Strength and difficulties questionnaire (SDQ) as an early predictor of school dropout: results from a longitudinal study in 24,988 Dutch children.

Putrik, P.[1,2], Kant, IJ[3], Hoofs, H[3], Reijs, R[2], Jansen, MWJ[1,2]

[1] Department of Health Services Research, Care and Public Health Research Institute (CAPHRI), Faculty of Health, Medicine and Life Sciences (FHML), Maastricht University, Maastricht, the Netherlands
[2] GGD Zuid Limburg, Academic Collaborative Centre for Public Health Limburg, Heerlen, the Netherlands
[3] Department of Epidemiology, Care and Public Health Research Institute (CAPHRI), Faculty of Health, Medicine and Life Sciences (FHML), Maastricht University, Maastricht, the Netherlands

Key words: school dropout, SDQ, prediction model, youth health services, early prevention

Abstract

Possible journals (please feel free to add):

- Lancet Public health (IF 16) – too ambitious?
- Social Science and Medicine (IF 3.6)
- Journal of Public health (IF 1.8)
- Social Science Research (IF 1.9)
- Public Health (IF 1.7)
- Children and Youth Services Review (IF 1.5)
- There is a number of other 'public health' journals (International journal of PH, European journal of PH…)

Introduction (694 words)

School dropout is an economic, social, and individual problem, and has long-term consequences including difficulty to engage in the labor market, subsequent socio-economic deprivation and increased risk to engage in criminal activities.(1-3) Over the last decades, a lot has been done in the Netherlands to reduce school dropout, including the regulation about the obligatory schooling till 18 years old and a number of other ongoing programs.(4) After a significant decline in school dropout in the years preceding 2015 (from 2.8 to 1.7%), an increasing trend has been documented between 2015 and 2019 (towards 2%), indicating that the problem of school dropout requires sustained attention.(5)

A growing body of research suggests that dropping out is but the final stage in a dynamic and cumulative process of disengagement from school.(6) Many factors have been proven predictive of school dropout, with most of them having a long exposure before the dropout happens. A substantial body of research has been done in US context, as well as in a few European countries.(7) Known predictors of dropout include family and neighborhood socio-economic status, academic achievement throughout the school years, school and teacher support, behavioral problems, and physical and psychological health problems.(7, 8) When recognized early, some factors can be amended (e.g. additional support with educational load) or the negative impact of non-modifiable factors can be reduced (e.g. engaging the child from deprived families in structural social and sport activities). This underlines the value of efforts for early prediction of school dropout.

An important question is which organizations or parties should play a role in recognition and prevention of school dropout. As of now, Dutch schools can signal the problems with the family (where possible) and contact social services supported by the municipalities. Student is then provided tailored supervision and support depending on the situation. However, the Achilles heel of this approach is that help is being activated when clear signs of upcoming problem are in sight, which might be too late. An earlier risk stratification of the student population, for example, at the transition between the primary school and secondary school, could potentially allow more effective efforts towards strengthening the ties with school and supporting successful graduation. The Dutch Youth Health Care (YHC), a municipal services that monitors the development of children and adolescents in the Dutch population, could play a role in this process. YHC regularly sees children between birth and age 18, with very frequent contacts in the first 4 years, several planned appointments in the primary and secondary school years as well as structural collaboration with schools in questions related to health and upbringing.

An important instrument used by the YHC is the Strengths and Difficulties Questionnaire (SDQ).(9) The SDQ is a widely used questionnaire which is often used as a screenings instrument for the initial assessment of psychosocial problems ((10, 11)) and was previously validated in the Dutch population. (12-14) Using the SDQ, an increased risk of emotional problems, behavioral problems, hyperactivity, peer problems and lack of pro-social behavior can be determined. To our knowledge, SDQ has never been explored in relation to future school dropout. We hypothesize that high SDQ score is predictive of school dropout.

In the Dutch region of South Limburg, the SDQ has been consistently used as a screening instrument for nearly two decades. Combining the SDQ with other routinely available to YHC professionals background information of the children and their families with school success data could give a new insight into the complex problem of school dropout. If a reasonable prediction model can be developed that signals children with a high risk of school dropout years ahead of the time child would leave school, it could help to steer efforts to improve child's chances to graduate and enter the labour market with a qualification. Such model could empower YHC to take a proactive role in prevention of school dropout in partnerships with school and family.

Therefore, the objective of this study was (1) to assess whether SDQ scores are predictive of school dropout and (2) to develop a school dropout prediction model using relevant child and family factors available to YHC professionals at regular consultations for age 10 and 14.

