The article has been peer-reviewed, but does not include the publisher's layout, page numbers and proof-corrections

Childcare Providers' Nominations of Preschool Children at Risk for Mental Health Problems: Does it Discriminate Well Compared to the Caregiver-Teacher Report Form (C-TRF)?

Kenneth Stensen*, Stian Lydersen*, Frode Stenseng**, Jan Lance Wallander*,***, Håvard Tveit*, and May Britt Drugli*, ****

* Regional Centre for Child and Youth Mental Health and Child Welfare, Faculty of Medicine and Health Science, Department of Mental Health, Norwegian University of Science and Technology (NTNU)

** Department of Education and Lifelong Learning, Faculty of Social and Educational Sciences, Norwegian University of Science and Technology (NTNU) and Queen Maud University College, Trondheim

*** University of California, Merced

**** Centre of the Study of Educational Practice, Inland Norway University of Applied Sciences

Correspondence regarding this article should be addressed to Kenneth Stensen, Regional Centre for Child and Youth Mental Health and Child Welfare, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. Email: kenneth.stensen@ntnu.no

Postal address: Regional Centre for Child and Youth Mental Health and Child Welfare, P.O. Box 8905, MTFS, N-7491 Trondheim

Abstract

Childcare providers are vanguards in identifying children at risk for mental health problems. Thus, the aim of the current study was to investigate the accuracy of childcare providers' nominations of children at risk for mental health problems against a well-established comparator, the Caregiver-Teacher Report Form. Findings from the present study, including 1430 children aged one to six years old and 169 childcare providers from 57 childcare centers, indicates that nominations in the form of concerns should be taken seriously and followed up with additional screening or assessment and consideration for referral. However, nominations also created a considerable portion of false positives. These results suggest that when childcare providers become concerned about a child, it may be beneficial to apply a psychometrically sound screening instrument to decrease the rate of false positive nominations. This may help childcare providers to act more promptly by confirming or discarding their initial concerns.

Introduction

At the community level, childcare providers' observational accuracy may play an important role in identifying children with mental health problems and connecting them with relevant health services (Berkhout et al., 2012; Eklund et al., 2009). Given childcare providers potential role in early identification of young children in need of follow-up assessment for mental health services, surprisingly little research has been carried out on their accuracy in identifying children at risk for mental health problems. Since the childcare centers constitute a promising arena for early identification of young children (age one to six years) at risk for mental health problems, more attention should be directed towards childcare providers' perception of such problems (Poulou, 2015). There is a broad consensus that early childhood is a crucial time for identifying risk for later mental health problems. Thus, interventions should be initiated before negative developmental patterns emerge (Chen, 2010; Doyle et al., 2009; Dougherty et al., 2015; Essex et al., 2009; Heckman, 2006; Heo & Squires, 2011; Kauffman, 1999; Poulou, 2015; Raver et al., 2009; de Wolff et al., 2013). Because mental health problems in the form of early behavioral (externalizing) and emotional (internalizing) problems have been found to be a precursor for later maladjustment, it is important to identify young children with high, recurrent, and continued externalizing and internalizing problems (Basten et al., 2016; Briggs-Gowan et al., 2006; Essex et al., 2009; Fanti & Henrich, 2010; Gilliom & Shaw, 2004). However, it is difficult to identify problem behaviors in a period where development proceeds rapidly (Keenan et al., 1998).

For most children, the acquisition of behavioral and emotional self-regulation proceeds normally, but for some children mental health problems may emerge that are severe enough to cause concerns (Powell et al., 2006). Globally, data consistently show that 20% of children display mental health problems (Belfer, 2008) and a pooled prevalence estimates

show that 13% to 20% of children meet diagnostic criteria for a psychiatric disorder (Charach et al., 2020; Polanczyk et al., 2015; Vasileva et al., 2020). In addition, approximately 33% of 1 to 7 years old who meets diagnostic criteria for one psychiatric disorder also meets diagnostic criteria for at least one additional psychiatric disorder (Vasileva et al., 2020). Lavigne and colleagues (1996) found that boys in the preschool period were more likely to have a psychiatric disorder compared to girls, mainly in the form of externalizing disorders such as ODD, CD, and ADHD, while no gender effects were observed for emotional disorders (anxiety, depression etc.).

In Norway, the prevalence estimates are somewhat lower with 15% to 20% of the preschool children exhibiting some mental health problems (Skogen et al., 2014) and 7% having a symptom load that would qualify for a psychiatric disorder (Wichstrøm et al., 2012). Parents, childcare providers, and primary school teachers from Nordic countries tend to report lower symptom scores for emotional and behavioral problems on dimensional measures compared to other countries (Heiervang et al., 2008; Rescorla et al., 2012; Rescorla et al., 2014). This tendency has also been demonstrated in Norway (Drugli & Stensen, 2019; Larson & Drugli, 2011), especially for childcare providers' report of children's internalizing symptoms (Berg-Nielsen et al., 2012). Even though some emotional and behavioral problems may be transient in nature, it has been reported that preschoolers who meet diagnostic criteria for a psychiatric disorder at age 3 are five times more likely to still meet diagnostic criteria at age 6 (Bufferd et al., 2012). Additionally, approximately 50% of preschoolers who meets criteria for a psychiatric disorder still have a diagnose psychiatric disorder in middle childhood or early adolescence (Finsaas et al., 2018). As approximately half of all children with emotional or behavioral problems are not identified before school entry (Glascoe & Marks, 2011) and only one tenth of Norwegian four-years old with an emotional or behavioral disorder receives professional help (Wichstrøm et al., 2014), opportunities for early

intervention may be lost for these children who may have benefited from receiving help for their problems.

