



ELSEVIER

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

## Child Abuse &amp; Neglect

journal homepage: [www.elsevier.com/locate/chiabuneg](http://www.elsevier.com/locate/chiabuneg)

## Validation and further development of a risk assessment instrument for child welfare

Annemiek Vial<sup>a,\*</sup>, Claudia van der Put<sup>a</sup>, Geert Jan J.M. Stams<sup>a</sup>, Marc Dinkgreve<sup>b</sup>, Mark Assink<sup>a</sup>

<sup>a</sup> Research Institute of Child Development and Education, University of Amsterdam, Nieuwe Achtergracht 127, 1018 WS, Amsterdam, the Netherlands

<sup>b</sup> Youth Protection Amsterdam Area, Overschiestraat 57, 1062 HN, Amsterdam, the Netherlands

### ARTICLE INFO

#### Keywords:

Predictive validity  
Risk assessment  
Child maltreatment  
Child welfare  
Cumulative risk  
Instrument

### ABSTRACT

**Background:** Although many child maltreatment risk assessment instruments have been implemented in child welfare organizations, thorough studies on their predictive validity are scarce. **Objective:** To examine (1) the predictive validity of a risk assessment instrument that has been widely implemented in the Netherlands, and to examine (2) whether the actuarial risk estimation could be improved and simplified to widen the instrument's applicability to different organizations serving different populations.

**Participants and setting:** The sample comprised risk assessments ( $N = 3,681$ ) performed for families enrolled at one of five child welfare agencies in the Netherlands between January 2015 and December 2017.

**Methods:** In a follow-up period of at least one year, child maltreatment was operationalized as whether or not child protection orders, residential care, or hotline reports occurred. Area Under the Curve values were calculated to determine the predictive accuracy of the risk classifications. Chi-square Automatic Interaction Detection was used to develop a new risk classification based on a new cumulative risk variable.

**Results:** The original risk classification and the newly developed and simplified risk classification showed a similar discriminative accuracy for the different outcome measures: Area Under the Curve values were .68 and .69 for child protection orders, .62 and .63 for residential care, and .58 and .60 for hotline reports, respectively.

**Conclusions:** The original and new risk classification of the instrument had a medium predictive validity with the latter being simpler, more widely applicable, and based on more valid risk factors.

## 1. Introduction

Over time, many risk assessment instruments have been developed to improve child welfare workers' decision making. However, little research on these instruments has been conducted, and the studies that have been conducted leave room for improvement

\* Corresponding author at: Research Institute of Child Development and Education, University of Amsterdam, P.O. Box 15780, 1001 NG, Amsterdam, the Netherlands.

E-mail address: [A.Vial@UvA.nl](mailto:A.Vial@UvA.nl) (A. Vial).

<https://doi.org/10.1016/j.chiabu.2021.105047>

Received 9 August 2020; Received in revised form 17 March 2021; Accepted 22 March 2021

Available online 7 April 2021

0145-2134/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

(Barlow, Fisher, & Jones, 2012; D'andrade, Austin, & Benton, 2008; Knoke & Trocmé, 2005). A recent meta-analytic review identified 27 different risk assessment instruments for child maltreatment and revealed an overall moderate predictive validity (Van der Put, Assink, & Boekhout van Solinge, 2017). For most of these instruments, only one study on predictive validity was available. As the discriminative accuracy of most risk assessment instruments is still insufficient, it is essential to improve these instruments in validation research, so that in the end clinical decision making is improved. The ARIJ is a risk assessment instrument that has been widely implemented in the Netherlands (Actuarial Risk Assessment Instrument for Child Protection; Van der Put, Assink, & Stams, 2016), and is used by many child welfare workers to assess the risk for (the recurrence of) child maltreatment. In this study, we prospectively examined the predictive validity of the ARIJ risk assessments performed under practice conditions. Moreover, we examined whether and, if so, how the predictive validity of the ARIJ could be improved.

In 2015, the ARIJ was developed based on a clinical instrument for risk assessment that was at that time widely used in the Netherlands: the Check List of Child Safety (LIRIK; Ten Berge & Eijgenraam, 2009). The predictive validity of the LIRIK was poor, and the instrument did not perform better than chance in predicting child welfare recidivism (Area Under the Curve [AUC] = .53; Van der Put, Assink et al., 2016). As a large body of research had shown that actuarial instruments outperform clinical instruments in risk assessment (e.g., Baird & Wagner, 2000; Barber, Shlonsky, Black, Goodman, & Trocmé, 2008; Camasso & Jagannathan, 2000; Chaffin & Valle, 2003; Chan, 2012; Coohey, Johnson, Renner, & Easton, 2013; D'andrade et al., 2008; Dorsey, Mustillo, Farmer, & Elbogen, 2008; Johnson, 2011; Van der Put et al., 2017; Van der Put, Hermanns, van Rijn-van Gelderen, & Sondejker, 2016), and a Dutch actuarial risk assessment instrument was not yet available, an actuarial risk classification was developed using the items of the LIRIK that each assessed a potential risk factor for child maltreatment. This risk classification (ARIJ) was only based on the risk factors that significantly predicted child welfare recidivism and performed significantly better than the LIRIK in predicting relapse with a moderate predictive validity (AUC = .63).

Still, the predictive validity was only moderate, and therefore there was a need to further improve the instrument. In the initial development of the ARIJ, multiple so-called "experimental" items were added to the instrument, which all assess a potential risk factor for child maltreatment. These items were not taken into account in the actuarial algorithm but would provide valuable information for future validation research and an opportunity to improve the instrument's validity (Van der Put, Assink et al., 2016). Each experimental item assesses a risk factor that showed to be predictive of child maltreatment in literature reviews, but were not included in the LIRIK, and thus could not be included in the actuarial risk calculation of the ARIJ.