Methods (1060 words)

Study design

This study was a longitudinal cohort study in South Limburg province of the Netherlands. Data on SDQ scores was collected during regular assessments by Youth Health Service of children aged 10 and 14. Information on child's and family's socio-economic status (at child's age 10 and 14), school type (age 14) and school dropout status (age 17) was obtained from national population registries through Statistics Netherlands The data linkage was executed using the individual number of municipal administration (GBA) or, where this number was not available, a combination of birth date, gender, and postal code of the child. The data linkage was performed by a trusted third party (Statistics Netherlands), the data has been pseudo anonymized before it was made accessible for the researchers. The study was approved by medical ethical committee of Maastricht Academic Hospital and Maastricht University (METC azM/UM 2020-1573).

Participants

Children born between 1996-2001 who (1) lived and attended school (excluding 'special education' for children with physical and intellectual disabilities) in South Limburg at age 10 and/or 14 were included in this study. YHC intends to see all children at these ages for a regular health check. During such appointments YHC collects information about psychological and physical health, provides advice and counselling to child and family, and initiates help trajectories with other partners where necessary. In practice, not all children are being reached out for the health checks due to capacity and organizational limitations.

Study outcome

School dropout defined as having left the school at least once without diploma by the age of 17 was the study outcome. This data is being annually collected by a Dutch governmental organization responsible for organization and funding of education system (DUO). Each student who was enrolled in the education system in preceding year is categorized as (1) enrolled in study (2) graduated or (3) left without diploma on October, 1st of the next academic year. Binary variable 'enrolled or graduated' vs 'dropped out' was created. The cut-off of 17 y.o. was chosen to ensure that data was available from children from each of the 6 birth cohorts (1996-2001). While it is possible that someone classified as 'dropped out' in this way can still enroll in another study and graduate later, having had dropped out at least once by age of 17 can be seen as a proxy to social vulnerability and is an indicator of potential difficulties in the future.(2)

Independent variables

The SDQ is a questionnaire consisting of 25 Likert scale questions. Each question is scored as "Not true", "Somewhat true" and "Certainly true". The 25 questions are split into five subscales; emotional problems, behavioral problems, hyperactivity, peer problems and pro-social behaviors, with each subscale consisting of 5 questions each. The total SDQ score (range 0-40, higher scores indicate more difficulties) is the sum of the subscales except pro-social behaviors.(9, 11) SDQ score was modelled as a continuous score, as well as using the proposed cut-offs points to categorized results into 'normal', 'borderline' and 'high' SDQ.(9, 11, 15) The questionnaire was completed by the parents of children aged

10, and by children themselves at age of 14. Child's gender and immigration background (Dutch, first and second generation immigrant) was obtained from municipal population registry.

The highest achieved education of parents was defined as the highest education level (categorized as low, medium or high) achieved by one of the parents. If education of one of the parents was missing, the highest achieved education of the second parent was used. A binary variable 'at least one parent receiving a state welfare subsidy' was created using data from the socio-economic registry if at least one of the parents had an unemployment or long-term welfare subsidy as main source of income when child was age 10 and 14. 'Parents registered on the same address' vs 'not registered at the same address' when child was age 10 and 14 was used as a proxy to capture single-parent households.

Type of secondary education at age 14 was obtained from national education registry. Secondary education level is being assigned at the age of 12 based on the results of the standardized test (CITO) and primary school advice. The secondary education level was categorized as (1) bridge class (for a small number of children that could not yet be allocated to one of the levels) (2) secondary vocational education (VMBO) (3) secondary professional education (HAVO) (4) secondary scientific education (VWO).

Statistical analyses

First, data were summarized as means (SD) and frequencies of the distributions, as appropriate. Next, multivariable multilevel logistic models with total SDQ score and 5 subscores (separate models) and random intercepts for school have been computed for predictors at age 10 and 14. Models with and without random intercepts for school were compared by means of LR test. Interactions between SDQ and type of secondary school were tested in the model with predictors at age 14. Additionally, interactions with gender were explored in both models given that SDQ scores could have differential impact on the outcome in boys and girls.(16) Backwards selection procedure was used for variable selection ($p<0.10$).

Hosmer and Lemeshow Test was used to assess goodness of fit.(17) The test assesses whether or not the observed event rates match expected event rates in subgroups of the model population. Models for which expected and observed event rates in subgroups are similar are called well calibrated. Calibration plots were used to graphically explore the model fit.(18)

A Receiver Operating Characteristic (ROC) curve was plotted using the predicted probability of school dropout and the observed dropout. The area under the ROC curve (AUC) determined the discriminative ability of the model, i.e. how well the model can differentiate between those with high risk of outcome and those with low risk of outcome.(19, 20) The AUC takes a value between 0.5 (equivalent to flipping a coin) and 1.0 (perfect prediction). Random K-fold validation of the AUC curve has been performed (k=10).(21) A series of sensitivity analyses were performed with a number of categorizations of SDQ scores (15), as well as quintiles and deciles of the distribution. Further, age at the moment of completing the questionnaire (due to organizational reasons, some kids were younger or older than 10 or 14 during the measurement) was factored into the model and model sensitivity to this factor was explored.