In contrast to clinical assessment, screening is only an indication of the presence or absence of a given condition or criteria. Early screening may lead to more children in need of services being identified and referred for a more thorough assessment. However, relying solely on the clinical judgement of health professionals may leave many children in need of help unidentified. For example, pediatricians and health nurses working without standardized screening instruments demonstrates a low accuracy in identifying children with developmental and/or mental health problems (Sheldrick et al., 2011; Skovgaard et al., 2008). There are several universal screening instruments available for the identification of young children at risk for mental health problems (see Bagner et al.,2012, or Lavigne et al., 2016a for a review), but the most cost effective approach is the nomination method. This method simply involves the respondent making a judgment call of nominating any child that he or she feels meets a given criterion (e.g., developmental concerns, at risk for mental health problems, one or more risk factors). Thus, the nomination method may be regarded as screening in the form of a subjective judgement call, something that childcare providers and teachers in the primary school should be used to from their work. The nomination method can also be seen as a pre-screener that can direct attention towards children for whom the respondent is uncertain about.

The accuracy of screening tests is usually measured by its sensitivity, which is the correct identification of those with the target condition, and specificity, which is the correct identification of those without the target condition. Misclassification may lead to wasted resources and possible stigmatization for those who are screened positive falsely (which results from low specificity), whereas a false negative (which results from a low sensitivity) result may deprive children of receiving appropriate help. One recommendation is that

screening tests should at least exhibit 70% sensitivity and between 70-80% specificity to be worthwhile (Glascoe, 2005). However, the accuracy of the nomination method has been examined less for young children than for school-aged children.

Research on school-aged children shows that teachers in primary school are more likely to nominate children with externalizing than internalizing symptoms (Loades & Mastroyannopoulou, 2010; Soles et al., 2008) and their accuracy is also better for externalizing than internalizing symptoms (Dwyer et al., 2006). The accuracy of primary school teachers' recognition of mental health problems is improved when presented with gender stereotypically cases, for instance boys with externalizing problems or girls with internalizing problems (Loades & Mastroyannopoulou, 2010). In addition, school-aged children identified by teachers as needing mental health services has exhibited significantly more adjustment problems than non-nominated peers (Layne et al., 2006; Roeser & Midgley, 1997). However, others report low to moderate sensitivity and specificity using the nomination method in identifying school age children with anxiety and depression problems (Dadds et al., 1997; Moor et al., 2000). Most of the research to date has used a cross-sectional design to investigate the accuracy, mainly how well the teacher nominations correspond with concurrent reports from other informants, such as parents and self-reports. In one of the few prospective studies, Dwyer and colleagues (2006) reported that teacher nominations was a poor screening instrument for detecting children with internalizing problems (sensitivity 34%, specificity 75%), but more successful for detecting those with externalizing problems (sensitivity 69%, specificity 78%) measured one year after baseline among 4-8 years old. Ollendick and colleagues (1990) found that children nominated by teachers as aggressive or withdrawn were outperformed by those nominated as well-adjusted on outcomes such as academic grades and social behavior through a five-year period.

Given the potential utility of childcare providers' nominations in identifying children at risk for mental health problems, an examination of the appropriateness of this approach for screening or pre-screening purposes is warranted. The aim of the current study is therefore to investigate the accuracy of childcare providers' nomination of children at risk for mental health problems against a well-established psychometric scale comparator (Caregiver-Teacher Report Form). Additionally, this study is warranted by the lack of research on childcare providers' ability to discriminate young children at risk for mental health problems from those who are not at risk, especially including the youngest children. As children's age and gender may have an impact on childcare providers' nomination accuracy, this will be explored as well. Coupled together with results from school-aged children, the following hypothesis will be tested: (1) childcare providers' nominations discriminate with high accuracy young children in the clinical range of the comparator from those who are not, (2) the accuracy is higher for externalizing problems compared to internalizing problems, and (3) childcare providers show better accuracy for boys than girls, as well as for older compared to younger children in the childcare centers.

Methods

Data are from the baseline, collected in 2012-2014, of the study Children in Central Norway, which aimed to enhance the competence of childcare providers addressing young children's mental health and improve relational quality between childcare providers and children. The study was approved by the Regional Committee for Medical and Health Research Ethics.

Procedure and participants

Parents with children in childcare centers, serving children from age one to six years old, in three municipalities in Central Norway received recruitment letters with information regarding the project together with an informed consent form. Information was also provided

in parent meetings before the project started. The recruitment letter provided the option for parents to consent either by logging in with a personal invitation code or by returning the consent form to the childcare center. Parental consent gave the childcare provider in the childcare center who was most familiar with the child permission to complete a survey regarding that parent's child. Childcare providers with a bachelor degree (three years of higher education) in early childhood education gave consent electronically via the survey with their own invitation codes. Participation was voluntary and parental consent could be withdrawn at any time without reprisal until the participation registry was deleted. Of the invited parents, 1631 (77%) consented to enroll their child in the study and the childcare providers reported on 1431 children (68%). The gender distribution was 51% boys and 49% girls with a mean age of 45 months. One-hundred and sixty-nine childcare providers participated (7% males) and they usually reported on 6 to 12 children each.

Measures

Childcare providers' nomination

The childcare providers were asked to make a global judgment concerning each child's risk status by answering "*yes*" or "*no*" to indicate whether they perceived a child with any developmental concerns. This question was located at the start of the survey before the standardized questionnaires were presented. If "*yes*" was answered, childcare providers could specify their nomination with checking one or more reasons for nomination (aggression, attention, emotional, social, motoric, language, home). However, only those nominated with specification of aggression, attention, emotional, or social were considered in the analyses to be nominated at risk to match the types of problems addressed in the comparator, the Caregiver-Teacher Report Form (C-TRF) (see below).

The C-TRF

Childcare providers completed the C-TRF (Achenbach & Rescorla, 2000), which contains 100 items describing problem behaviors for children ages 1.5 to 5 years. Each item has three response options: "*not true (as far as you know)*", "*somewhat or sometimes true*" and "*very often or often true*" corresponding with a score from zero to two. The C-TRF contains the following subscales: emotional reactive (7 items), anxious/depressed (8 items), withdrawn (10 items), somatic complaints (7 items), attention problems (9 items), aggressive behavior (25 items), and other problems (34 items). A total problem score (ranging from zero to 200) can be calculated by adding the scores across all items. In addition, two broadband scales can be calculated by adding certain subscales, namely internalizing problems (emotional reactive, anxious/depressed, withdrawn, and somatic complaints) and externalizing problems (attention and aggression problems). The subscales and the broader scales of internalizing problems, externalizing problems, and total problems can then be used to create an individual problem profile to investigate if the scores surpass the selected cutoff point(s), which indicates that further referral may be needed.

