Normering en actualisering van het Halt-signaleringsinstrument (Halt-SI) en beschrijving van de actuele populatie Halt-deelnemers

Summary

Onderzoek uitgevoerd in opdracht van het WODC,

Ministerie van Justitie en Veiligheid

Dr. C. E. van der Put, Universiteit van Amsterdam

I. J. Stolwijk, MSc, Universiteit van Amsterdam

Amsterdam, juli 2022

Summary

Introduction

On a yearly basis, 10 to 15 thousand juveniles receive a Halt sanction. The Halt sanction is an extrajudicial intervention with the aim of making juveniles aware of their behavior and to prevent recurrence of criminal behavior (Halt Manual, 2020). The underlying idea is that by confronting the juvenile with the consequences of his or her behavior and offering the juvenile behavioral alternatives, the juvenile is prevented from engaging in delinquent behavior again. However, previously conducted effect research has found no effect of the Halt sanction on reducing crime and behavioral problems among youth (Ferwerda, et al., 2006). A possible explanation for this finding is that the principles of the Risk, Need, Responsivity (RNR) model were not yet sufficiently taken into account in the Halt sanction at the time of the study. The RNR model is widely used in correctional services because several meta-analytic studies showed that the effectiveness of correctional interventions is greatest when working according to the principles of this model (Andrews & Bonta, 2010a, 2010b; Andrews & Dowden, 1999).

In order to apply the principles of the RNR model, instruments are needed to assess the risk of recidivism and to identify the criminogenic needs. To this end, Halt uses the Halt-Signaling Instrument (Halt-SI), which is administered during the initial interview. The purpose of the Halt-SI is to gain insight into: (a) the level of recidivism risk, (b) the presence of dynamic risk factors with respect to the juvenile's delinquent behavior, and (c) signs of underlying psychosocial problems and victimization of domestic violence and child abuse. Based on the outcome of the Halt-SI, it is determined what type of Halt sanction will be given (i.e., which Halt-modules and learning assignments), as well as whether any additional referral to social services and/or a report to Child Protective Services (CPS) is necessary.

Since the development of the Halt-SI in 2009, no research has been conducted into its psychometric properties and therefore the present study examined the predictive validity of the Halt-Signaling instrument (Halt-SI) with respect to predicting reoffending (research aim 1). In addition, the current study examined how the Halt-SI could be updated with respect to risk- and needs assessment, referral to modules and learning assignments, and the identification of psychosocial problems and victimization of domestic violence and child abuse (research aim 2). The third objective was to map the composition of the current population of Halt participants in order to gain more insight into the underlying problems of Halt participants and the severity of the offenses for which youth are referred to Halt.

Research questions

Given the aforementioned research objectives, the current study aimed to answer the following research questions:

<u>Aim 1 Insight into the predictive validity of the Halt-SI</u>

1. How strong is the association between individual Halt-SI items and recidivism (judicial recidivism and renewed police contacts)?
2. What is the predictive validity of the Halt-SI with the current weighting of items based on US data (sumscore and risk classification)?

3. (a) What is the predictive validity of the new scoring based on Dutch data that takes into account the strength of the bivariate association between items and recidivism and (b) What is the predictive validity of the sumscore based on this scoring?
4. What is the predictive validity of the sumscore if items are equally weighted?
5. What is the predictive validity of the sumscore if a logistic regression analysis is used to determine the weighting of items?
6. (a) Which method of weighting has the highest predictive validity and (b) what are the new optimal cut-off scores for identifying risk categories (no/little risk, some risk, high risk)?
7. To what extent are there differences between: (a) boys and girls, (b) 12-13 year olds and 14-18 year olds and (c) regular Halt and Halt-plus participants, in the strength of the relationship between (I) the individual Halt-SI items and recidivism and (II) the Halt-SI sumscore/Halt-SI risk classification and recidivism? (In other words, to what extent is a separate scoring method necessary for these different subgroups?)
8. What is the predictive validity of the current assessment of: (a) other psychosocial problems and (b) victimization of child abuse and/or domestic violence with regard to predicting future child protection measures and residential youth care?
9. To what extent can these assessment be improved by actuarial risk classification?