Analyses were performed in Stata 16.(22) Statistical significant was assumed at $\alpha=5\%$.

Results (953 words)

Descriptive statistics

In total, 11,589 and 18,955 children born between 1996-2001 completed the SDQ questionnaire sent by Youth Health Service at age 10 and 14 (Table 1), respectively. In both groups, approx. 51% were

boys, 9% of children came from households where at least one of the parents had state subsidy as main source of income, and 11% of children came from families where one or both parents had low education. In 20% of children education of parents could not be obtained due to missings in the national registry. Among children measured at age 10, 22% of parents were not living together, and this proportion was slightly higher among 14 year-olds (25%). From secondary school children (age 14), 934 (5%) were in the bridge class, 7,772 (40%) were enrolled in secondary vocational education (VMBO), 4,705 (24%) into secondary professional education (HAVO), and 5,731 (30%) were enrolled in the highest level study, namely, secondary scientific education (VWO). Of note, 74% of 'bridge class' students were registered at lowest level of the secondary school (VMBO) 2 years after the measurement. Mean(SD) SDQ at age 10 was 5.81 (5.11) and at age 14 mean scores were 9.34 (5.01)

To explore the coverage achieved by Youth Health Care (in terms of response to the questionnaire that is part of the regular assessment), total number of children born between 1999 and 2001 and living and studying in South Limburg was compared to number of children reached by Youth health care (data before 1999 was not available). On average, questionnaire was completed for 63% of 10 years-old children. Similar analyses was performed for age 14 assessment (for birth cohorts 2000-2001) resulting in 58% and 26% reach. A comparison of dropout rates at age 17 among children that completed and did not complete the SDQ questionnaire revealed a difference (4.2 vs 6.6% and 3.8 vs 5.7% for age 10 and 14 measurements, respectively).

Table 1. Characteristics of the study population

Variable	Subsample with measurements at age 10 Mean (SD) [range]/ N(%)* n=11,589	Subsample with measurements at age 14 Mean (SD) / N(%)* n=18,955
Males	5,934 (51.2)	9,841 (51.0)
Highest education of parents		
Low	1,309 (11.3)	2,286 (11.8)
Middle	3,511 (30.3)	6,025 (31.2)
High	4,368 (37.7)	6,735 (34.9)
Unknown	2,401 (20.7)	4,257 (22.1)
At least one parent on social welfare subsidy	991 (8.6)	1,755 (9.1)
Parents not living together	2,512 (21.7)	4,720 (24.5)
Secondary school program level		
Bridge class	na	934 (4.8)
Secondary vocational education (VMBO)	na	7,772 (40.3)
Secondary professional education (HAVO)	na	4,705 (24.4)
Secondary scientific education (WVO)	na	5,731 (29.7)
Missing	na	161 (0.8)
SDQ total score	5.81 (5.11) [0-39]	9.34 (5.01) [0-32]
SDQ Emotions subscale	1.63 (1.89) [0-10]	1.96 (1.90) [0-10]
SDQ Behavior subscale	0.73 (1.19) [0-10]	1.40 (1.19) [0-9]
SDQ Hyperactivity subscale	2.52 (2.45) [0-10]	3.83 (2.24) [0-10]
SDQ Peers subscale	0.93 (1.48) [0-10]	2.15 (2.05) [0-10]
SDQ pro-social behavior subscale	9.06 (1.35) [0-10]	8.29 (1.55) [0-10]
School dropout at age 17		
Graduated or studying	10,566 (91.0)	18,230 (96.0)

Dropped out	508 (4.4)	680 (3.6)
Missing	515 (4.4)	85 (0.5)

* For continuous / categorical variables

Predictors of school dropout at age 10

In the model with factors measured at age 10, SDQ total score was a significant predictor of future school drop out with OR 1.07 [1.05;1.09] for each additional point of SDQ score. Boys were more likely to drop out than girls (OR 1.37 [1.11;1.69]). Family socio-economic disadvantage had a pronounced effect on kids chances to have left the school without diploma by the age of 17. In particular, children of parents with low education had OR 2.37 [1.72;3.28] to drop out compared to children of highly educated parents (Table 2). Children of parents whose education could not be obtained from the registry (approx. 20% of all children) were not statistically significantly different from children whose parents had high or middle education, indicating no specific pattern in the missing data in relation to the outcome. Immigration background did not show a significant association with the outcome, nor did it confound the relationship between other factors with the outcome. Interaction with gender was significant but not relevant after stratification. Hosmer and Lemeshow test did not indicate good fit (p-0.03), with upper groups being poorly approximated by the predicted model counts (Figure S1, left). The AUC value of this model was 0.697 (Figure 1) indicating that in approx. 70% of cases that model could provide an accurate prediction of future drop out. Sensitivity and specificity trade-off is presented in figure S2 (online appendix). Positive and negative predictive values for the two potential cut-offs (0.07 and 0.20, with range being 0.01-0.44) are presented in Table 5. SDQ have resulted in slight improvement of the prediction properties of the model (AUC without SDQ = 0.66). Further calculations of marginal probabilities of drop out for children with a number of risk factors (male, low educated parents, parents not living together, at least one parent on welfare subsidy) revealed that accounting for SDQ score substantially calibrated the predictions which ranged from 6% to 16% for children with lowest and highest SDQ scores, respectively (Table 4).