The C-TRF has exhibited a test-retest (mean interval of 8 days) Pearson's correlation coefficient of .88 for the Total Problem scale and .81 mean across all scales. The cross-informant correlation on the Total Problem scale is reported to be .72 for pairs of childcare providers. The developers have provided thorough psychometric information regarding the instrument in the manual and elsewhere (Achenbach & Rescorla, 2000; Rescorla, 2005). The validity, reliability, and factor structure of the C-TRF have also proven to be excellent across cultures (de Groot et al., 1994; Koot et al., 1997; Ivanova et al., 2007; Ivanova et al., 2010; Ivanova et al., 2011; Liu et al., 2011; Rescorla et al., 2012; Rescorla et al., 2014; Verhulst & Koot, 1992). The 90th percentile defines the clinical range of the C-TRF total problem score and has shown to discriminate well between referred and non-referred children (Achenbach &

Rescorla, 2000; Rescorla, 2005). To apply empirically derived cutoff values are a viable and commonly used option when diagnostic information is not available (e.g. information from structured diagnostic interviews).

The C-TRF was selected as the comparator in the present study because of its extensive use in research and clinical settings, as well as its well-documented psychometric properties. In addition, the C-TRF and its parent-reported counterpart, the Child Behavior Checklist (CBCL), are commonly used in validation and accuracy studies as comparator for screening instruments (Lavigne et al, 2016a). In Norway, the C-TRF is mainly used by special health services as a first-assessment instrument, often administered as a part of children's clinic admission. For the C-TRF broadband scales, the Cronbach's alphas in the present sample are α= .85 for Internalizing problems, α= .94 for Externalizing problems, and α= .94 for Total problems

We defined, as instructed in the manual (Achenbach & Rescorla, 2000) children with a score at or above the 90th percentile on the C-TRF's Total Problem, Internalizing, or Externalizing scale to be at elevated risk for mental health problems in respective domains. In addition, children in the top 2% on at least one subscale (except somatic complaints) but who were not rated in the clinical range (90th percentile) on any of the three C-TRF scales were also considered to be at elevated risk. For the Total Problem scale, the top 2% on any subscale (excluding somatic complaints) was included in the clinical range, while for the Internalizing and Externalizing scale only the top 2% on the corresponding subscales was included. Following recommendations by Achenbach and Rescorla (2000), this was done because the subscales compromise a smaller and more homogeneous sets of problem, which are believed in need of a more stringent cutoff value to suggest that professional help is needed. By doing so, we ensured that children scoring very high on a specific set of problems were included in the clinical range on the Total Problem, Internalizing, and Externalizing scales, even though

they might have scored below these cutoff values on the broader scales. Because childcare providers tend to score boys higher than girls on the C-TRF (Achenbach & Rescorla, 2000; Drugli & Stensen, 2019; Kristensen et al., 2010; Rescorla, 2005), this procedure was based on the present samples norms separately for girls and boys to establish gender specific cutoffs, so that girls mental health problems not were overlooked when defining the clinical range. The cutoff values used are shown in Table 1. As age effects for the C-TRF are generally very small (Drugli & Stensen, 2019; Rescorla, 2005), the gender specific cutoff values were applied to the entire age span in the present sample.

Statistical analyses

We calculated the sensitivity and specificity, as well as the rate of false positive and false negative cases using the following formula:

Sensitivity (or true positive rate) = true positive/(true positive+false negative)

Specificity (or true negative rate) = true negative/(false positive+true negative)

False positive rate = false positive/(true positive+false positive)

False negative rate = false negative/(false negative+true negative)

In addition, the positive predictive value (PPV) and negative predictive value (NPV) at sample prevalence of target condition was calculated using the following formula:

$$PPV = \frac{Sensitivity * Prevalence}{Sensitivity * Prevalence + (1 - Specificity)(1 - Prevalence)}$$

$$NPV = \frac{Specificity * (1 - Prevalence)}{(1 - Sensitivity) * Prevalence + Specificity * (1 - Prevalence)}$$

This was done separately for each age group (ages 1-2 and 3-6), overall and separately for each gender. Childcare centers in Norway are usually organized by children's age. Thus, the

sample were divided into two age groups to reflect this. Independent sample t-tests were performed to investigate age and gender differences of childcare providers' nominations. One child was excluded from the study due to missing age information, while none of the rest had missing data. The analyses were performed in SPSS25. Wilson 95% confidence intervals were computed where relevant, using STATA15.

Insert Table 1 about here

Results

Significantly more boys compared to girls were nominated by childcare providers ($p=.018$), as well as significantly more children from the 3-6 years old age group compared to the 1-2 years old group ($p=<.001$). As seen in Table 2, in the 1-2 years old group 13% boys were nominated and 9% girls, while for the 3-6 years old group the nomination rates were 23% for boys and 17% for girls. Table 3 shows the proportion of children found in the non-clinical and the clinical range of the C-TRF, where the proportion found in the clinical range are from 8% to 12% depending on age and gender.

Insert Table 2&3 about here

Sensitivity and specificity analyses

As shown in Table 4 (nominations against the C-TRF's Total Problem scale), although the overall sensitivity of childcare providers' nominations against the C-TRF's Total Problem scale was 57% for the 1-2 years old and 81% for the 3-6 years old children, only the 1-2 years old girls were nominated with a low 44% sensitivity. Boys 1-2 years old yielded a sensitivity of 71% and both genders of the 3-6 years old children exceeded 78% sensitivity. The false positive rate ranged from 41% to 57%, with the highest rate found for 3-6 years old boys and the lowest for 1-2 years old boys. The specificity ranged from 86% to 95% with false negative

rates below 7%. The positive predictive value (PPV) ranged from 43% to 59%, while the

negative predictive value (NPV) ranged from 93% to 98%

Insert Table 4 about here

As shown in Table 5 (nominations against the C-TRF's Internalizing and

Externalizing scales), childcare providers nominations compared to the scores in the clinical

range of the C-TRF's Internalizing scale showed a sensitivity ranging from 53% (girls 1-2

years old) to 83% (boys 3-6 years old). The lowest rate of false positives was found for boys

1-2 years old (52%) and the highest for boys 3-6 years old (70%). The specificity ranged from

83% (boys 3-6 years old) to 95% (girls 1-2 years old) and rates of false negatives ranged from

2% to 4% across age and gender. The PPV ranged from 30-48% and the NPV 96% to 98%.