Several large meta-analytic reviews have identified many risk factors for different forms of child maltreatment (Assink et al., 2019; Mulder, Kuijper, van der Put, Stams, & Assink, 2018; Stith et al., 2009). Overall, these reviews have revealed that in particular parent-related risk factors are strong predictors of child maltreatment. Moreover, the sum of the risk factors that are present in a child and its family (i.e., the cumulative risk) has been identified as most predictive of child maltreatment (Brown, Cohen, Johnson, & Salzinger, 1998; Lamela & Figueiredo, 2018; Li, Chu, Ng, & Leong, 2014; MacKenzie, Kotch, Lee, Augsberger, & Hutto, 2011; Patwardhan, Hurley, Thompson, Mason, & Ringle, 2017; Thornberry et al., 2014; Yang & Maguire-Jack, 2018). Cumulative risk has traditionally been described as a linear model in the sense that each additional risk factor increases the risk of future child maltreatment equally (Appleyard, Egeland, van Dulmen, & Alan Sroufe, 2005; Horan & Widom, 2015). However, recent studies have shown evidence for a nonlinear model of cumulative risk (Lamela & Figueiredo, 2018; Patwardhan et al., 2017). This nonlinear quadratic model describes a threshold in the sum of the identified risk factors above which the risk of child maltreatment increases exponentially. This threshold can be used in actuarial risk calculations to identify high-risk groups.

To determine the validity of risk factors or risk classifications, it is necessary to measure occasions or episodes of (future) child maltreatment. However, this is complex and imposes methodological challenges (Mathews, Pacella, Dunne, Simunovic, & Marston, 2020). Prevalence studies on child maltreatment have shown that the number of identified child maltreatment cases greatly depends on the type of measurement that is used (Euser et al., 2013; Stoltenborgh, Bakermans-Kranenburg, Alink, & van IJzendoorn, 2015). Self-reports reveal higher rates of child maltreatment than official reports or substantiated reports (Cyr et al., 2013; Fergusson, Horwood, & Woodward, 2000; Finkelhor, Ormrod, Turner, & Hamby, 2005; MacMillan, Jamieson, & Walsh, 2003). Studies on risk assessment instruments often use multiple outcome measures of child maltreatment, as there is no ideal measure available (Van der Put et al., 2017). The most used outcome measures are: (number of) new child maltreatment reports, substantiated child maltreatment reports, (number of) child maltreatment investigations, supervision orders, or out-of-home placements.

The ARIJ was developed using child welfare recidivism as a measure of future child maltreatment. In the current study, we compared multiple outcome measures in an attempt to measure child maltreatment as accurately as possible. The outcome measures used were reports of child maltreatment or domestic violence, substantiated child maltreatment or domestic violence reports, supervision orders, and out-of-home placement (into foster or residential care). These measures also include child endangerment not related to child maltreatment. For example, a child could be placed in residential care due to severe psychological problems. For clarity, we refer to child maltreatment in this manuscript, but it is important to keep in mind that this also includes child safety problems not related to child maltreatment.

The ARIJ is currently the only actuarial risk assessment instrument for child maltreatment available in the Netherlands, which resulted in an increasing number of agencies using the ARIJ. The ARIJ was originally developed for child protection agencies, but also other types of agencies (e.g. community outreach services) started to use the ARIJ. Therefore, it is even more essential to study the instrument's validity and if possible, improve its validity for an expanding number of users. We therefore included families of five different organizations in the current study, to determine if the instrument can be used by the full chain of child welfare organizations. In sum, the aim of this study was (1) to examine the predictive validity of the items and the risk classifications of the ARIJ and (2) to examine whether the actuarial risk classification could be improved by adding additional risk factors, and (3) to examine whether the risk classification could be simplified to widen the applicability of the instrument to populations served by different types of child

welfare organizations. These aims were studied using multiple prospective measures of future child maltreatment. Additionally, ARIJ risk assessments were retrieved from multiple organizations that serve different populations, to study these aims in an ecologically valid way.

## 2. Method

### 2.1. Sample

Risk assessments performed with the ARIJ were obtained from five Dutch child welfare agencies serving different populations of child and family care needs. In specific, these agencies provide child protection services, child and family support services, or community outreach services. Data on all risk assessments that were performed for children and their families that registered at the agencies between January 2015 and December 2017 were obtained. The selection criteria that were used to determine which risk assessments were included in the study are described in Table 1. This selection procedure resulted in 3681 eligible risk assessments. The children for whom these assessments were performed were on average 9.3 years old ( $SD = 5.6$ ).

### 2.2. Instrument

The ARIJ risk assessment instrument can be used to assess future child endangerment (Van der Put, Assink et al., 2016). The instrument comprises 30 items in total, which are presented in three classes of items: (1) 9 items on the current child safety situation, (2) 14 items on empirically derived risk factors for (the recurrence of) child maltreatment, and (3) 8 so-called “experimental” items that assess potential risk factors for future child maltreatment, and which were added to the originally developed instrument for research purposes. See Table 2 for a short description of each item of the ARIJ. All items can be responded to with one of three response categories: “Yes” (the risk factor is assessed to be present), “No” (the risk factor is assessed to be absent), and “Unknown” (insufficient information is deemed to be available to determine whether a risk factor is absent or present).

Based on the professional’s responses to all items, the instrument estimates an overall risk of future child maltreatment. The instrument also estimates a so-called dynamic risk, which is based on the responses of a professional to only those items in which a dynamic risk factor is assessed (23 of the 30 items). Both risks are expressed as a low, medium, or high risk. In a study on reliability of the ARIJ, the items generally showed moderate reliability, whereas the reliability of the risk outcomes was substantial to almost perfect (Authors’ citation). Additionally, the reliability of the instrument showed to be consistent for cases and professionals of two different agencies as well as for raters with different experience levels.

### 2.3. Outcome measures

Future child endangerment was measured using four different official and prospective outcome measures. The first outcome measure was the presence of a new child protection order. In the Netherlands, an agency can notify the child protection council if there are concerns for a child’s safety, because caregivers are not able or willing to cooperate with child welfare officials. The council will investigate the child’s safety and may submit a request for a court order to be issued, if deemed necessary. If the court rules that the child needs to be protected, a child protection order will be issued. The second outcome measure was the presence of residential care (including foster and residential care). This care can be provided voluntarily and in agreement with the child’s legal guardians but can also be issued by a court. For these first two measures, the national data provided by Statistics Netherlands were used. All Dutch child welfare agencies are obliged to report to Statistics Netherlands. These two outcome measures therefore contain the most complete information on child protection orders and residential care available.

The third outcome measure was the presence of a report of child maltreatment or domestic violence at the hotline services. This measure only comprised official reports made for the child and its family, if someone consulted the hotline services about the family, but not made a report, this was not included in the outcome measure. The fourth measure was the presence of a by the hotline services substantiated report of child maltreatment or domestic violence. The data for these last two measures were provided by the hotline

**Table 1**  
Risk Assessment Selection Criteria.

- 
- Only risk assessments performed for children living in their original living situation with their caregivers were included<sup>1</sup>
  - Only risk assessments with no more than 5 missing items or 8 ‘unknown’ items were included
  - Only risk assessments with a follow-up time<sup>2</sup> of at least one year were included
  - If assessments were performed for multiple children in a family, only the assessment for the youngest child was included
  - If multiple assessments were performed for a single child, only the first performed risk assessment was included
- 

*Note.* Only risk assessments that meet the criteria described in this table were included in this study.

