

Concurrent validity of the Strengths and Difficulties Questionnaire in an indigenous pre-school population

Paula Kersten • Alain C. Vandal • Hinemoa Elder • Robyn Tauroa • Kathryn M. McPherson

P. Kersten (✉)

School of Health Sciences, Westlaine House, Falmer Campus, University of Brighton, Village Way, Brighton

BN1 9PH, UK & Faculty of Health & Environmental Sciences, AUT University, Auckland, New Zealand

Email p.kersten@brighton.ac.uk. Tel +44 273 643483

Alain C. Vandal

Department of Biostatistics and Epidemiology, AUT University, New Zealand & Ko Awatea Health Intelligence

and Informatics, Counties Manukau District Health Board, Auckland, New Zealand

Hinemoa Elder

Te Whare Mātai Aronui, Te Whare Wānanga o Awanuiārangi, Auckland, New Zealand

Robyn Tauroa • Kathryn M. McPherson

Centre for Person Centred Research, AUT University, Auckland, New Zealand

Kathryn M. McPherson

The Health Research Council of New Zealand, Auckland, New Zealand

Acknowledgements

We thank the tamariki, families and whānau in this study for their participation.

Disclosure of Potential Conflicts of Interest

Funding: This study was supported by the Ministry of Health of New Zealand (grant number 341088). The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the funder.

Conflict of Interest: The authors declare that they have no conflict of interest.

ABSTRACT

The Strengths and Difficulties Questionnaire (SDQ) is a frequently used tool for universal screening of pre-schoolers' behavioural and emotional problems. However, evidence for its concurrent validity is equivocal and has not been tested in a Māori population. We aimed to evaluate the concurrent validity of the Strengths and Difficulties Questionnaire (SDQ) in Māori pre-schoolers (tamariki), aged 4 and 5. We carried out a prospective study of 225 tamariki (46% female) for whom a recent SDQ was available from the New Zealand Ministry of Health's Before School Check database. A trained nurse carried out a standardised wellbeing and behavioural assessment for these children. Sensitivity, specificity, positive predictive and negative predictive values were calculated, using optimal total Difficulty scale threshold values published for the SDQ (parent version SDQ-P; teacher version SDQ-T). Primary outcome: an assessment-based child referral to Child and Adolescent Mental Health Services or to a Paediatric outpatient service. Secondary outcomes: assessment-based parental referral to a parenting programme and combined referral. The optimal thresholds for child referral were low for the SDQ-P (13) and SDQ-T (7). Child referral SDQ-P: sensitivity 62%, specificity 83%, positive predictive value 0.35, negative predictive value 0.94. Child referral SDQ-T: sensitivity 77%, specificity 78%, positive predictive value 0.31, negative predictive value 0.96. The findings demonstrate optimal threshold values for referral for Māori on the SDQ-P and SDQ-T are much lower when compared to published thresholds (17 versus 16). Sensitivity values were also low. A surveillance approach for the assessment of psychosocial problems is recommended for pre-schoolers.

Keywords: Validation, Strengths and Difficulties Questionnaire, Māori, Pre-school, Screening

INTRODUCTION

Universal health programmes are offered in many countries from birth and throughout childhood. Such programmes aim to enhance health by offering screening, assessment, disease prevention strategies and health education of children and their families, and are often offered by nurses and health visitors in primary care (Blair & Hall, 2006). Blair and Hall (2006) argued that universal provision of such services has key advantages: by ensuring all parents are aware of services and support available since they are often the first ones to notice any problems in the development, hearing or vision of their children; promoting engagement in accessing routine health care by all, maximising involvement of those parents who may otherwise be difficult to locate; and preventing some parents from being seen as being targeted as “bad parents”. This is especially important in terms of behavioural and emotional problems in early childhood (e.g. pre-schoolers) because such problems can impact upon transition into primary school, lead to on-going problems in middle-childhood and adulthood, and affect educational achievement (Bierman et al., 2013; Eivers, Brendgen, & Borge, 2010; Kim-Cohen et al., 2009; Kim-Cohen et al., 2003; Poulou, 2015; White, Connelly, Thompson, & Wilson, 2013). Consequently, identification of children with behavioural problems at pre-school age is a priority in many countries in order to ensure support programmes can be put in place to enhance educational and health outcomes (Doughty, 2005). Indeed, it has been shown that in the course of delivering universal health programmes, nurses working with young children and their families identify many children with psychosocial problems and they often manage children with major mental health difficulties (Wilson et al., 2008).

