of the present study. A third strength was the inclusion of recursive partitioning methods to analyse our data. By this we could handle predictor variables with few categories and possible hided multidimensional interactions, and still generate behavioural patterns that was easy to interpret.

In spite of the strengths and the importance of the findings generated, several limitations must be mentioned. Although the statistical analyses have several strengths, applications of recursive partitioning methods in psychology also reveal common misperceptions and pitfalls. For example, Luellen et al. (2005) suspected that ensemble methods could overfit (i.e., adapt too closely to random variations in the learning sample). By adding recursive partitioning, a high number of threes were generated. This is important, in that the selection of a splitting variable will strongly depend on the distribution in the learning sample. By repeating the procedure as in the present study, the prediction is expected to be more correct than when based on a single tree.

Inclusion of very few predictor variables when assessing predictions over a nine-years period as in the present study, is another obvious limitation of the present study. A stronger model could have been obtained by including results from psychometric test assessing vigilance and distractibility, similar to the one developed by Cassuto et al. [Cassuto et al., 2013], or more ecological valid virtual reality test as the described by Pelham et al. [Pelham et al., 2011]. Inclusion of only teacher reports may also be considered as a limitation. Although parent reports obviously are of importance, teacher reports were selected due to the focus of the present study and studies showing that teacher reports are stronger predictors of academic success than parent resports (). An even more sophisticated analysis would be to include repeated inattention measures to understand the trajectory from early symptoms of inattention to function in adolescence and adulthood. Its importance was shown in a study by Pingault and collaborators [Pingault et al., 2014], showing that increase in symptoms of inattention during childhood really matters when it comes to school graduation failure. Such studies are important and should include analysis of behavioural patterns, because a specific pattern of vigilance and distraction was suggested by the present study.

Clinical and methodological implications and future directions

Knowledge about the future function of children identified as inattentive by their primary school teachers should lead to remediation procedures. Poor inattention is not only predictive of poor academic outcome, but also linked to a wide range of other present and future problems.The present study made us aware of the importance of a behavioural pattern including both the ability to sustain attention and to handle distractibility. However, only about 10% of the variance was explained by the inattention measures, leaving several other factors still unknown as predictors of later academic achievements that should be explored in future studies. The importance of cognitive factors has already been mentioned, but a model should clearly take into account the mulitcontextual importance of family, school, emotional, social and behavioural characteristics of the participants, as well as individual domains for adolescents' school success. Furthermore, there is a call for studies comparing results in countries with different attitudes and pressure on academic success. Large-scaled studies including a large number of participants and a wide range of variables should consider using the statistical methods advocated for in the present study, Based on findings of weights and behavioural patterns of important to educational success, there is a call for studies of remediation procedures for problems related to inattention.

Acknowledgments

The present study was supported by the Centre for Child and Adolescent Mental Health and Welfare, Uni health, Uni Research, Bergen, Norway, and was also funded by the University of Bergen, the Norwegian Directorate for Health and Social Affairs, and the Western Norway Regional Health Authority. We are grateful to the children, parents and teachers participating in the Bergen Child Study (BCS) and members of the BCS project group for making the study possible.

References

[Bellanti and Bierman, 2000] Bellanti, C. J. and Bierman, K. L. (2000). Disentangling the impact of low cognitive ability and inattention on social behavior and peer relationships. conduct problems prevention re search group. *J Clin Child Psychol*, 29(1):66–75.

[Berry et al., 2014] Berry, A. S., Demeter, E., Sabhapathy, S., English, B. A., Blakely, R. D., Sarter, M., and Lustig, C. (2014). Disposed to distraction: genetic variation in the cholinergic system influences distractibility but not time-on-task effects. *J Cogn Neurosci*, 26(9):1981–1991.

[Biederman et al., 2005] Biederman, J., Kwon, A., Aleardi, M., Chouinard, V.-A., Marino, T., Cole, H., Mick, E., and Faraone, S. V. (2005). Absence of gender effects on attention deficit hyperactivity disorder: findings in nonreferred subjects. *Am J Psychiatry*, 162(6):1083–1089.

[Bilder and Loughin, 2015] Bilder, C. R. and Loughin, T. M. (2015). *Analysis of Categorical Data with R*. CRC Press.

[Breiman, 2001] Breiman, L. (2001). Random forests. *Machine Learning*, 45:5–32.

[Cassuto et al., 2013] Cassuto, H., Ben-Simon, A., and Berger, I. (2013). Using environmental distractors in the diagnosis of adhd. *Front Hum Neurosci*, 7:805.

[Connors et al., 2012] Connors, L. L., Connolly, J., and Toplak, M. E. (2012). Self-reported inattention in early adolescence in a community sample. *J Atten Disord*, 16(1):60–70.

[Criminisi and Shotton, 2013] Criminisi, A. and Shotton, J., editors (2013). *Decision Forests for Computer Vision and Medical Image Analysis*. Springer-Verlag.

[Garner et al., 2013] Garner, A. A., O?connor, B. C., Narad, M. E., Tamm, L., Simon, J., and Epstein, J. N. (2013). The relationship between adhd symptom dimensions, clinical correlates, and functional impairments. *J Dev Behav Pediatr*, 34(7):469–477.

[Gershon, 2002] Gershon, J. (2002). A meta-analytic review of gender differences in adhd. *J Atten Disord*, 5(3):143–154.