Aim II Updating the Halt-SI

10. What recommendations can be made regarding updating the Halt-SI needs assessment (shortening, expanding, or modifying current items)?
11. What recommendations can be made regarding updating the risk assessment of the Halt-SI in a way that increases predictive validity or at least does not reduce it?
12. Which recommendations can be made with regard to updating the assessment of psychosocial problems and victimization of domestic violence and child abuse so that this identification is (better) adapted to current insights in the field of adolescent development?

Aim III To understand the composition of the current population of Halt participants

13. a) How can the current population of Halt participants be characterized in terms of demographic characteristics, offending behavior, recidivism risk, dynamic risk factors and psychosocial problems (including indications of victimization of domestic violence and child abuse)? And (b) What shifts have occurred in these characteristics in recent years?
14. What differences are there in psychosocial characteristics (demographics, risk and protective factors, psychosocial problems) between Halt participants with no/little, some, or a high risk of recidivism? (Based on the insights from these analyses, the use of Halt activities during the Halt sanction can be more tailored).

Methods

The following two research groups were used for this study:

I Halt-participants 2017-2018

To answer the research questions from Aims I and II (investigating predictive validity and updating the Halt-SI), data collected with the Halt-SI during the period January 2017 through December 2018 (N = 23,735)

were used. This is because these research questions require a sufficiently long follow-up period to also include recidivism data and data on future child protection measures and residential youth care.

II Halt-participants 2019-2021

To map the current population of Halt participants (Aim III), an additional research group was examined, consisting of Halt participants during the period January 2019 through April 2021 (N = 24,113 youth).

Halt-SI

The Halt-SI consists of three parts (see Appendix I). The first part is aimed at identifying dynamic risk factors and estimating the risk of recidivism. This part consists of 17 items that include school performance, truancy, leisure activities, peer interaction, obedience, substance use, and impulsivity. After the items from this part of the Halt-SI are scored, the scores from all the individual dynamic risk factors are summed, resulting in a sumscore. Based on this sumscore, it is then determined in which risk group the juvenile falls, using the following three classifications: 0 to 26 points (little/no risk), 27 to 61 points (some risk) and 62 points or more (high risk). This classification is based on US data.

The second part focuses on identifying signs of: (a) additional psychosocial problems and (b) child abuse and/or domestic violence. This part consists of items that reflect concerns that are not (directly) related to the risk of recidivism, namely: mood problems, victimization of neglect, victimization of physical abuse, witness of domestic violence, and signs of sexual abuse. Based on these items, an assessment is made of signs of additional psychosocial problems (classification into 'none/little', 'some', and 'a lot') and of signs of domestic violence and/or child abuse (classification into 'no', 'yes, in the past', and 'yes, at this moment).

The third part of the Halt-SI examines to what extent referral to social services and/or a report to CPS is necessary. Referral to social services is based on the presence of dynamic risk factors, signs of other psychosocial problems and/or a request for help from the youth/parents. A report to CPS is based on signs of domestic violence and/or child abuse.

Outcome measures

In this study, the following three outcome measures were used (in a follow-up period of 2 years): (a) recidivism based on police data, (b) recidivism based on judicial data, and (c) future psychosocial problems and victimization of child abuse and/or domestic violence (child protection measure, youth outpatient care or residential youth care)..

All outcome measures were obtained from Statistics Netherlands (in Dutch: Centraal Bureau voor de Satistiek [CBS]). Data from the Halt-SI were linked to CBS data. These data were obtained anonymously, and were analyzed in the secure digital working environment of the CBS.

Analyses

Part 1 Predictive validity of the Halt-SI

The predictive validity of the current Halt-SI sum score and risk classification (based on US data) for predicting recidivism was examined by calculating Area Under the receiver-operating-characteristic Curve (AUC) values. In addition, AUC values were calculated to examine the association between the Halt-SI items and recidivism.