Table 2. Association between socio-demographic and family factors and SDQ at age 10 with school dropout at age 17. Results from multilevel logistic regression model.

Factor (measured at age 10)	OR [95% CI]
SDQ total score	1.07 [1.05; 1.09]
Male vs female	1.37 [1.11; 1.69]
Highest education parents	
Low vs high	2.37 [1.72; 3.28]
Middle vs high	1.51 [1.14; 2.00]
Unknown vs high	1.31 [0.94; 1.84]
Parents not living together	1.59 [1.25; 2.01]
At least one parent receiving state welfare subsidy	1.57 [1.16; 2.13]

Predictors of school dropout at age 14

Significant interaction was detected between SDQ and type of secondary school (p<0.01), therefore analyses were stratified by secondary school level. Quite diverse patterns by school level were revealed. SDQ remained a predictor of dropout among the secondary vocational education (VMBO) and secondary professional education (HAVO) students, and was not relevant for bridge class or the highest

secondary school level (VWO). Differences between boys and girls were only observed in VMBO students, and parental disadvantage (low education and single households) were only predictive of school dropout among secondary vocational education (VMBO) students and to some extent in the bridge class. Parents reliance of social subsidies remained an important factor across all levels except the small group of children attending bridge class (Table 3). Immigration background did not show a significant association with the outcome, nor did it confound the relationship between other factors with the outcome. Interaction with gender was not statistically significant. ROC of model among secondary vocational education (VMBO) students is presented in Figure 1 (AUC value 0.69). Hosmer and Lemeshow model fit test yielded p-value of 0.65. AUC value of models in other strata's was between 0.54-0.57, indicating very poor predictive performance. Sensitivity and specificity trade-off is presented in figure S2 (online appendix). Positive and negative predictive values for the two potential cut-offs (0.07 and 0.20, with range being 0.02-0.35) are presented in Table 5. Adding SDQ to the prediction model yielded a slight improvement above the model with only gender and socio-economic family risk factors (AUC 0.66). Marginal predictions for a disadvantaged VMBO student (education parents=low, parents not living together, at least one parent on welfare subsidy) could be refined to a range 10-16% (as opposed to an average 13%) when accounting for lowest and highest decile SDQ, respectively (Table 4).

Table 3. Association between socio-demographic and family factors and SDQ at age 14 with school dropout at age 17, stratified by school type. Results from multilevel logistic regression model.

Factor (measured at age 14)	Bridge class n=924	Secondary vocational education (VMBO) n=7,720	Secondary professional education (HAVO) n=4,681	Secondary scientific education (VWO) n=5,696
	OR [95% CI]			
SDQ total score	1.03 [0.99;1.08]	1.04 [1.03;1.06]	1.06 [1.01; 1.11]	1.07 [1.00; 1.14]
Male vs female	1.29 [0.78;2.13]	1.61 [1.34;1.92]	1.24 [0.80; 1.94]	0.84 [0.45; 1.55]
Highest education parents				
Low vs high	2.19 [1.02;4.71]	2.00 [1.50;2.65]	1.67 [0.81;3.46]	0.34 [0.04; 2.58]
Middle vs high	1.56 [0.77;3.14]	1.31 [1.00;1.70]	0.86 [0.48;1.52]	0.42 [0.16; 1.11]
Missing vs high	1.09 [0.47;2.53]	1.39 [1.03;1.88]	1.25 [0.69; 2.26]	1.09 [0.52; 2.28]
Parents not living together	1.63 [0.97;2.72]	1.37 [1.14; 1.66]	1.26 [0.77; 2.07]	1.47 [0.70; 3.06]
At least one parent receiving state welfare subsidy	1.15 [0.59;2.23]	1.62 [1.28;2.04]	2.07 [1.07;4.01]	2.97 [1.11; 7.98]