<sup>1</sup> For example, risk assessments performed for residential of foster care situations were excluded.

<sup>2</sup> The follow-up time was corrected for the time a child spent in foster care or residential care during the follow-up period. This correction was necessary to prevent that the child was incorrectly registered as safe during this period, because no ‘new’ child maltreatment in the child’s original living situation could occur when a child was placed in out-of-home care. To determine the follow-up time, the number of days spent in care was subtracted from the total number of days in the follow-up time.

**Table 2**  
Prevalence of Risk Factors and the Risk Outcomes of the ARIJ (N = 3,681).

Risk factors assessed in the ARIJ	%	
<i>Current child safety situation</i>		
Recent physical abuse of the child	5.7	
Recent sexual abuse of the child	1.4	
Recent psychological abuse of the child	8.8	
Recent neglect of the child	10.4	
Recent domestic violence in the family	20.7	
Concerns about parental protection and safety	22.7	
Concerns about insufficient basic care	6.1	
Concerns about rules and boundaries set by the caregivers	33.3	
Concerns about stability in caregiving	37.2	
<i>Risk factors</i>		
Caregiver has a history of abusing a child	12.7	
Caregiver has a psychiatric disorder	25.8	
Caregiver has an addiction	8.1	
Caregiver has a mental disability	4.8	
Caregiver is physically absent	6.4	
Caregiver is emotionally absent	21.4	
Caregiver was maltreated as a child	12.5	
Caregiver has been violent before	14.1	
Caregivers have a problematic relationship	44.1	
Family conflict	38.2	
A history of domestic violence	31.1	
Parental stress about financial problems	38.0	
Social isolation	15.0	
Child is younger than 5 years	27.9	
<i>Experimental items</i>		
Authoritarian caregiver	10.6	
Caregiver has anger management problems	23.0	
Caregiver underestimates the parenting problems	19.1	
Caregiver is unable or unwilling to change their behavior	14.9	
Caregiver perceives the child as a problem	11.9	
Caregivers experience stressful life events	49.7	
Caregiver has a history of criminal behavior	26.6	
<i>Outcomes of the actuarial risk classifications of the ARIJ</i>		
Overall Risk	Low	44.4
	Medium	16.2
	High	39.3
Dynamic Risk	Low	40.3
	Medium	36.0
	High	23.7

*Note.* The prevalence of the response category “Yes” for all the items of the ARIJ and the actuarial risk classifications (low, medium, and high). For most items, the prevalence of the response category “Unknown” was less than 10 %. For the following items the prevalence of the response category “Unknown” was between 10 % and 20 %: Caregiver has a history of abusing a child, caregiver has an addiction, caregiver has a mental disability, caregiver has anger management problems, and caregiver is unable or unwilling to change their behavior. For three items the prevalence of the response category “Unknown” was higher than 20 %: Caregiver has been violent before (21 %), caregiver has a psychiatric disorder (23 %), and caregiver was maltreated as a child (40 %).

services in Amsterdam.

All outcome measures were registered in the period from January 2015 until July 2019. To ensure that the outcome measures were not the result of the risk assessment, we started measuring them three months after the date of the risk assessment. If an outcome measure was present in these first three months after the assessment, it was not deemed to be a “new” episode of child maltreatment. All outcome measures were registered as 1 = child was endangered in the follow-up period and 0 = child was not endangered in the follow-up period.

#### 2.4. Data analyses

First, the prevalence was calculated for each item assessed in the ARIJ. Second, the correlations (phi-coefficients;  $r_{\phi}$ ) between the items (both the original and experimental items) and each of the four outcome measures were calculated. Third, a new cumulative risk variable was created by adding up all the items that were significantly associated with at least one of the outcome measures. Another new cumulative risk variable was based on only the items measuring dynamic risk factors that were significantly associated with at least one of the outcome measures. Fourth, the discriminative accuracy of the original ARIJ risk classifications were determined by calculating the “Area Under the Curve” (AUC). The AUC values were calculated separately for each outcome measure. An AUC value of

.72 or higher corresponds with a large effect size of  $d = .80$  and is therefore viewed as an excellent predictive validity (Rice & Harris, 2005). An AUC value between .63 and .72 corresponds with a medium predictive validity.

The fifth step was to examine whether or not it was possible to improve the overall risk classification by developing a new risk classification. This was done by performing CHAID-analyses (Chi-square Automatic Interaction Detector). In these CHAID-analyses, subsets of assessments with different levels of risk (i.e., different levels of prevalence of the outcome measures) are identified based on the new cumulative risk variable. In other words, the CHAID-analyses identified optimal cut-off scores in the cumulative risk variable.

Only the cumulative risk variable was used to identify different risk levels so that the impact of individual risk factors on the risk classification is minimal. If the impact of individual risk factors on the risk classification is high, a bias could occur when the instrument is used by different agencies, as the prevalence of specific risk factors can vary largely across agencies. Since the instrument is being used by an increasing number of different types of agencies, it is essential that the risk classification produces stable risk estimates across these agencies. An additional reason for only taking the cumulative risk variable into account is that the risk calculation is simplified. This improves the usability of the instrument, as it is easier for professionals to calculate the risk themselves, and it is clearer how child maltreatment risk can be reduced.

Each CHAID-analysis was an *exhaustive* CHAID using chi-square testing and the likelihood ratio method. For each CHAID analysis, ten bootstrap samples were generated to cross-validate the results. A CHAID-analysis was performed for each outcome measure, after which the results were used to develop a new optimal risk classification for all outcome measures combined. For this new risk classification, AUC-values were calculated to determine discriminative accuracy. Last, additional CHAID-analyses were performed to develop a new dynamic risk classification based on the sum of all the dynamic risk factors. Once again, AUC values were calculated to determine the discriminative accuracy of this dynamic risk classification.