Childcare providers nominations compared to scores in the clinical range on the C-

TRF's Externalizing scale similarly yielded the highest sensitivity for boys 3-6 years old

(83%) and the lowest for girls 1-2 years old (24%). The false positive rates ranged from 52%

(boys 1-2 years old) to 75% (girls 1-2 years old) and the specificity ranged from 84% (boys 3-

6 years old) to 93% (boys 1-2 years old). Rates of false negatives ranged from 2% to 8% with

the highest rate found for girls 1-2 years old and the PPV ranged from 25% to 48%, while the

NPV ranged from 92% to 98%.

Insert Table 5 about here

Discussion

The aim of the current study was to investigate how accurately a childcare providers'

nomination could discriminate children below or within the clinical range of a well-validated

psychometric scale, the C-TRF. Overall, childcare providers' nominations of children at risk

were relatively well reflected in their scores of the C-TRF, but there were variations related to

child gender, age range, and type of behavior problems. Childcare providers nominated

significantly more boys and older preschool children compared to girls and younger preschool children, and they were more accurate in discriminating normal from abnormal behavior for older children compared to younger children, and boys compared to girls. For older children, childcare providers show about the same accuracy in discriminating children with internalizing and externalizing problems, while for the younger children a lower accuracy is demonstrated, particularly for girls with externalizing problems. Childcare providers' nominations also created a considerable portion of false positives, especially for the oldest age group.

How accurate are childcare providers in discriminating?

Childcare providers are trained in making subjective decisions in their professional life and the nomination method can be seen as a subjective screening instrument. In childcare centers, childcare providers meet with multiple children over a prolonged time and can potentially build up a reference base for normal and abnormal behavior. According to Glascoe (2005), a screening instrument should exhibit at least 70% sensitivity and a specificity above 70%, preferably above 80%. The specificity as seen in tables 4 and 5 is well above this recommendation, indicating that children below the clinical range are generally not nominated by childcare providers. The overall sensitivity on the other hand, is more mixed, indicating that childcare providers generally have more problems nominating children found in the clinical range of the C-TRF. In other words, childcare providers do a better job at identifying children in the non-clinical range than identifying those in the clinical range. Taken together, these results may suggest that childcare providers have a better reference base for normal than abnormal behavior.

As most children are found within normal developmental parameters, universal approaches to promote competency and healthy development may be sufficient. However, for some children more selective or targeted interventions are necessary to ensure healthy

development. This assumes that the children who will benefit of such interventions are identified. Even though childcare providers' nominations identify most of the children in the clinical range of the C-TRF, approximately half of the nominees are children with normal parameters. Thus, there is a considerable improvement potential for childcare providers to increase their accuracy in distinguishing normal from abnormal behavior. Misclassifications in the form of false positives may stigmatize and encumber children within normal parameters with unnecessary screening and assessment, while false negatives may deny children with a clinical level of mental health problems the help they need. Childcare providers' nominations should always be followed by a psychometrically sound screening instrument, while the non-nominated children are found in the normal range of the C-TRF most of the time. This said, even if the rate of false negatives is small, behind every false negative case there is a child with a clinical level of mental health problems that childcare providers have no concerns for. Thus, efforts to reduce the false negative rates to an absolute minimum should be pursued.

Age and gender differences

Overall, the sensitivity reached the level of recommendation for 3-6 years old, but not for 1-2 years old. The exception are boys in the 1-2 years old group measured against the Total Problems scale, which barely is above the recommended level. Actually, the sensitivity obtained by the nomination method for 3-6 years old is about the same level as a recent cultural validation of the screening instrument Ages and Stages Questionnaire: Social-Emotional (ASQ:SE) (Stensen et al., 2018). However, the ASQ:SE exhibits considerably fewer false positive cases compared to the childcare providers nominations in the present study (Squires et al., 2002; Stensen et al., 2018). For the 1-2 years old group, it seems that childcare providers find it difficult nominating children in the clinical range of the C-TRF, especially girls. A possible explanation is that symptom expression is different for girls and maybe childcare providers perceive them differently compared with boys.

Regarding the age difference in sensitivity, it might be that childcare providers lack the adequate knowledge or observational skills to classify the youngest children in childcare centers with an elevated symptom load or that precursors for mental health problems are subtler compared to older children. Also, childcare providers would on average have spent less time with the youngest children, who in many cases would just recently have started in the childcare center. This said, precursors for emotional and behavioral problems are identifiable in the first two years of life. For example, it was reported by Keenan and colleagues (1998) that a difficult temperament when children were 18 months old significantly correlated with both genders internalizing problems at age 3 and 5 years old. Additionally, early non-compliance in girls and aggression in boys were related in later externalizing problems. One possible explanation may be that childcare providers are more reluctant to nominate younger children, maybe perceiving a longer timeframe for development to normalize before entering school. On the other hand, childcare providers seem overly eager to nominate the older children, resulting in a considerable portion of false positives. If childcare providers nominate a significant portion of children, the sensitivity can consequently be artificially high due to inflation, as many of the cases would be false positives. For example, if the childcare providers had nominated all the children in the current study, the sensitivity would be 100% (all children in the clinical range nominated), but about 90% of those defined in the non-clinical range would be false positives.