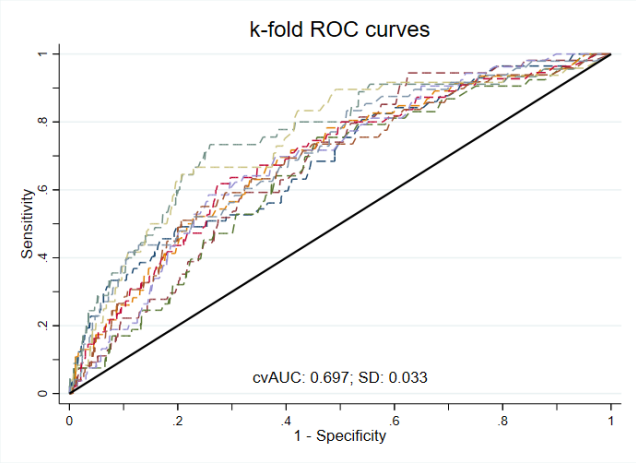

Figure 1. ROC curves for models with predictors at age 10 (left) and age 14 (secondary vocational education) level (right)

Table 4. Marginal probabilities of school-drop out by deciles of SDQ. Derived from regression models.

SDQ	Marginal probability [95% Confidence interval]	
	Age 10 *	Age 14 **
Bottom 10th decile*	0.07 [0.05; 0.10]	0.10 [0.07; 0.13]
Median*	0.10 [0.07; 0.13]	0.13 [0.10; 0.16]
Top 10th decile*	0.16 [0.11; 0.20]	0.16 [0.12; 0.20]
No SDQ in the model	0.13 [0.09; 0.16]	0.13 [0.10; 0.17]

* Prediction for male, education parents=low, parents not living together, at least one parent on welfare subsidy; SDQ bottom 10th decile=0, median=5, top 10th decile=13;
** Prediction for male, secondary vocational education (VMBO) study level, education parents=low, parents not living together, at least one parent on welfare subsidy; SDQ bottom 10th decile=4, median=10, top 10th decile=17

Table 5. Positive and negative predictive values.

Cut off	Model for age 10		Model for age 14 Secondary vocational education	
	PPV	NPV	PPV	NPV
0.07	13.82%	96.98%	11.30%	95.48%
0.20	21.43%	95.57%	16.76%	93.90%

PPV – positive predictive value (true positives/positive calls); NPV – negative predictive value (true negatives / negative calls)

Discussion (1240 words)

In this paper we (1) assessed whether SDQ at age 10 and 14 is predictive of future school dropout, and (2) explored whether a useful prediction model for school dropout can be developed using the factors routinely accessible to YHC professionals during regular health check contact moments for children aged 10 and 14.

As hypothesized, higher SDQ was indeed predictive of school dropout, with OR to drop out between 1.04-1.07 per each point of SDQ. Combining SDQ with all socio-economic and health factors available to YHC at contact moment, a prediction model could be developed for children attending the lowest level secondary school at age 14 (VMBO) as well as for a total group of 10 year-olds. Performance of both prediction models was moderate with AUC approx. 0.7. The strengths of these prediction models is that they only include a few health and socio-economic factors that are being routinely registered in YHC records. While a large number of false positives makes current models unsuitable for any intensive and costly interventions, the simplicity of the models and the fact that they concern a prediction 7 and 3 years ahead of the complex social outcome, the models could be used as a first step in risk stratification among the children population and an indication to add school performance to the regular consultation agenda.

In addition to slight improvement in prediction (AUC improvement by 0.03 in both final models), SDQ offers the YHC professionals an additional grip to assess risk of school dropout. Such assessment can facilitate risk differentiation within the 'disadvantaged group' of children with adverse family circumstances that are known to be predictive of school dropout. Namely, children with 'adverse' background who also have high (i.e. top decile) SDQ have more than double risk of school dropout compared to similar peer with normal SDQ (probability ranges between 16 and 6%, respectively). In the context of limited resources, this offers an opportunity for a more targeted interventions.

Future research involving all relevant stakeholders including schools, families and YHC professionals should evaluate potential barriers and facilitators for its use in practice as a first step in risk stratification and subsequently developing intervention strategies to address or mitigate risk factors. A particular caution is needed around communication about the goals and procedures of such prevention strategies, to avoid stigmatization around SDQ scores that may otherwise compromise the validity of the data and discount potential benefits.

Two important predictors of school dropout were not included in the current models, namely diagnosed mental health problems and school performance. While mental health problems (e.g. ADHD, ADD) are partially captured by SDQ scores available to YHC professionals and used in the model, a more accurate anamnesis of mental health diagnoses is expected to improve the prediction.(23) School achievement and grades is another known strong predictor of future early school leaving.(24) At the moment, these data are not routinely available to YHC professionals and therefore including them in the model has no practical value. Future research and field-work efforts in tackling early prevention of school dropout should include attempts to make the reliable data on children's mental health and school achievement available to YHC at the moment of consultation, at least to the group that has been classified as 'at risk' by the initial prediction model.