For standardization based on Dutch data, a new scoring method was calculated for each item based on the strength of the items' association with recidivism. For each response category of each item, the difference between the (average) recidivism and the recidivism per category was calculated (see section 3.3 for an example of the method). One half of the sample was used to determine this scoring (training sample). The other half of the sample (validation sample) was used for: (a) determining the predictive validity of the total score based on this scoring method and (b) determining the strength of the relationship between the items and recidivism.

Then, on the basis of the AUC values, the predictive validity of a number of other weighting methods was examined, namely: (a) equal weighting, (b) weighted weighting with protective points (c) weighting based on a logistic regression analysis and (d) weighting in which protective points are assigned to the protective response categories of the items (instead of the score '0' as is the case in the current Halt-SI). For all analyses, the total study sample was again randomly divided into two groups of equal size, one to perform the analysis on (training sample) and one to calculate the AUC value of the result of the analysis (validation sample).

A Chi-squared Automatic Interaction Detector (CHAID) analysis was used to determine the optimal cut-off scores for identifying the 'no/little risk', 'some risk', and 'high risk' groups. The CHAID-analysis was performed mainly for illustration purposes and was based on the scoring with equal weighting of items.

In order to investigate to what extent there are differences between: (a) boys and girls, (b) 12-13 year olds and 14-18 year olds and (c) regular Halt and Halt-plus participants in the predictive validity of the individual Halt-SI items and the Halt-SI total score/Halt-SI risk classification, AUC values were calculated separately for the mentioned subgroups.

The predictive validity of the current assessment of: (a) additional psychosocial problems and (b) victimization of child abuse and/or domestic violence, with respect to predicting future child protection measures and (residential or outpatient) youth care, was also examined based on AUC values. To explore the extent to which the assessment could be improved, it was first examined which Halt-SI items were associated with future child protection measures or residential youth care. Next, two risk classifications were developed based on CHAID analyses: (1) a classification for assessment increased risk of future psychosocial problems and (2) a classification for assessment increased risk of future child abuse and/or domestic violence).

Part II Updating the Halt-SI

For the updating of the Halt-SI it was examined (a) which items can be removed from the Halt-SI while maintaining predictive validity, (b) to what extent items should be added to the Halt-SI and (c) to what extent changes in content are needed in the current items of the Halt-SI. The answer to this question was based on the results of (a) the current study, (b) two focus groups and (c) previously conducted studies on the National Set of Instruments fort he Juvenile Justice Method (in Dutch: Landelijk Instrumentarium Jeugdstrafrecht [LIJ]).

Part III Composition of the current Halt population

First, descriptive analyses were conducted to map how Halt participants can be characterized in terms of demographic characteristics, delinquent behavior, dynamic risk factors, psychosocial problems, and classification based on recidivism risk (distribution of participants across the groups 'no/little risk', 'some risk', and 'high risk'), and any shift in these characteristics that occurred over the past four years. In addition, latent class analyses were conducted on the group of current Halt participants (study group II) with the aim of identifying groups with common characteristics.

In order to gain insight into the differences in background characteristics (demographics, risk and protective factors, psychosocial problems) between Halt participants with no/little, some or high risk of recidivism, the prevalences and mean scores of the risk factors were obtained for the different risk groups.

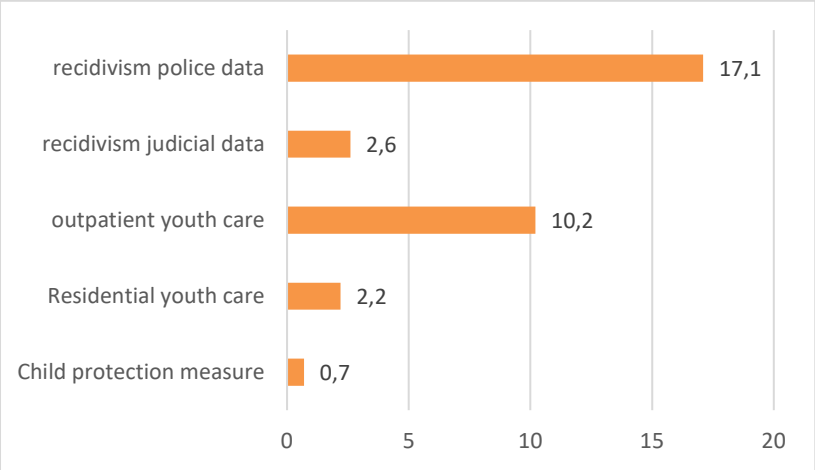
Results

Part I Predictive validity of the Halt-SI

Prevalence of the outcome measures

The figure below shows the prevalence of the outcome measures in a 2-year follow-up period after the Halt-SI was administered.