Internalizing and externalizing problems

Contrary to what has been established in prior research with school-aged children (Dwyer et al., 2006), for the 3-6 years old group, childcare providers' exhibit about the same level of sensitivity for internalizing problems and externalizing problems. For this age group, childcare providers' nominations are above Glascoe's (2005) recommendation for both types of behavior problems independent of gender. For 1-2 years old on the other hand, the

sensitivity is below the recommendation for both types. As mentioned above, this could be due to less time spent with the child, lack of knowledge or observational skills, or subtler symptom expression for younger children that childcare providers find harder to catch. The largest gender discrepancy in sensitivity is also found in the 1-2 years old group on the externalizing scale. Even though boys usually are rated higher on the C-TRF by childcare providers (Achenbach & Rescorla, 2000; Drugli & Stensen, 2019; Kristensen et al., 2010). Basten et al. (2016) found no gender differences regarding internalizing and externalizing problem profiles for children 1.5 years old reported by mothers. This said, others have found that parents report significantly more externalizing problems in boys compared to girls among older preschoolers (Chen, 2010) and that childcare providers rate boys substantially higher on aggression across cultures compared to girls (LaFreniere et al., 2002). This may suggest that childcare providers perception could be influenced by both age and gender expectations (i.e. that girls exhibit fewer externalizing problems than boys and problems may be perceived more normative for younger than older children), which again may influence whom they nominate and how they rate children. Moreover, Norwegian childcare providers and primary school teachers seem to have a more normative perception of children's internalizing problems compared to other countries, making them more reluctant to state such behaviors as problematic (Berg-Nielsen et al., 2012, Heiervang et al., 2007). Consequently, as externalizing problems are more prevalent for boys, more boys would be classified as false positives since childcare providers are more eager to state such behaviors as problematic. In addition and in accordance with previous research with school-aged children (Loades & Mastroyannopoulou, 2010), childcare providers are most accurate when presented with gender stereotypical problems (e.g., boys with externalizing problems and girls with internalizing). In other words, it seems likely that childcare providers operate with different threshold for stating concern depending on children's age and gender.

Considerations

When considering if the sensitivity is acceptable, one must also consider the rate of false positives. In the current study, it might be that the higher nomination rate for older children gives an overly optimistic sensitivity when it also shows that about every other nomination is a false positive. How high a rate of false positives is acceptable depends largely on the context and aim. A high rate of false positives may result in unnecessary referrals and an overload of the support system, as well as creating unnecessary stress for children and parents. In addition, it can also create the Pygmalion effect (Rosenthal & Jacobson, 1968), where a childcare providers' negative perception or attitude towards a child may lead to self-fulfilling prophecies in form of increased risk for negative relations and consequently more behavior problems. Childcare providers reported conflict level with children have been shown to influence how they rate children's problem behavior, as well as increasing the discrepancy in rating agreement between childcare providers and parents. Three out of four times, childcare providers and parents disagree regarding children with high severity of problems (Berg-Nilsen et al., 2012), underlining the importance of including both childcare providers and parents in the screening process.

Another issue in need of consideration is the necessity of screening all children in childcare centers. As seen in the current study, the specificity and NPV is generally high and the rate of false negatives low, indicating that when childcare providers are not concerned, children are generally confirmed with score below the clinical range. For purposes of early intervention, it is more important to obtain a low rate of false negatives than false positives. Moreover, it has been reported that children screened as false positives carries more psychosocial risk than children screened as true negatives (Glascoe, 2001; Jensen & Watanbe, 1999). Thus, some children may carry more risk than others even though they do not display a clinical level of symptoms. If childcare providers do not nominate and state their concern, the

opportunity for early identification and intervention may diminish for those children who are actually at risk.

Clinical implications

In the current study, the positive predictive value was about 40% to 50% with a 10% prevalence of clinically elevated mental health problems. If childcare providers' nomination had been the only form of screening before referral, the mental health services would have wasted half of their time evaluating children with a non-clinical level of mental health problems. However, if childcare providers nominations are regarded rather as a first step pre-screener in a sequential screening process, their concern should be followed by the completion of a psychometrically sound screening instrument to confirm or disconfirm their concerns (cf. Lavigne et al. (2016a) for a review on classification accuracy for various screening instruments). If using this instrument confirms a positive pre-screener, contact with mental health services should be initiated for a more thorough evaluation. Using screening instruments as a dialog tool with parents may also be beneficial as emotional and behavioral problems may be context specific.

Sequential screening may help in managing the high rates of false positives commonly found in populations with a low prevalence of mental health problems, reflected by a low positive predictive value (PPV) (Lavigne et al., 2016b). When the prevalence of problems decline, the PPV declines, as does the ability of a screening tests correctly to detect true cases (Lavigne et al., 2016b). The result of sequential screening is a higher prevalence for the latter stages, thus reducing the measurement errors associated with low prevalent problems and the rate of false positives (Young & Takala, 2018). In addition, the high specificity and low rate of false negatives obtained in the current study suggest that childcare providers' nominations may be suitable to direct attention to the uncertain cases and at the same time accurately rule out those children without the targeted condition, thus making the standardized screening

instruments more efficient in the latter stages of the screening sequence due to higher prevalence. This said, there are still some children with a clinical problem level that goes unidentified, underlining the importance of collaboration between childcare providers, parents, and mental health professionals to ensure the best possible identification rate.

Strengths and limitations

Previous studies have mainly focused on school-aged children when investigating the accuracy of the nomination method. This study adds to the knowledge of how this method works with childcare providers nominating young children in childcare centers against a well-established comparator. Another feature of this study is the inclusion of the full age span of children enrolled in childcare centers. However, several possible limitations need to be mentioned.

First, the choice of cutoff values on the comparator, the C-TRF, does not necessarily need to be the optimal cutoff for the current sample. However, if the cutoff value in the present study had been lowered further, the sensitivity would have increased and consequently the rate of false positives dropped. Additionally, this would also have led to a drop in specificity and an increase in the rate of false negatives. As mentioned before, which rates are acceptable depends largely on the aim and the ability of the support systems to act. Also, the rate of false positives might be inflated for the internalizing and externalizing testing because the childcare providers' nominations integrate both emotional and behavioral concerns. Consequently, one might expect a higher rate of false positives when testing an overall judgment against more specific sets of problems. In addition, applying the C-TRF's Total Problem scale gives an indication of symptom load, but gives little direction to which problem domains, represented by the specific subscales, attention should be directed. Future studies with larger sample sizes should investigate the accuracy of more specific concerns against specific sets of problems, as this could potentially provide further insight in childcare

providers ability to identify children at risk for mental health problems and their skills in symptom recognition.