### 3. Results<sup>1</sup>

#### 3.1. Prevalence of the risk factors and risk outcomes

Table 2 shows the prevalence of each risk factor that was assessed with the ARIJ as well as the prevalence of the risk outcomes for both the overall and dynamic actuarial risk classification. Of the factors related to the current child safety situation, the following three were the most prevalent: concerns about stability in caregiving (37.2 %), concerns about rules and boundaries set by the caregivers (33.3 %), parental protection and safety (22.7 %), recent domestic violence in the family (20.7 %), recent neglect of the child (10.4 %), and recent psychological abuse of the child (8.8 %). Recent sexual abuse of the child was the least prevalent factor related to the current child safety situation with 1.4 %. The most prevalent risk factors were: caregivers experience stressful life events (49.7 %), caregivers have a problematic relationship (44.1 %), parental stress about financial problems (38.0 %), a history of domestic violence (31.1 %), caregiver has a history of criminal behavior (26.6), family conflict (28.2 %), the child is younger than 5 years (27.9 %), and caregiver has a psychiatric disorder (25.8 %). The prevalence of the outcomes of the overall actuarial risk classification was 44 % for low risk, 16 % for medium risk, and 40 % for high risk, whereas the prevalence was 40 %, 36 %, and 24 % for the outcomes of the dynamic actuarial risk classification.

#### 3.2. Future child maltreatment

Future child maltreatment was assessed by measuring four outcomes in the follow-up period: child protection orders, residential care, hotline reports of child maltreatment or domestic violence, and substantiated reports of child maltreatment or domestic violence. Hotline reports were the most prevalent outcome (15.2 %), followed by child protection orders (8.1 %), residential care (6.1 %), and substantiated reports (1.8 %).

After investigating these separate outcomes, we decided to leave out substantiated hotline reports from further analyses for three reasons. First, the prevalence of this outcome showed to be very low, which is very likely to be a disproportionately low representation of actual cases of child maltreatment. Second, for only approximately 20 % of the reports made to the hotline services an investigation is started, which can possibly confirm the report. For the other 80 % of the reports, a child could have been abused without the services substantiating the abuse. An example of an unsubstantiated report of child abuse is when other services, such as child protection agencies, are already involved in the case that is reported. Third, the associations between the risk factors assessed with the ARIJ and the outcome substantiated hotline reports were not significant ( $r_{\phi} = -.04-.01$ ). Therefore, substantiated reports were not suitable to be considered as a dependent variable in further analyses. This finding is in line with previous research that also questioned the usability of child maltreatment substantiation as a measure of child maltreatment (Hussey et al., 2005).

#### 3.3. Predictive validity of the original ARIJ risk classifications

To determine the predictive validity of the original overall risk classification and the and dynamic risk classification of the ARIJ, AUC values were calculated for the three outcome measures (see Table 3). The overall risk classification did have the highest

<sup>1</sup> All results are based on calculations by the authors of this paper using non-public microdata from Statistics Netherlands and data provided by the different agencies.

discriminative accuracy for the outcome measure child protection orders (AUC = .68), followed by residential care (AUC = .62), and hotline reports (AUC = .58). For the dynamic risk classification, the AUC values were somewhat lower with .62, .60, and .56, for child protection orders, residential care, and hotline reports, respectively.

### 3.4. Associations between risk factors and the outcome measures

Table 4 shows the association between all (potential) risk factors assessed with the ARIJ and the outcome measures child protection order, residential care, and hotline reports. Most assessed factors were significantly associated with all three outcome measures, and all assessed factors were significantly associated with at least one of the outcome measures including all seven experimental items. Therefore, all factors assessed in the ARIJ are risk factors for child maltreatment. The risk factor “Child is younger than 5 years” was negatively associated with residential care. This can be explained by the fact that a very large part of residential care is only available for older children.

### 3.5. The development of a new overall risk classification

Based on the above described associations between the items and the outcome measures, a cumulative risk variable was created by adding up the response “Yes” (scored with 1 indicating that a risk factor was assessed as present) to all ARIJ items that assessed a true risk factor for child maltreatment. All original ARIJ items as well as the experimental items were included in this newly created cumulative risk variable. As a result, the cumulative risk variable describes the sum of 30 risk factors of which 23 were dynamic risk factors.

Next, three CHAID analyses were performed to determine optimal cut-off scores for the cumulative risk variable for the three outcome measures. AUC values were calculated to determine the discriminative accuracy of the different cut-off scores. From the results of the CHAID-analyses and the AUC values optimal cut-off scores of the cumulative risk variable for all outcomes combined could be derived. Fig. 1 shows these cut-off scores along with prevalence of the three outcome measures in each of the three groups. This figure shows that the new risk was low if three or less risk factors were present, which holds for 37.5 % of the children in the sample. In this risk group, 2.5 % of the children received a child protection order, 2.9 % received residential care, and for 10.2 % of the children a hotline report was made in the follow-up period. If four to seven risk factors were present, the risk for child maltreatment was classified as medium, and if eight or more risk factors were present, the risk was classified as high. The prevalence of the risk levels low, medium, and high were approximately the same, with 37.5 % for low risk, 30.6 % for medium risk, and 31.9 % for high risk.

### 3.6. The development of a new dynamic risk classification

For developing a new dynamic actuarial risk classification CHAID analyses were again performed, after which the predictive validity of the different cut-off scores was determined. Here, the cumulative risk variable was only based on a sum of the dynamic risk factors. Instead of three, five risk groups were created, so that relatively small changes in risk over time can also be measured. Fig. 2 shows the new optimal cut-off scores in the cumulative dynamic risk variable. If there were no dynamic risk factors present then dynamic risk is low, if one or two dynamic risk factors were present the dynamic risk was low-medium, if three or four dynamic risk factors were present then dynamic risk was medium, if five, six, or seven dynamic risk factors were present then dynamic risk was medium-high, and if eight or more dynamic risk factors were present then dynamic risk was high.

### 3.7. Predictive validity of the new risk classifications

To determine the predictive validity of the new risk classifications, AUC values were calculated for the three outcome measures separately (see Table 5). The new overall risk classification did have the highest predictive validity for the outcome measure child protection orders (AUC = .69), followed by residential care (AUC = .63) and hotline reports (AUC = .60). The new dynamic risk classification showed a slightly higher predictive validity: AUC = .70 for child protection orders, AUC = .65 for residential care, and AUC = .60 for hotline reports.

**Table 3**

Predictive Validity of the Original Actuarial Risk Classification (N = 3,681).