While both models showed comparable predictive value, the model with predictors at 10 years old may be preferred over the model with predictors at 14 years old. In general, earlier prevention is desirable when it is possible, and assessing risk of drop out at age 10 offers a larger window of opportunity to engage with school and family, and, where possible, anticipate better on the choice of secondary school level. On the contrary, 14 years represent puberty age where interventions to re-engage with school

may be more challenging. The children of this age have been already differentiated by the levels of education (ranging from vocational education to scientific programs, and the division is done based on performance at primary school and primary school advice). Majority of participants of secondary vocational education will stream into senior vocational programs (MBO), and these are known to be at higher risk of future dropout occurring at and after this transition.(25) Nonetheless, the proposed model allows to further differentiate within this already vulnerable group.

Interestingly, students with migrant background were not at higher risk of school dropout after correction for other confounders that included proxies for parental socio-economic status. Similar findings have been observed in earlier research and may indicate that higher dropout in migrant groups occur primarily due to socio-economic family disadvantage.(26) However, this finding should be interpreted with caution given the rapidly changing migration landscape in recent years. So this factor should not be discounted in future research efforts to improve and validate the model. Other factors that were predictive of school dropout in our model, namely male gender, low parental education and reliance on state subsidies as main source of family income, as well as living in a single-parent household, have been also identified as predictors of dropout by several other studies.(7, 26, 27)

The findings of this research should be interpreted in a view of few limitations. Firstly, given a young age of the cohort at the moment of assessment (the most recent data on school dropout available for ages 17 to 22), definition of school dropout used in this study is not in line with the formal definitions(5, 28), and is therefore only a proxy for potential problems with obtaining start qualification in the future.(2) Secondly, extensive linkage with national population registries allowed to explicitly assess the potential selection bias in the study population. In particular, children who did not complete the questionnaire sent by YHC (and thus for whom SDQ was not available) had somewhat higher future school dropout rates. This is not a surprising finding given that most vulnerable population groups are known to have lower participation rates in public health programs. Further, SDQ is completed by 14-years-old teenagers themselves (as opposed to parents completing the questionnaire for 10-years-old), which may introduce an additional response bias in this group, when those facing most behavioral and social challenges choose to not participate. Of note, organizational changes occurring in YHC in years 2014-2015 (transition to another digital dossier software) contributed to lower reach in those years, and it is possible that low reach in those years was a temporal setback. Last but not least, the current model did not account for ongoing dropout policies at school level and intensity of already provided support at different schools. Nonetheless, within-school correlation was factored in through random intercept per school with resulting intra-class correlation of 7% which suggests not strong clustering in the data.

In conclusion, we have demonstrated that school dropout can be to some extent anticipated based on a few factors including SDQ score and adverse family circumstances, as early as age of 10. The developed prediction model can contribute to early risk stratification where window of opportunity exists for interventions aimed to strengthen ties with school. Current model can provide a first signal of potential problem and set the school performance on the agenda in consultations between YHC, parents, children and school. Future research involving all relevant stakeholders is needed to improve the model as well as assess barriers and facilitators for using the prediction model in daily YHC practice as a first step for a deeper analysis of the problem at stake.

Online supplementary material

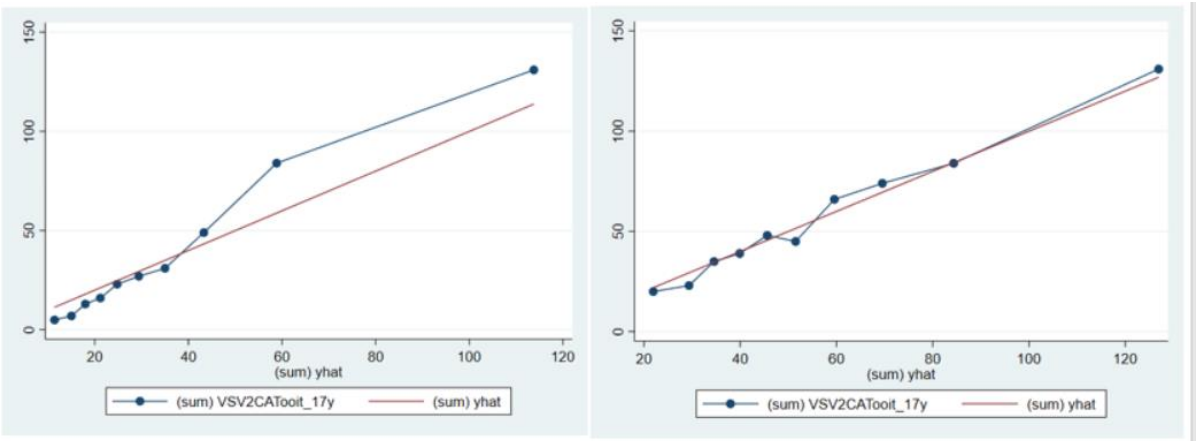

Figure S1. Graphical presentation of Hosmer and Lemeshow test (predicted vs observed counts). Model age 10 (left) and model age 14 (secondary vocational school (VMBO), right)