Second, inflation of accuracy estimates may also occur when using the same respondent for both the nomination and the comparator without any significant separation in time between completions of these tasks. However, as childcare providers' perception of problem behaviors in young children may be important for referral, it seems appropriate to use them as sole respondents in the current study. There might also be a priming bias at play, because the decision to nominate or not might influence how the childcare providers respond on the C-TRF. Because of these limitations, these results may present an upper bound of sensitivity and specificity.

Third, a possible limitation could be the comparator itself, as it does not exhibit 100% perfect discrimination (Lavigne et al., 2016a). However, the status and extensive use of the C-TRF among clinicians and researchers, as well as its psychometric properties, makes it a commonly used "gold standard" when investigating other instruments. The C-TRF itself is usually tested against structured diagnostic interviews, which as of date are considered the "gold standard" for other "gold standards". Thus, future studies should investigate the accuracy of childcare providers' nominations against diagnostic information. This said, diagnostic interviews may not always be convenient or applicable in larger studies, as it demands more resources compared to questionnaire comparators.

Last, findings from this study may not be automatically generalized to other countries, as prevalence of problems, organization of childcare centers, and childcare providers' education may differ from Norway. However, findings from the current study underlines the potential in listening to childcare providers concerns to direct attention towards children in need of screening, and maybe referral for further assessment.

Conclusion

The nomination method appears promising as a first step pre-screening in a screening sequence, as the childcare providers nominate a large portion of young children who are at risk for mental health problems, with exception of the youngest girls. In other words, if childcare providers have a hunch about a child's risk status, it seems wise to investigate it further. However, the childcare providers' nominations also have a considerable potential for improvement, as it creates a considerable rate of false positives that must be dealt with, for example by applying psychometrically sound screening instruments as part of a sequential screening process. On the other hand, when childcare providers do not nominate a child as at risk, that child is usually not found in the clinical range on the C-TRF. This raises the question of the necessity of screening all children in childcare centers, but rather use childcare providers' nominations to direct attention towards the uncertain cases and then apply a standardized screening instrument. Given the rapid developmental processes that occurs during the preschool period, childcare providers need adequate knowledge of age-appropriate normal and abnormal development, observational training, as well as access to appropriate screening instruments to be able to accurate classify children at risk for mental health problems, especially the youngest children.

References

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA Preschool Forms & Profiles.* Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Bagner, D. M., Rodríguez, G. M., Blake, C. A., Linares, D., & Carter, A. S. (2012). Assessment of Behavioral and Emotional Problems in Infancy: A Systematic Review. *Clinical Child and Family Psychology.* 15: 113-128

Basten, M., Tiemer, H., Althoff, R. R., van de Schoot, R., Jaddoe, V. W. V., Hofman, A., …, & van der Ende, J. (2016). The Stability of Problem Behavior Across the Preschool years: An Empirical Approach in the General Population. *Journal of Abnormal Child Psychology.* 44: 393-404

Belfer, M.L. (2008). Child and adolescent mental disorders: The magnitude of the problem across the globe. *Journal of Child Psychology and Psychiatry.* 49: 226-236

Berg-Nielsen, T.S., Solheim, E., Belsky, J., & Wichstrøm, L. (2012). Preschoolers' Psychosocial Problems: In the Eyes of the Beholder? Adding Teacher Characteristics as Determinants of Discrepant Parent-Teacher Reports. *Child Psychiatry and Human Development.* 43: 393-413

Berkhout, L., Hoekman, J., & Goorhuis-Brouwer, S. M. (2012). Psychosocial health of 4-year-old boys and girls, as observed by parents and teachers. *European Journal of Developmental Psychology.* 9: 500-507

Briggs-Gowan, M. J., Carter, A. S., Bosson-Heenan, J., Guyer, A. E., & Horwitz S. M. (2006). Are Infant-Toddler Social-Emotional and Behavioral Problems Transient? *American Academy of Child and Adolescent Psychiatry.* 45: 849-858

Bufferd, S. J., Doughert, L. R., Carlson, G. A., Rose, S., & Klein, D. N. (2012). Psychiatric Disorders in Preschoolers: Continuity From Ages 3 to 6. *American Journal of Psychiatry*. 169: 1157-1164

Charach, A., Mohammadzadeh, F., Belanger, S. A., Easson, A., Lipman, E. L., McLennan, J. D., …, & Szatmari, P. (2020). Identification of Preschool Children with Mental Health Problems in Primary Care: Systematic Review and Meta-analysis. *Journal of the Canadian Acedemy of Child and Adolescent Psychiatry*. 29: 76-105

Chen, J. J. (2010). Gender differences in externalizing problems among preschool children: implications for early childhood educators. *Early Child Development and Care*. 180: 463-474

Dadds, M. R., Spence, S. H., Holland, D. E., Barrett, P. M., & Laurens, K. R. (1997). Prevention and early intervention for anxiety disorders: A controlled trial. *Journal of Consulting and Clinical Psychology*. 65: 627-635

Dougherty, L. R., Leppert, K. A., Merwin, S. M., Smith, V. C., Bufferd, S. J.,& Kushner, M. R. (2015). Advances and Directions in Preschool Mental Health Research. *Child Development Perspectives*. 9: 14-19

Doyle, O., Harmon, C. P., Heckman, J. J., & Tremblay, R. E. (2009). Investing in early human development: Timing and economic efficiency. *Economics and Human Biology*. 7: 1-6

Drugli, M. B. & Stensen, K. (2019). Psykiske vansker hos norske barnehagebarn. *Første steg*. 46-49

Dwyer, S. B., Nicholson, J. M., & Battistutta, D. (2006). Parent and teacher identification of children at risk of developing internalizing or externalizing mental health problems: A comparison of screening methods. *Prevention Science*. 7: 343-357

Eklund, K., Renshaw, T. L., Dowdy, E., Jimerson, S. R., Hart, S. R., Jones, C. N., & Earhart, J. (2009). Early Identification of Behavioral and Emotional Problems in Youth: Universal Screening versus Teacher-Referral Identification. *The California School Psychologist*. 14: 89-95