Prevalence of the outcome measures (%) with a follow-up period of 2 years



The figure shows that in a follow-up period of 2 years after administration of the Halt-SI, 17.1% of the juveniles had a renewed police contact (in the role of suspect of an offence), 2.6% of the juveniles had a judicial contact (conviction of an offence), 10.2% of the juveniles received outpatient youth care, 2.2% of the youngsters received residential youth care (a.o. foster care) and 0.7% had a child protection measure imposed (a.o. guardianship and supervision).

Predictive validity of the Halt-SI for predicting recidivism

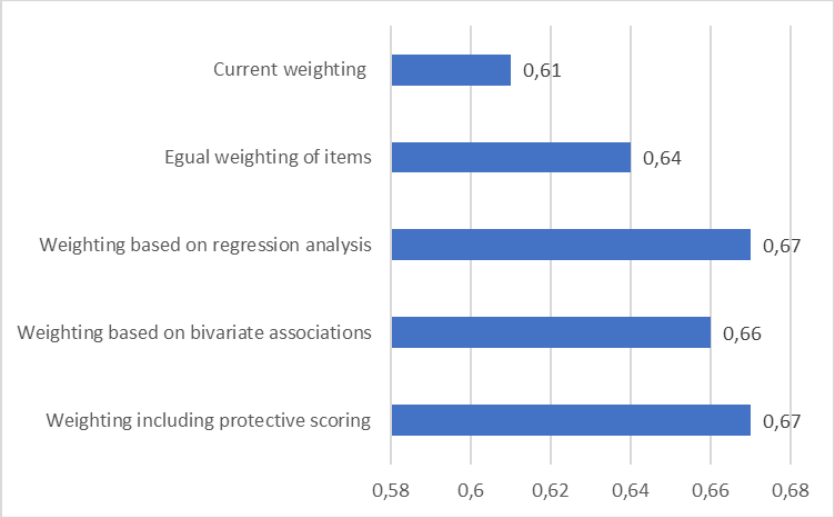
The results showed that although most of the items were significantly related to recidivism, these relationships were in general (very) weak. The predictive validity of the current weighted total score of the Halt-SI (based on US data) for predicting recidivism was weak, with an AUC of .61 for recidivism based on

police contacts and .60 for recidivism based on judicial contacts (both small effects). Moreover, the predictive validity of the current risk classification was found to be very weak with an AUC of .54 for recidivism based on police contacts and .54 for recidivism based on justice contacts (these fall below the .56 threshold of a small effect).

Predictive validity of the new weighting methods for predicting recidivism

The figure below shows the AUC values of the total score based on the different weighting methods investigated, namely: (a) the current weighting (based on US data), (b) equal weighting of the items, (c) weighting based on a logistic regression analysis (based on Dutch data), (d) a weighting based on the bivariate relationship between items and recidivism (based on Dutch data) and (e) a weighting based on the bivariate relationship between items and recidivism in which protective scoring is also assigned (based on Dutch data). In these analyses, recidivism based on police contacts was used as the outcome measure.

AUC-waarden totaalscore gebaseerd op verschillende puntentellingen



The results showed that all four new weighting methods examined had significantly higher predictive validity (medium effects) than the predictive validity of the current scoring method score (small effect). These AUC values are comparable to AUC values of other instruments used internationally for predicting juvenile recidivism.

Assessment of care needs

Predictive validity of the current assessment of care needs

The predictive validity of both the assessment of additional psychosocial problems and the assessment of child abuse and domestic violence was found to be weak for the different outcome measures (AUC values lower than the medium effect limit of .64).