	Outcome measure								
	Child protection order			Residential care			Report		
	AUC	SE	95 % CI	AUC	SE	95 % CI	AUC	SE	95 % CI
Overall Risk	.68**	.015	[.65–.70]	.62**	.018	[.59–.66]	.58**	.013	[.56–.61]
Dynamic Risk	.62**	.015	[.59–.65]	.60**	.019	[.56–.64]	.56**	.013	[.54–.59]

Note. AUC values of the original risk classifications for the different outcome measures.

AUC = area under the curve; SE = standard error; CI = confidence interval.

\*\*  $p$ -value < 0.01.

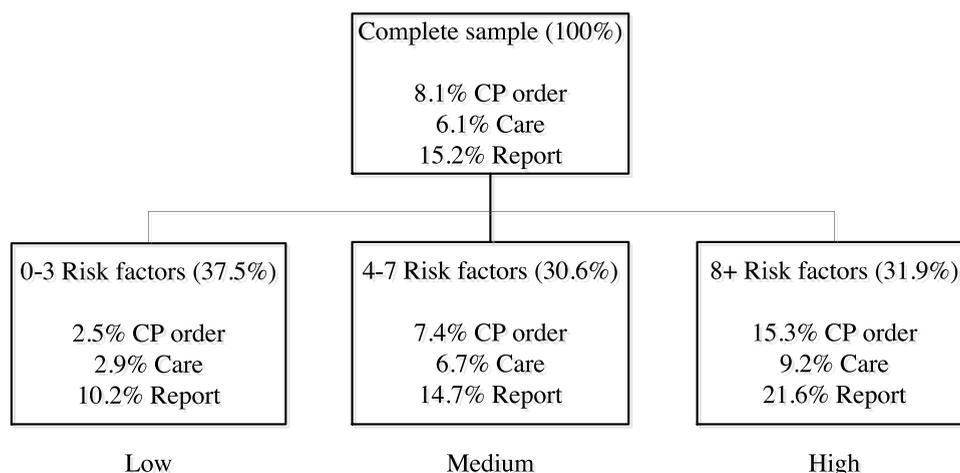
**Table 4**  
Associations (Phi-Coefficients) between the Items and the Outcome Measures (N = 3,681).

Items ARIJ	CP order $r_{\phi}$	Care $r_{\phi}$	Report $r_{\phi}$
<i>Current child safety situation:</i>			
Recent physical abuse of the child	.08**	.14**	.04*
Recent sexual abuse of the child	.02	.06**	.01
Recent psychological abuse of the child	.11**	.11**	.02
Recent neglect of the child	.13**	.12**	.05**
Recent domestic violence in the family	.12**	.06**	.10**
<i>Concerns about parental protection and safety</i>			
Concerns about insufficient basic care	.16**	.15**	.09**
Concerns about rules and boundaries	.06**	.06**	.06**
Concerns about stability	.07**	.15**	.08**
Concerns about stability	.18**	.07**	.08**
<i>Risk factors</i>			
Caregiver has a history of abusing a child	.08**	.08**	.08**
Caregiver has a psychiatric disorder	.05*	.04*	.02
Caregiver has an addiction	.06**	.04*	.07**
Caregiver has a mental disability	.07**	.04*	.08**
Caregiver is physically absent	.03	.07**	.00
Caregiver is emotionally absent	.12**	.09**	.03
Caregiver was maltreated as a child	.14**	.12**	.12**
Caregiver has been violent before	.12**	.01	.09**
Caregivers have a problematic relationship	.12**	-.01	.10**
Family conflict	.17**	.11**	.09**
A history of domestic violence	.13**	.05**	.09**
Stress about financial problems	.03	.02	.08**
Social isolation	.05**	.03*	.03
Child is younger than 5 years	.08**	-.03*	.08**
<i>Experimental items</i>			
Authoritarian caregiver	.08**	.08**	.02
Caregiver has anger management problems	.16**	.09**	.09**
Caregiver underestimates the parenting problems	.11**	.03*	.07**
Caregiver is unable or unwilling to change their behavior	.19**	.09**	.04*
Caregiver perceives the child as a problem	.06**	.18**	.00
Parental stress	.07**	.02	.10**
Caregiver has a history of criminal behavior	.15**	-.02	.08**

Note. Each experimental item assesses a risk factor that showed to be predictive of child maltreatment in literature reviews, but were not included in the LIRIK, and thus could not be included in the actuarial risk calculation of the ARIJ. CP order = child protection order; Care = residential care; Report = hotline report.

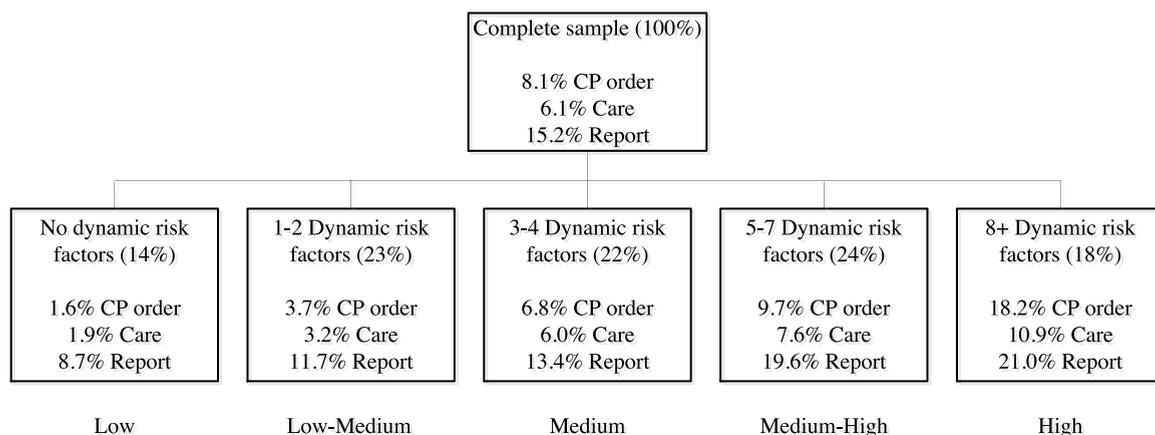
\* p-value < 0.05.

\*\* p-value < 0.01.



**Fig. 1.** Optimal Cut-off scores in the New Cumulative Risk Variable and the Prevalence of the Outcome Measures for each Outcome Measure and Risk Level (N = 3,681).

Note. CP order = child protection order, Care = residential care, Report = hotline report.