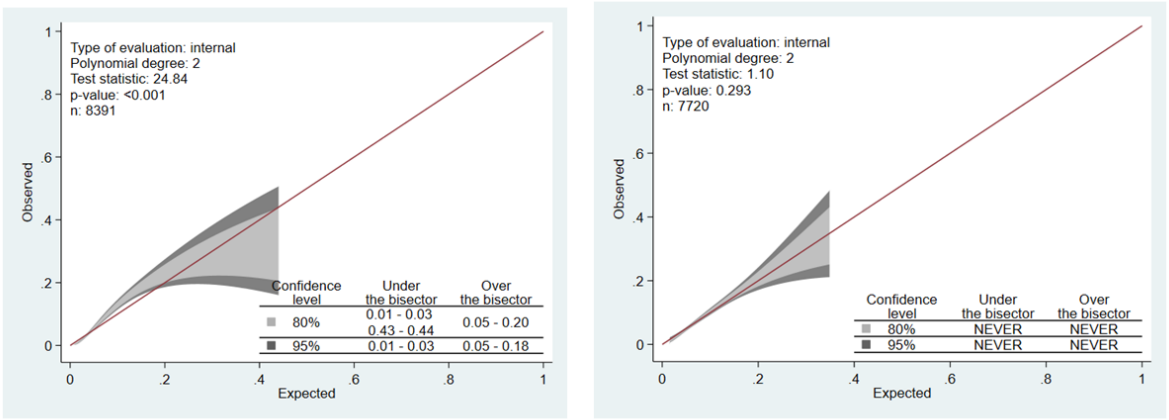

Figure S2. Calibration plots for model age 10 (left) and model age 14 (secondary vocational school (VMBO)) (right)

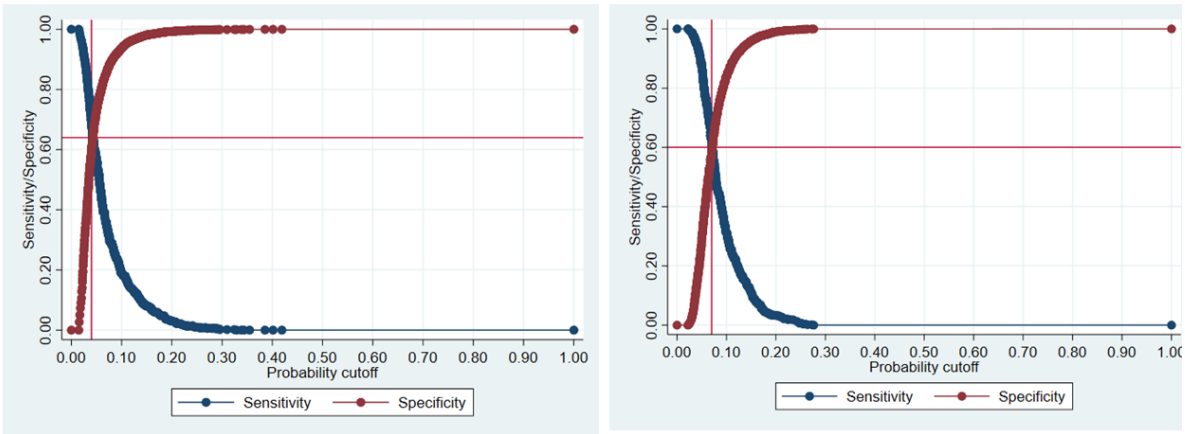

Figure S3. Sensitivity and specificity of model age 10 (left) and model age 14 (secondary vocational school (VMBO) (right)