Predictive validity of the actuarial assessment of care needs

The predictive validity of the care assessment could be improved by using a sum score of all Halt-SI items (instead of only the items from the 'care part') for assessment psychosocial problems and future child abuse. The predictive validity of the sum score was found to be higher than the predictive validity of the current care assessment (medium effects for all outcome measures when using the sum score).

Part II Recommendations for updating the Halt-SI

Recommendations on risk classification

Scoring

With respect to scoring, it is first recommended that no longer a distinction is made between boys and girls and different age groups because there are not many differences between these subgroups in the strength of the association between the items and recidivism.

Choice of weighting method

With respect to the weighting method, it is recommended to choose the method with equal weights, despite the lower predictive validity compared to the other weighting methods. The most important reasons for this are that (a) this is in line with Ritax A and B of the LIJ and (b) with equal weights it is easier to add items (such as the suggestions as presented in Table 14 and 15), because in this case it is not necessary to base weightings on the strength of the associations between the items and recidivism and (c) because the score with equal weights is easier to use in practice (more insightful and easier to explain to juveniles and their parents). Although this scoring has a lower predictive validity, the main goal of the Halt-SI is not to estimate recidivism as well and as precisely as possible, but to match the content of the Halt sanction as closely as possible to the identified risk factors.

Establishing the cut-off values for risk classification

After a decision is made regarding the weighting method, the cut-off values can be set for the risk groups (no/low, some and high risk). Here it is not only important to look at the statistically optimal cut-off values, but also to take into account what risk level is considered acceptable and what the consequences of the risk classification are, that is, adhering to the risk- and the need principle from the RNR model (e.g. number of juveniles who receive intensive treatment or are referred to specific modules). Here it is also relevant to consider whether: (a) the groups are sufficiently different from each other in terms of recidivism risk and (b) whether a risk classification into three groups is actually desirable or whether a division into a different number of groups is more adequate given the (policy) decisions that are made on the basis of the division. For example, a classification into three groups is only useful if there are actually three different variants in terms of intensity of Halt treatment.

Recommendations regarding the needs assessment

Changes and adjustments items Halt SI

Based on the results (see also section 3.11) and the focusgroups, we recommend removing at least the following items from the Halt-SI:

- Item 1 (number of days with daytime activities)
- Item 7b (age of first suspension)
- Item 8 (leisurely activities within clubs/associations)

Furthermore, we recommend implementing the proposed changes as shown in Appendix IV.

Adding items to the Halt-SI

Tables 14 and 15 list items that are eligible to be added to the Halt-SI. We advise to make a choice from these, whereby we recommend to limit the number of items to be added (maximum around 5-7 items) because otherwise the instrument will become too long. It is particularly important to consider which items have the most added value in the context of the implementation of the Halt sanction (choice of

modules/learning assignments). If it is desired to add more items, it should be examined to what extent current items can be replaced by new items. Suggestions for this were made in the focus groups (see Appendix VI).

Recommendations regarding assessment of care needs

For the identification of psychosocial problems it is recommended to use the sum score with all risk factors (equal weighting) because the sum score is both a good predictor for future child protection measures and for future (residential and outpatient) youth care. The exact cut-off values for the risk classification should be determined in consultation with Halt. For this purpose, Figure 4 can be used as input.

Part III Composition of the current population of Halt-participants

Characteristics of the current population of Halt participants

Halt participants (in the period January 2018 through April 2021) are on average 15.7 years old and consist of 75% boys. Participants are referred to Halt about equally often due to misdemeanors and felonies. They are most often referred to Halt for a property crime (26%), an "other" crime (24%), school absenteeism (12%) or a public order crime (11%). The following risk factors are most often present among Halt participants: no leisure activities within clubs/associations (30%), impulsiveness (26%), truancy (23%), inadequate punishment by the caregiver(s) for undesirable behavior (20%), and disobedience (19%). Of the Halt participants, 91% fall into the "no/little risk" group, 8.5% into the "some" risk group, and only 0.5% into the "much" risk group (according to the current risk classification). Furthermore, some to a lot of psychosocial problems occur in 17% of Halt participants, with mood problems being relatively common (11%). Finally, 3% of the participants were referred to social services and 0.7% reported to CPS.