Essex, M. J., Kraemer, H. C., Slattery, M. J., Burk, L. R., Boyce, W. T., Woodward, H. R., & Kupfer, D. J. (2009). Screening for childhood mental health problems: outcomes and early identification. *Journal of Child Psychology and Psychiatry.* 50: 562-570

Fanti, K. A. & Henrich C. C. (2010). Trajectories of Pure and Co-Occuring Internalizing and Externalizing Problems From Age 2 to Age 12: Findings From the National Institute of Child Health and Human Development Study of Early Child Care. *Developmental Psychology*. 46: 1159-1175

Finsaas, M. C., Bufferd, S. J., Dougherty, L. R., Carlson, G. A., & Klein, D. N. (2018a). Preschool Psychiatric Disorders: Homotypic and Heterotypic Continuity through Middle Childhood and Early Adolescence. *Psychological Medicine*. 48: 2159-2168

Gilliom, M. & Shaw, D. S. (2004). Codevelopment of externalizing and internalizing problems in early childhood. *Development and Psychopathology*. 16: 313-333

Glascoe, F. P. (2001). Are Overreferrals on Developmental Screening Tests Really a Problem? *Archives of Pediatric and Adolescent Medicine*. 155: 54-59

Glascoe, F. P. (2005). Screening for Developmental and Behavioral Problems. *Mental Retardation and Developmental Disabilities*. 11: 173-179

Glascoe, F. P. & Marks, K. P. (2011). Detecting children with developmental-behavioral problems: The value of collaborating with parents. *Psychological Test and Assessment Modeling*. 53: 258-279

de Groot, A., Koot, H. M., & Verhulst, F. C. (1994). Cross-Cultural Generalizability of the Child Behavior Checklist Cross-Informant Syndromes. *Psychological Assessment*. 6: 225-230

Heckman, J. J. (2006). Skill Formation and the Economics of Investing in Disadvantaged Children. *Science*. 312: 1900-1902

Heiervang, E., Goodman, A., & Goodman, R. (2008). The Nordic advantage in child mental health: separating health differences from reporting style in a cross-cultural comparison of psychopathology. *Journal of Child Psychology and Psychiatry*. 49: 678-685

Heo, K. H., & Squires, J. (2011). Cultural adaption of a parent completed social emotional screening instrument for young children: Ages and stages questionnaire-social emotional. *Early Human Development*. 88: 151-158

Ivanova, M.Y., Achenbach, T.M., Rescorla, L. A., Dumenci, L., Almqvist, F., Bathiche, M., …, & Verhulst, F. (2007). Testing the Teacher's Report Form syndromes in 20 societies. *School Psychology Review*. 36: 468-483

Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Harder, V. S, Ang, R. P., Bilenberg, N, … & Verhulst, F. C. (2010). Preschool Psychopathology Reported by Parents in 23 Societies: Testing the Seven-Syndrome Model of the Child Behavior Checklist for Ages 1,5-5. *New Research*. 49: 1215-1224

Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Bilenberg, N., Bjarnadottir, G., Denner, S., …, & Verhulst, F. C. (2011). Syndromes of Preschool Psychopathology Reported by Teachers and Caregivers in 14 Societies Using the Caregiver-Teacher Report Form (C-TRF). *Journal of Early Childhood and Infant Psychology*. 7: 87-103

Jensen, P. S. & Watanabe, H. (1999). Sherlock Holmes and Child Psychopathology Assessment Approaches: The Case of the False-Positive. *Journal of American Academy of Child and Adolescent Psychiatry*. 38: 138-146

Kauffman, J. M. (1999). How we prevent the prevention of emotional and behavioral disorders. *Exceptional Children*. 65: 448–468

Keenan, K., Shaw, D., Delliquadri, E., Giovannelli, J., & Walsh, B. (1998). Evidence for the Continuity of Early Problem Behaviors: Application of a Developmental Model. *Journal of Abnormal Child Psychology*. 26: 441-454

Koot, H., van den Oord, J., Verhulst, F. C., & Boomsma, D. (1997). Behavioral and emotional problems in young preschoolers: Cross-cultural testing of the validity of the Child Behavior Checklist/2-3. *Journal of Abnormal Child Psychology.* 25: 183-196

Kristensen, S., Henriksen T. B., & Bilenberg, N. (2010). The Child Behavior Checklist for Ages 1.5-5 (CBCL/1½–5): Assessment and analysis of parent and caregiver-reported problems in a population-based sample of Danish preschool children. *Nordic Journal of Psychiatry*. 64: 203-209

LaFreniere, P., Masataka, N., Butovskaya, M., Chen, Q., Dessen, M. A., Atwanger, K., …, & Frigerio, A. (2002). Cross-Cultural Analysis of Social Competence and Behavior Problems in Preschoolers. *Early Education & Development*. 13: 201-220

Lavigne, J. V., Gibbons, R. D., Christoffel, K. K., Arend, R., Rosenbaum, D., Binns, H., …, & Isaacs, C. (1996). Prevalence rates and correlates of psychiatric disorders among preschool children. *Journal of the American Academy of Child and Adolescent Psychiatry*. 35: 204-214

Larsson, B. & Drugli, M. B. (2011). School competence and emotional/behavioral problems among Norwegian school children as rated by teachers on the Teacher Report Form. *Scandinavian Journal of Psychology*. 52: 553-559

Lavigne, J. V., Meyers, K. M., & Feldman, M. (2016a). Systematic Review: Classification Accuracy of Behavioral Screening Measures for Use in Integrated Primary Care Settings. *Journal of Pedriatic Psychology*. 5: 1-19

Lavigne, J. V., Feldman, M., & Meyers, K. M. (2016b). Screening for Mental Health Problems: Addressing the Base Rate Fallacy for a Sustainable Screening Program in Primary Care. *Journal of Pediatric Psychology*. 41: 1081-1090

Layne, A. E., Bernstein, G. A., & March J. S. (2006). Teacher Awareness of Anxiety Symptoms in Children. *Child Psychiatry and Human Development*. 36: 383-392