[Gillberg, 2010] Gillberg, C. (2010). The essence in child psychiatry: Early symptomatic syndromes eliciting neurodevelopmental clinical examinations. *Research in Developmental Disabilities*, 31(6):1543–1551.

[Gray et al., 2014] Gray, S. A. O., Carter, A. S., Briggs-Gowan, M. J., Jones, S. M., and Wagmiller, R. L. (2014). Growth trajectories of early aggression, overactivity, and inattention: Relations to second-grade reading. *Dev Psychol*, 50(9):2255–2263.

[Heiervang et al., 2007] Heiervang, E., Stormark, K. M., Lundervold, A. J., Heimann, M., Goodman, R., Posserud, M.-B., Ullebø, A. K., Plessen, K. J., Bjelland, I., Lie, S. A., and Gillberg, C. (2007). Psychiatric disorders in norwegian 8- to 10-year-olds: an epidemiological survey of prevalence, risk factors, and service use. *J Am Acad Child Adolesc Psychiatry*, 46(4):438–447.

[Holmberg and Bolte, 2014] Holmberg, K. and Bolte, S. (2014). Do symptoms of ADHD at ages 7 and 10 predict academic outcome at age 16 in the general population? *J Atten Disord*, 18(8):635–645.

[Kessler et al., 2005] Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E., Howes, M. J., Jin, R., Secnik, K., Spencer, T., Ustun, T. B., and Walters, E. E. (2005). The World Health Organization Adult ADHD Self-Report Scale (ASRS): a short screening scale for use in the general population. *Psychol Med*, 35(2):245–256.

[Lundervold et al., 2011] Lundervold, A. J., Adolfsdottir, S., Halleland, H., Halmøy, A., Plessen, K., and Haavik, J. (2011). Attention network test in adults with ADHD–the impact of affective fluctuations. *Behav Brain Funct*, 7:27.

[Pelham et al., 2011] Pelham, Jr, W. E., Waschbusch, D. A., Hoza, B., Gnagy, E. M., Greiner, A. R., Sams, S. E., Vallano, G., Majumdar, A., and Carter, R. L. (2011). Music and video as distractors for boys with ADHD in the classroom: comparison with controls, individual differences, and medication effects. *J Abnorm Child Psychol*, 39(8):1085–1098.

[Pingault et al., 2014] Pingault, J.-B., C?t?, S. M., Vitaro, F., Falissard, B., Genolini, C., and Tremblay, R. E. (2014). The developmental course of childhood inattention symptoms uniquely predicts educational attainment: a 16-year longitudinal study. *Psychiatry Res*, 219(3):707–709.

[Polderman et al., 2010] Polderman, T. J. C., Boomsma, D. I., Bartels, M., Verhulst, F. C., and Huizink, A. C. (2010). A systematic review of prospective studies on attention problems and academic achievement. *Acta Psychiatr Scand*, 122(4):271–284.

[Rescorla et al., 2007] Rescorla, L., Achenbach, T. M., Ivanova, M. Y., Dumenci, L., Almqvist, F., Bilenberg, N., Bird, H., Broberg, A., Dobrean, A., Düpfner, M., Erol, N., Forns, M., Hannesdottir, H., Kanbayashi, Y., Lambert, M. C., Leung, P., Minaei, A., Mulatu, M. S., Novik, T. S., Oh, K.-J., Roussos, A., Sawyer, M., Simsek, Z., Steinhausen, H.-C., Weintraub, S., Metzke, C. W., Wolanczyk, T., Zilber, N., Zukauskiene, R., and Verhulst, F. (2007). Epidemiological comparisons of problems and positive qualities reported by adolescents in 24 countries. *J Consult Clin Psychol*, 75(2):351–358.

[Sonuga-Barke, 2005] Sonuga-Barke, E. J. S. (2005). Causal models of attention-deficit/hyperactivity disorder: from common simple deficits to multiple developmental pathways. *Biol Psychiatry*, 57(11):1231–1238.

[Sonuga-Barke and Halperin, 2010] Sonuga-Barke, E. J. S. and Halperin, J. M. (2010). Developmental phenotypes and causal pathways in attention deficit/hyperactivity disorder: potential targets for early intervention? *J Child Psychol Psychiatry*, 51(4):368–389.

[Sørensen et al., 2008] Sørensen, L., Hugdahl, K., and Lundervold, A. J. (2008). Emotional symptoms in inattentive primary school children: a population-based study. *J Atten Disord*, 11(5):580–587.

[Strobl et al., 2009] Strobl, C., Malley, J., and Tutz, G. (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol Methods*, 14(4):323–348.

[Swanson et al., 2001] Swanson, J. M., Kraemer, H. C., Hinshaw, S. P., Arnold, L. E., Conners, C. K., Abikoff, H. B., Clevenger, W., Davies, M., Elliott, G. R., Greenhill, L. L., Hechtman, L., Hoza, B., Jensen, P. S., March, J. S., Newcorn, J. H., Owens, E. B., Pelham, W. E., Schiller, E., Severe, J. B., Simpson, S., Vitiello, B., Wells, K., Wigal, T., and Wu, M. (2001). Clinical relevance of the primary findings of the mta: success rates based on severity of adhd and odd symptoms at the end of treatment. *J Am Acad Child Adolesc Psychiatry*, 40(2):168–179.

[West et al., 2014] West, G. L., Mendizabal, S., Carriere, M.-P., and Lippe, S. (2014). Linear age-correlated development of inhibitory saccadic trajectory deviations. *Dev Psychol*, 50(9):2285–2290.