**Fig. 2.** Optimal Cut-off Scores in the New Cumulative Dynamic Risk Variable and the Prevalence of the Outcome Measures for each Outcome Measure and Risk level (N = 3,681).

Note. CP order = child protection order; Care = residential care; Report = hotline report.

**Table 5**  
Predictive Validity of the New Risk Classifications (N = 3,681).

	Outcome measure								
	Order			Care			Report		
	AUC	SE	95 % CI	AUC	SE	95 % CI	AUC	SE	95 % CI
New overall risk	.69**	.015	[.66-.72]	.63**	.018	[.59-.66]	.60**	.013	[.57-.62]
New dynamic risk	.70**	.015	[.67-.73]	.65**	.018	[.61-.68]	.60**	.013	[.57-.62]

Note. AUC values of the new risk classifications for the different outcome measures. Order = child protection order; Care = residential care; report = hotline report; AUC = area under the curve; SE = standard error; CI = confidence interval.

\*p-value < 0.05.

\*\* p-value < 0.01.

#### 4. Discussion

The ARIJ is used by many Dutch organizations to assess the risk for future child maltreatment, but a thorough validation study had not yet been performed. Therefore, the first aim of this study was to examine the predictive validity of the items and the risk classifications of the ARIJ in an ecologically valid way by examining ARIJ risk assessments that were retrieved from multiple organizations serving different populations. In this study, three prospective outcome measures of future child maltreatment were examined: whether or not a child protection order, residential care, or a hotline report occurred. The predictive accuracy of the ARIJ risk classification varied from slightly higher to lower (nonsignificant differences) than the accuracy assessed at the time of its development (AUC = .63; Van der Put, Assink et al., 2016). For child protection orders the predictive accuracy was the highest with an AUC value of .68, following by residential care (AUC = .62), and hotline reports (AUC = .58). The predictive accuracy of the dynamic risk classification turned out to be equal or slightly lower (nonsignificant differences) than the accuracy assessed at the time of its development (AUC = .62; Van der Put, Assink et al., 2016). Again, the accuracy was highest for child protection orders (AUC = .62) followed by residential care (AUC = .60), and hotline reports (AUC = .56).

The next aim of this study was to examine whether the actuarial risk classification could be improved by adding risk factors and to examine whether the risk classification could be improved to widen the applicability of the instrument to populations served by different types of child welfare organizations. To do this, we first examined the predictive validity of all risk factors that were assessed with the ARIJ, which included the experimental factors that were added to the first version of the instrument to study their validity. All factors showed to be true risk factors for child maltreatment, as they were all significantly but weakly associated with at least one of the outcome measures. Most factors were even associated with all outcome measures. This result is in line with previous research showing that the factors as assessed with the ARIJ are predictive of child maltreatment (Assink et al., 2019; Cash, 2001; Hindley, Ramchandani, & Jones, 2006; Mulder et al., 2018; Stith et al., 2009; Van der Put et al., 2017).

Since all factors assessed with the ARIJ were identified as true risk factors, a new cumulative risk variable was created by adding up all the risk factors that were determined to be present in a case. In contrast to the risk classifications of the original ARIJ, the new risk classifications were only based on the new cumulative risk variables, so that risk classification was only based on the number of present risk factors and applicable for the populations served by the different organizations using the instrument. The impact of individual risk factors was not considered in the risk classifications. The predictive accuracy of the newly developed overall and dynamic risk classifications was slightly higher (nonsignificant differences) than the accuracy of the original ARIJ risk classifications. The AUC values of

the new overall risk classification were .69 (child protection orders), 0.63 (residential care), and .60 (hotline reports), and these values were .70, .65, and .60 for the dynamic risk classification, respectively. Given these results, a new version of the ARIJ risk assessment instrument could be released in which several risk factors were added and the risk classifications renewed.

Despite the lower number of risk factors in the dynamic risk classification, its predictive accuracy (AUC = .70 for child protection orders) was very similar to the predictive accuracy of the new overall risk classification (AUC = .69 for child protection orders). This can be explained by the number of risk groups in each risk classification. The dynamic risk classification comprised five risk groups, whereas the overall risk classification comprised only three risk groups. Generally, the predictive accuracy of a risk classification increases as the number of risk groups increases, as it gives a more accurate depiction of the scale risk has in reality. Therefore, we also examined the predictive value of an overall risk classification comprising five risk groups, which resulted in an AUC value of .71. Moreover, the predictive value of the overall cumulative risk variable, which is in fact a risk classification comprising 30 risk groups as there are 30 risk factors, revealed an even higher AUC of .72. However, for an optimal usability of the risk assessment instrument in clinical practice, it is important to reduce the number of risk groups without substantially compromising the predictive value of the instrument. Consequently, the final risk classification comprised three risk groups, and showed a slightly lower predictive accuracy (AUC = .69) than the statistically most optimal classification.

An important question is how the predictive accuracy of the ARIJ compares to the predictive accuracy of other risk assessment instruments. In 2017, Van der Put et al. (2017) examined the predictive accuracy of child maltreatment risk assessment instruments in a meta analytic review. The predictive accuracy of the new ARIJ risk classifications is comparable to the predictive accuracy of other actuarial risk assessment instruments for child maltreatment (AUC = .70; Van der Put et al., 2017). When comparing the risk classification of the ARIJ in more detail to risk classifications of other instruments, it is notable that the number of risk factors that need to be present for a high-risk classification is relatively high in the ARIJ risk classification. Previous studies identified a high-risk classification when at least three risk factors (Patwardhan et al., 2017), four risk factors (Brown et al., 1998), or six risk factors (Lamela & Figueiredo, 2018) were assessed to be present. In the current study, the risk was classified as high when 8 or more risk factors were present in a case. This relatively high cut-off may be explained by a conceptual overlap between the risk factors assessed in the ARIJ. This overlap also became apparent in a recent study in which the interrelatedness of risk factors for child maltreatment was explored using network analysis (Authors' citation). A relatively large number of ARIJ items concern the current child safety situation (7 items), or problems related to domestic violence (4 items). Since all items proved to be significantly related to child maltreatment, they were all included in the cumulative risk variable. As a result, a relatively high number of risk factors need to be assessed as present to obtain a high-risk classification. It would be interesting to look into this cut-off score more closely in future research, and to examine how a possibly exponential increase of risk is related to the cumulative risk variable of the risk classifications.