Liu, J., Cheng, H., & Leung, P. W. L. (2011). The Application of the Preschool Child Behavior Checklist and the Caregiver-Teacher Report Form to Mainland Chinese Children: Syndrome Structure, Gender Differences, Country Effect, and Inter-Informant Agreement. *Journal of Abnormal Child Psychology*. 39: 251-264

Loades, M. E. & Mastroyannopoulou, K. (2010). Teachers' Recognition of Children's Mental Health Problems. *Child and Adolescent Mental Health*. 15: 150-156

Moor, S., Sharrock, G., Scorr, J., McQueen, H., Wrate, R., Cowan, J., & Blair, C. (2000). Evaluation of a teaching package designed to improve teachers' recognition of depressed pupils - a pilot study. *Journal of Adolescence*. 23: 331-342

Ollendick, T. H., Greene, R. W., Weist, M. D., & Oswald, D. P. (1990). The Predictive Validity of Teacher Nominations; A Five-Year Followup of At-Risk Youth. *Journal of Abnormal Child Psychology*. 18: 699-713

Polanczyk, G. V, Salum, G. A., Sugaya, L. S., Caye, A., & Rohde, L. A. (2015). Annual Research Review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of Child Psychology and Psychiatry*. doi:10.1111/jcpp.12381

Poulou, M. S. (2015). Emotional and Behavioural Difficulties in Preschool. *Journal of Child and Family Studies*. 24: 225-236

Powell, D., Dunlap, G., and Fox, L. (2006). Prevention and Intervention for the Challenging Behaviors of Toddlers and Preschoolers. *Infants & Young Children*. 19: 25-35

Raver, C. C., Jones, S. M., Li-Grining, C., Zhai, F., Metzger, M. W., & Solomon, B. (2009). Targeting Children's Behavior Problems in Preschool Classrooms: A Cluster-Randomized Controlled Trial. *Journal of Consulting and Clinical Psychology*. 77: 302-316

Rescorla, L. A. (2005). Assessment of Young Children Using the Achenbach System of Empirically Based Assessment (ASEBA). *Mental Retardation and Developmental Disabilities Research Review*. 11: 226-237

Rescorla, L. A., Achenbach, T. M., Ivanova, M. Y., Bilenberg, N., Bjarnadottir, G., Denner, S., … & Verhulst, F. C. (2012). Behavioral/Emotional Problems of Preschoolers: Caregiver/Teacher Reports From 15 Societies. *Journal of Emotional and Behavioral Disorders*. 20: 68-81

Rescorla, L. A., Bochicchio, L., Achenbach, T. M., Ivanova, M. Y, Almqvist, F., Begovac, I., … & Verhulst, F. C. (2014). Parent-Teacher Agreement on Children's Problems in 21 Societies. *Journal of Clinical Child and Adolescent Psychology*. 43: 627-642

Roeser, R. W., & Midgley, C. (1997). Teachers' views of issues involving students' mental health. *The Elementary School Journal*, 2, 115-133

Rosenthal, R. & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*. 3: 16-20

Sheldrick, R. C., Merchant, S., & Perrin, E. C. (2011). Identification of Developmental-Behavioral Problems in Primary Care: A Systematic Review. *Pediatrics*. 128: 356-363

Skogen, J. C., Kjeldsen, A., Knudsen, A. K., Myklestad, I., Nesvåg, R., Reneflot, A., & Major. E. (2014). *Psykisk helse hos barn og unge* (Folkehelserapporten 2014:4). Oslo: Nasjonalt Folkehelseinstitutt

Skovgaard, A. M., Houmann, T., Christiansen, E., Olsen, E. M., Landorph, S. L., Lichtenberg, A., & Jorgensen, T. (2008). Can a general health surveillance between birth and 10 months identify children with mental disorder at 1 ½ year? A case-control study nested in cohort CCC 2000. *European Child and Adolescent Psychiatry*. 17: 290-298

Soles, T., Bloom, E. L., Heath, N. L., & Karagiannakis, A. (2008). An exploration of teachers' current perceptions of children with emotional and behavioural difficulties. *Emotional and Behavioural Difficulties*. 13: 275-290

Squires, J., Bricker, D., & Twombly, E. (2002). *The ASQ:SE User's Guide*. Paul H. Brookes Publishing Co.

Stensen, K., Stenseng, F., Lydersen, S., de Wolff, M. S., Wallander, J. L., & Drugli, M. B. (2018). Screening for mental health problems in a Norwegian preschool population. A validation of the Ages and Stages Questionnaire: Social-emotional (ASQ:SE). *Child and Adolescent Mental Health*. 23: 368-375

Vasileva, M., Graf, R. K., Reinelt, T., Petermann, T., & Petermann, F. (2020). Research review: A meta-analysis of the international prevalence and comorbidity of mental

disorders in children between 1 and 7 years. *Journal of Child Psychology and Psychiatry*.

DOI: 10.1111/jcpp.13261

Verhulst, F. C., & Koot, H. M. (1992). *Child psychiatric epidemilogogy: concepts, methods and findings*. Beverly Hills: Sage Publications

Wichstrøm, L., Berg-Nielsen, T.S., Angold, A., Egger, H.L., Solheim, E. & Sveen, T.H. (2012). Prevalence of psychiatric disorders in preschoolers, *Journal of Child Psychology and Psychiatry*. 53: 695–705

Wichstrøm, L., Belsky, J., Jozefiak, T., Sourander, A. & Berg-Nielsen T. S. (2014). Predicting Service Use for Mental Health Problems Among Young Children. *Pediatrics*. 133: 1054-1060

de Wolff, M. S., Theunissen, M. H., Vogels, A. G. C., & Reijneveld, S. A. R. (2013). Three Questionnaires to Detect Psychosocial Problems in Toddlers: A Comparison of the BITSEA, ASQ:SE, and KIPPPI. *Academic Pediatrics*. 13: 587-592

Young, N. D. & Takala, C. R. (2018). Sequential Screening to Improve Behavioral Health Needs in Primary Care. *Journal of American Child and Adolescent Psychiatry*. 57: 603-609