Development of Halt participants

Based on the results of the current study it cannot be said whether the group of Halt participants has become 'heavier' in recent years. Halt participants did more often commit a felony offence and less often an misdemeanor offence in recent years, compared to previous years. However, the sum score of the risk factors has remained unchanged. Some risk factors have increased slightly and others have decreased, but because of the corona measures and the transition to a new registration system at Halt it is difficult to determine the exact cause of this.

Subgroups of Halt participants

The latent class analyses showed that the differences between subgroups are mainly in the degree of occurrence of risk factors. Thus, groups do not differ from each other because of the presence of specific (combinations of) risk factors, but because of the extent to which all risk factors occur. This same picture has previously emerged from research on profiles based on LIJ data (Mensink, Hill, & Weijters, 2020).

Differences between low- and high-risk Halt participants

The percentage of girls and the mean age are lower in the high-risk groups than in the low-risk group. In the high risk groups, felony offences were committed more often while in the low/some risk group, misdemeanor offences were committed more often. In the low/some risk group, the offense committed relatively often involved mischief, public order offenses, or fireworks offenses. In the high-risk groups, the offences commited were mainly school absenteeism, vandalism, property crimes and violence. With regard to the risk factors, it can be seen that all factors occur more often in the high-risk groups than in the low-risk

group. An exception to this is the risk factor alcohol use, which is actually lower in the high-risk groups. Some risk factors are (very) common in the highest risk group, such as problem behavior at school (93.7%), truancy (77.6%), suspensions (76.9%), no leisure activities within clubs/associations (74.8%), no parental supervision (77.6%), disobedience (89.5%) and impulsiveness (90.9%). Psychosocial problems are also relatively common in the high-risk group: more than half of the youth show some or many signs of psychosocial problems and 35% show signs of child abuse or domestic violence (now or in the past).

Limitations and strengths of the study

A limitation of the current study was first of all that the (inter-rater) reliability of the Halt-SI has not yet been examined, while reliability is an important condition for good predictive validity. Another limitation relates to the outcome measures used in this study, namely official records of recidivism and care, while official records generally involve underestimation. Nevertheless, official records are considered the "gold standard" for recidivism research because alternative outcome measures (e.g., self-report) have other important drawbacks, such as practical feasibility. Another limitation was that not all the desired information was available to make a proper comparison between Halt participants over the past five years. One reason for this was the transition to a new registration system at Halt. As a result, for example, no (reliable) information was available for the last few years regarding the type of Halt participant (regular versus plus). Also, the classification into offence types was different in recent years compared to previous years so that no comparison could be made regarding, for example, the number of violent offences committed. In addition, the corona measures imposed in the years 2020 and 2021 made it difficult to obtain a good picture of any shifts that have occurred with respect to the composition of the Halt-population. Finally, a limitation is that the effects of the imposed Halt sanction could not be taken into account. The content of the Halt sanction is adjusted to the level of the risk of recidivism and the criminogenic factors present, which may have influenced the level of recidivism. This may have led to an underestimation of the actual predictive validity. However, a research design without intervention was not possible in the current study due to practical and ethical concerns.

Strengths of this research include the fact that a very large dataset was available (more than 47.000 juveniles), containing a lot of information (including all Halt-SI items) and that we were in a unique position to link these data at a personal level to CBS data on police and judicial contacts, imposed child protection measures (guardianship, supervision) and youth care with and without residence (including foster care and outpatient care) in a follow-up period of 2 years. In addition, advanced statistical techniques were applied in order to provide as complete a picture as possible of the predictive validity of the different components of the Halt-SI and the composition of the current population of Halt-participants.

In conclusion

The current study showed that the predictive validity of the various components of the Halt-SI can be improved by modifying the current weightings and risk classifications. Updating the Halt-SI will make the predictive validity of the Halt-SI acceptable and comparable to other risk assessment instruments used internationally. Furthermore, this study shows how the Halt-SI can be standardized and updated so that the RNR principles can be better applied.