The results of this study have some important implications for the use of the ARIJ instrument in practice. First, the risk classifications produce more stable and valid risk estimates across different agencies, because they are based on the sum of the risk factors assessed as being present (cumulative risk). Second, it is easier for professionals to calculate the risk for child maltreatment themselves, and it is easier to understand how the risk for child maltreatment can be reduced, as all risk factors now contribute to the risk estimate with an equal weight of the estimated risk for child maltreatment. Third and last, the dynamic risk classification now has five instead of three risk groups, enabling measurement of smaller but statistically meaningful changes in risk over time.

Despite the improvement of the dynamic risk classification, it is important to keep in mind that the ARIJ was not developed to use as an instrument for needs assessment. For a thorough and complete needs assessment, it is essential to assess the full range of dynamic risk factors, after which appropriate interventions and treatment can be determined. A new assessment instrument is now being developed for this purpose (Van der Put, Assink, Gubbels, Van Lent, & Stams, 2018). However, the dynamic risk classification of the ARIJ can be used to monitor (changes in) the risk for future child maltreatment over time. Additionally, new dynamic risk factors have been added to the instrument, and therefore the updated instrument now assesses a wider range of family needs. Further research should assess whether and how professionals of different agencies are able to incorporate these new factors in their treatment plan.

It is important to mention some limitations of the current study. First, the risk assessments that were studied were performed at the time of enrolment of the children and their families at the different agencies that participated in this study. As a result, insufficient information was available to assess the presence of a relatively large number of risk factors. It is inevitable that practitioners have more information on the families they supervise as time passes and are therefore better able to assess the presence of all risk factors in subsequent phases of case involvement. A related limitation is that after performing an assessment at enrollment, each agency will do everything in its power to reduce the risk for future child maltreatment, for example, by starting interventions with the aim to improve child safety. Assuming that these interventions focus on (some) risk factors as identified in the initial assessment, and that these interventions are at least effective to some extent, this will reduce the risk for future child maltreatment. Consequently, the predictive accuracy of the ARIJ risk classifications (and the risk factors) may be underestimated. Unfortunately, we did not have information on the care received by the families or whether this care was suitable given the results of the assessment performed at the time of registration. To avoid this intervention effect on the instrument's predictive accuracy that is to be estimated, future research should study the predictive accuracy of ARIJ risk assessments performed in the last phase of the agency's involvement with a family.

Second, official registrations of child maltreatment are often an underestimation of actual numbers of child maltreatment (Cyr et al., 2013; Euser et al., 2013; Fergusson et al., 2000; Finkelhor et al., 2005; MacMillan et al., 2003; Stoltenborgh et al., 2015). Additionally, the relation between occurrences of child maltreatment and the involvement of the child welfare system is complex (Jenkins, Tilbury, Mazerolle, & Hayes, 2017). Therefore, future research should study the predictive accuracy of the ARIJ using other sources than official registrations, such as self-reports or teacher reports, to avoid the bias of the child welfare system.

In sum, a new version of the ARIJ risk assessment instrument with a simplified risk classification based on cumulative risk was developed. Even though the risk classifications were simplified, the predictive accuracy of these new risk classifications was similar to

the original classifications. The newly developed risk classification did not have a predictive accuracy of .72 or higher, which is the criterion for excellent predictive validity. However, the cumulative risk variable did show a predictive accuracy of .72. Future research should study how the predictive accuracy of the risk classifications can be improved further and should use different types of child maltreatment measures, including measures that are not related to the child welfare system.

## Acknowledgments

We thank Altra (in particular Olaf Goorden), Jeugdbescherming Regio Amsterdam (in particular Inge Busschers), de Opvoedpoli (in particular Alfons Crijnen), Samen Doen Amsterdam (in particular Marjan Koopman), Spirit (in particular Carolien Konijn), and Veilig Thuis Amsterdam-Amstelland (in particular Frank Hazeleger and Benne Holwerda).

This work was supported by the Dutch organization for Health research and Development (ZonMW), grant number: 729300108.