References

1. CBS. Voortijdig schoolverlaten, werkloosheid en delinquentie: cumulatie van risicogedrag onder jongeren in Nederland 2020. Available from: https://www.cbs.nl/nl-nl/achtergrond/2011/51/voortijdig-schoolverlaten-werkloosheid-en-delinquentie-cumulatie-van-risicogedrag-onder-jongeren-in-nederland.
2. Rumberger RW, Lamb SP. The early employment and further education experiences of high school dropouts: a comparative study of the United States and Australia. Economics of Education Review. 2003;22(4):353-66.
3. De Ridder KA, Pape K, Johnsen R, Westin S, Holmen TL, Bjorngaard JH. School dropout: a major public health challenge: a 10-year prospective study on medical and non-medical social insurance benefits in young adulthood, the Young-HUNT 1 Study (Norway). J Epidemiol Community Health. 2012;66(11):995-1000.
4. Onderwijs in cijfers. Voortijdig schoolverlaters (https://www.rijksoverheid.nl/onderwerpen/vsv/minder-voortijdig-schoolverlaters) 2020 [cited 2020 August].
5. Onderwijs in cijfers. Landelijke vsv cijfers (https://www.onderwijsincijfers.nl/kengetallen/onderwijs-algemeen/leerlingen-en-studenten/prestaties-voortijdig-schoolverlaten/landelijke-vsv-cijfers) 2020.
6. National Research Council. High School Dropout, Graduation, and Completion Rates: Better Data, Better Measures, Better Decisions. Washington, DC: The National Academies Press; 2011.
7. Bowers A, Sprott R, Taff S. Do We Know Who Will Drop Out? A Review of the Predictors of Dropping out of High School: Precision, Sensitivity, and Specificity. The High School Journal. 2012;96(2).
8. Kane Salvador S. School Dropout Predictors. Annotated Bibliography. Accessed from: http://www.plancharlotte.org/sites/default/files/pdf/Dropout%20Predictors%20Bibliography.pdf. 2012.
9. Goodman R. The Strengths and Difficulties Questionnaire: a research note. J Child Psychol Psychiatry. 1997;38(5):581-6.
10. Goodman R, Ford T, Meltzer H. Mental health problems of children in the community: 18 month follow up. BMJ. 2002;324(7352):1496-7.
11. Goodman R. Psychometric properties of the strengths and difficulties questionnaire. J Am Acad Child Adolesc Psychiatry. 2001;40(11):1337-45.
12. Stone LL, Janssens JM, Vermulst AA, Van Der Maten M, Engels RC, Otten R. The Strengths and Difficulties Questionnaire: psychometric properties of the parent and teacher version in children aged 4-7. BMC Psychol. 2015;3(1):4.
13. Mieloo CL, Bevaart F, Donker MC, van Oort FV, Raat H, Jansen W. Validation of the SDQ in a multi-ethnic population of young children. Eur J Public Health. 2014;24(1):26-32.
14. Vogels AG, Crone MR, Hoekstra F, Reijneveld SA. Comparing three short questionnaires to detect psychosocial dysfunction among primary school children: a randomized method. BMC Public Health. 2009;9:489.
15. Theunissen M, Wolff M, van Grieken A, Mieloo C. Handleiding voor het gebruik van de Strengths and Difficulties Questionnaire binnen de Jeugdgezondheidszorg. TNO, 2016.
16. Fortin L, Lessard A, Marcotte D. Comparison by gender of students with behavior problems who dropped out of school. Procedia - Social and Behavioral Sciences. 2010;2(2):5530-8.
17. Hosmer D, Lemeshow S. Goodness-of-fit tests for the multiple logistic regression model. Communications in Statistics—Theory and Methods. 1980;9:1043–69.
18. Nattino G. Assessing the Calibration of Dichotomous Outcome Models with the Calibration Belt (https://www.stata.com/meeting/columbus18/slides/columbus18_Nattino.pdf). Stata Conference - July 19, 20182018.

19. Melo F. Receiver Operating Characteristic (ROC) Curve. In: Dubitzky W, Wolkenhauer O, Cho K, Yokota H, editors. Encyclopedia of Systems Biology Springer, New York, NY https://doiorg/101007/978-1-4419-9863-7_2422013.
20. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29-36.
21. Luque-Fernandez M, Maringe C, Nelson P. CVAUROC: Stata module to compute Cross-validated Area Under the Curve for ROC Analysis after Predictive Modelling for Binary Outcomes, Statistical Software Components S458324. 2017.
22. StataCorp. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC. 2019.
23. Fried R, Petty C, Faraone SV, Hyder LL, Day H, Biederman J. Is ADHD a Risk Factor for High School Dropout? A Controlled Study. J Atten Disord. 2016;20(5):383-9.
24. Balkis M. Academic amotivation and intention to school dropout: the mediation role of academic achievement and absenteeism. Asia Pacific Journal of Education 2018;38(2):257-70.
25. Allen J, Meng C. Risk Factors for School Absenteeism and Dropout: A Meta-Analytic Review. ROA-R-2010/9. Maastricht: 2010.
26. Traag T. EARLY SCHOOL-LEAVING IN THE NETHERLANDS A multidisciplinary study of risk and protective factors explaining early school-leaving. Accessed on 8 Sept 2020 from http://eslplus.eu/documents/Tanja_Traag_Early_school_leaving_in_the_Netherlands.pdf. In: Netherlands S, editor. 2012.
27. Gubbels J, van der Put CE, Assink M. Risk Factors for School Absenteeism and Dropout: A Meta-Analytic Review. J Youth Adolesc. 2019;48(9):1637-67.
28. European Comission. Early school leaving in Europe – Questions and answers 2020 [cited 2020 August]. Available from: https://ec.europa.eu/commission/presscorner/detail/en/MEMO_11_52.