## References

- Appleyard, K., Egeland, B., van Dulmen, M. H., & Alan Sroufe, L. (2005). When more is not better: The role of cumulative risk in child behavior outcomes. *Journal of Child Psychology and Psychiatry*, 46(3), 235–245.
- Assink, M., van der Put, C. E., Meeuwse, M. W. C. M., de Jong, N. M., Oort, F. J., Stams, G. J. J. M., ... Hoeve, M. (2019). Risk factors for child sexual abuse victimization: A meta-analytic review. *Psychological Bulletin*. <https://doi.org/10.1037/bul0000188>
- Barber, J. G., Shlonsky, A., Black, T., Goodman, D., & Trocme, N. (2008). Reliability and predictive validity of a consensus-based risk assessment tool. *Journal of Public Child Welfare*, 2(2), 173–195.
- Baird, & Wagner. (2000). The relative validity of actuarial-and consensus-based risk assessment systems. *Children and Youth Services Review*, 22(11–12), 839–871.
- Barlow, J., Fisher, J. D., & Jones, D. (2012). *Systematic review of models of analyzing significant harm*. Oxford University.
- Brown, J., Cohen, P., Johnson, J. G., & Salzinger, S. (1998). A longitudinal analysis of risk factors for child maltreatment: Findings of a 17-year prospective study of officially recorded and self-reported child abuse and neglect. *Child Abuse & Neglect*, 22(11), 1065–1078.
- Camasso, M., & Jagannathan, R. (2000). Modeling the reliability and predictive validity of risk assessment in child protective services. *Children and Youth Services Review*, 22, 873–896.
- Cash. (2001). Risk assessment in child welfare: The art and science. *Children and youth services Review*, 23(11), 811–830.
- Chaffin, M., & Valle, L. A. (2003). Dynamic prediction characteristics of the child abuse potential inventory. *Child Abuse & Neglect*, 27, 463–481.
- Chan, K. L. (2012). Evaluating the risk of child abuse: The child abuse risk assessment scale (CARAS). *Journal of Interpersonal Violence*, 27, 951–973.
- Coohey, C., Johnson, K., Renner, L. M., & Easton, S. D. (2013). Actuarial risk assessment in child protective services: Construction methodology and performance criteria. *Children and Youth Services Review*, 35, 151–161.
- Cyr, K., Chamberland, C., Clément, M.E., Lessard, G., Wemmers, J. A., Collin-Vézina, D., ... Damant, D. (2013). Polyvictimization and victimization of children and youth: Results from a population survey. *Child Abuse & Neglect*, 37(10), 814–820.
- D'andrade, A., Austin, M. J., & Benton, A. (2008). Risk and safety assessment in child welfare: Instrument comparisons. *Journal of Evidence-based Social Work*, 5(1–2), 31–56.
- Dorsey, S., Mustillo, S. A., Farmer, E. M. Z., & Elbogen, E. (2008). Caseworker assessments of risk for recurrent maltreatment: Association with case-specific risk factors and rereports. *Child Abuse & Neglect*, 32, 377–391.
- Euser, S., Alink, L. R., Pannebakker, F., Vogels, T., Bakermans-Kranenburg, M. J., & Van IJzendoorn, M. H. (2013). The prevalence of child maltreatment in the Netherlands across a 5-year period. *Child Abuse & Neglect*, 37(10), 841–851.
- Fergusson, D. M., Horwood, L. J., & Woodward, L. J. (2000). The stability of child abuse reports: A longitudinal study of the reporting behaviour of young adults. *Psychological Medicine*, 30(3), 529–544.
- Finkelhor, D., Ormrod, R., Turner, H., & Hamby, S. L. (2005). The victimization of children and youth: A comprehensive, national survey. *Child Maltreatment*, 10(1), 5–25.
- Hindley, Ramchandani, & Jones. (2006). Risk factors for recurrence of maltreatment: A systematic review. *Archives of disease in childhood*, 91(2), 744–752.
- Horan, J. M., & Widom, C. S. (2015). Cumulative childhood risk and adult functioning in abused and neglected children grown up. *Development and Psychopathology*, 27(3), 927–941.
- Hussey, J. M., Marshall, J. M., English, D. J., Knight, E. D., Lau, A. S., Dubowitz, H., ... Kotch, J. B. (2005). Defining maltreatment according to substantiation: Distinction without a difference? *Child Abuse & Neglect*, 29(5), 479–492.
- Jenkins, B. Q., Tilbury, C., Mazerolle, P., & Hayes, H. (2017). The complexity of child protection recurrence: The case for a systems approach. *Child Abuse & Neglect*, 63, 162–171.
- Johnson, W. L. (2011). The validity and utility of the California family risk assessment under practice conditions in the field: A prospective study. *Child Abuse & Neglect*, 35, 18–28.
- Knoke, D., & Trocme, N. (2005). Reviewing the evidence on assessing risk for child abuse and neglect. *Brief Treatment and Crisis Intervention*, 5(3), 310–327.
- Lamela, D., & Figueiredo, B. (2018). A cumulative risk model of child physical maltreatment potential: Findings from a community-based study. *Journal of Interpersonal Violence*, 33(8), 1287–1305.
- Li, D., Chu, C. M., Ng, W. C., & Leong, W. (2014). Predictors of re-entry into the child protection system in Singapore: A cumulative ecological-Transactional risk model. *Child Abuse & Neglect*, 38(11), 1801–1812.
- MacKenzie, M. J., Kotch, J. B., Lee, L. C., Augsberger, A., & Hutto, N. (2011). A cumulative ecological-Transactional risk model of child maltreatment and behavioral outcomes: Reconceptualizing early maltreatment report as risk factor. *Children and Youth Services Review*, 33(11), 2392–2398.
- MacMillan, H. L., Jamieson, E., & Walsh, C. A. (2003). Reported contact with child protection services among those reporting child physical and sexual abuse: Results from a community survey. *Child Abuse & Neglect*, 27(12), 1397–1408.
- Mathews, B., Pacella, R., Dunne, M. P., Simunovic, M., & Marston, C. (2020). Improving measurement of child abuse and neglect: A systematic review and analysis of national prevalence studies. *PLoS One*, 15(1), e0227884.
- Mulder, T. M., Kuiper, K. C., van der Put, C. E., Stams, G. J. J., & Assink, M. (2018). Risk factors for child neglect: A meta-analytic review. *Child Abuse & Neglect*, 77, 198–210.
- Patwardhan, I., Hurley, K. D., Thompson, R. W., Mason, W. A., & Ringle, J. L. (2017). Child maltreatment as a function of cumulative family risk: Findings from the intensive family preservation program. *Child Abuse & Neglect*, 70, 92–99.
- Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law and Human Behavior*, 29(5), 615.
- Stith, S. M., Liu, T., Davies, L. C., Boykin, E. L., Alder, M. C., Harris, J. M., ... Dees, J. E. M. E. G. (2009). Risk factors in child maltreatment: A meta-analytic review of the literature. *Aggression and Violent Behavior*, 14(1), 13–29.
- Stoltenborgh, M., Bakermans-Kranenburg, M. J., Alink, L. R., & Van IJzendoorn, M. H. (2015). The prevalence of child maltreatment across the globe: Review of a series of meta-analyses. *Child Abuse Review*, 24(1), 37–50.
- Ten Berge, I. J., & Eijgenraam, K. (2009). *Licht instrument risicotaxatie kindermishandeling (LIRIK) [Check List of Child Safety (CLCS)]*. Utrecht: Nederlands Jeugdinstituut.

- Thornberry, T. P., Matsuda, M., Greenman, S. J., Bears Augustyn, M., Henry, K. L., Smith, C. A., ... Ireland, T. O. (2014). Adolescent risk factors for child maltreatment. *Child Abuse & Neglect*, *38*, 706–722.
- Van der Put, C. E., Assink, M., & van Solinge, N. F. B. (2017). Predicting child maltreatment: A meta-analysis of the predictive validity of risk assessment instruments. *Child Abuse & Neglect*, *73*, 71–88.
- Van der Put, C. E., Assink, M., & Stams, G. J. J. (2016). Predicting relapse of problematic child-rearing situations. *Children and Youth Services Review*, *61*, 288–295.
- Van der Put, C. E., Hermanns, J., van Rijn-van Gelderen, L., & Sondejker, F. (2016). Detection of unsafety in families with parental and/or child developmental problems at the start of family support. *BMC Psychiatry*, *16*(1), 15.
- Van der Put, Assink, Gubbels, Van Lent, & Stams. (2018). Risico op kindermishandeling verlagen met ARIJ-needs. *Kind & Adolescent Praktijk*, *17*(3), 16-24. *Kind & Adolescent Praktijk*, *17*(3), 16–24.
- Yang, M. Y., & Maguire-Jack, K. (2018). Individual and cumulative risks for child abuse and neglect. *Family Relations*, *67*(2), 287–301.