One commonly used tool in universal programmes for children is the Strengths and Difficulties Questionnaire (SDQ) (R. Goodman, 1997). The SDQ asks about a child’s psychosocial attributes (positive and negative behaviours) with both parents and teachers completing the SDQ independently. The SDQ consists of 25 questions, each with 3 response options (not true, somewhat true, certainly true). The questions are grouped into five subscales, four measuring difficulties and one pro-social behaviour. Higher scores on the four subscales that report on difficulties reflect more significant problems, whereas lower scores on the prosocial subscale denote social behavioural difficulties. Scores from the first four subscales are summed to give an overall Difficulty score ranging from 0-40. SDQ thresholds were derived by comparing SDQ scores with psychiatric judgements of interviews with parents of just under 10,000 children in the UK (R. Goodman, Renfrew, & Mullick, 2000). Published score thresholds for each subscale, as well as the total Difficulties score, are used to classify children’s difficulties as “normal”, “borderline” and “abnormal”, with the specific aim to highlight whether children and parents should be offered further assessment and services (R. Goodman, 1997; Ministry of

Health, 2008). These thresholds are different for the parent version (SDQ-P) and the teacher version (SDQ-T). Enquiries through the SDQ website *Youth in Mind* have revealed that the SDQ has been translated in over seventy different languages and that SDQ translations are carried out in consultation with the SDQ team (personal communication, www.youthinmind.info). The questionnaire itself is readily available from this website. The SDQ has been used in American and Japanese research studies aiming to support parents' management of their children's emotions and behaviours (Fujiwara, Kato, & Sanders, 2011; Lakes, Vargas, Riggs, Schmidt, & Baird, 2011). It is routinely used with young people accessing mental health services in New South Wales, although questions have been raised about its validity with Aboriginal communities (Williamson et al., 2010). In New Zealand the SDQ is used as part of routine practice during the Before School Check (B4SC) with pre-schoolers aged 4 and 5, with over 50,000 taking place each year (Ministry of Health, 2008).

It is crucial that a tool such as the SDQ has proven ability to identify a good proportion of those children who really are in need of further assessment or treatment. In a recent systematic review of the SDQ, six studies were identified (median sample size of 500, range 86-7984) that examined the criterion validity of the SDQ (Kersten et al., 2016a). These were conducted in the UK, Australia, Spain and Bangladesh. They compared the scores from the total Difficulties scale and / or subscales to a "gold standard" clinical diagnostic interview with clinical samples, community samples, and children in care (Bekker, Bruck, & Sciberras, 2013; Ezpeleta, Granero, la Osa, Penelo, & Domènech, 2012; R. Goodman, Ford, Corbin, & Meltzer, 2004; R. Goodman et al., 2000; Mathai, Anderson, & Bourne, 2004; Mullick & Goodman, 2001). Four studies showed that the SDQ's sensitivity is inadequate (<70%) by criteria set out by an international outcome measurement group (Bekker et al., 2013; Ezpeleta et al., 2012; Mathai et al., 2004; Mokkink et al., 2010; Mullick & Goodman, 2001). Notably, one study reported sensitivity of 29.9% (Ezpeleta et al., 2012). Others reported sensitivity of 63% for "private household children" as rated by their parents, but 85% for "looked-after children" (i.e. children at foster homes or residential homes) as rated by their carers (R. Goodman et al., 2004). Goodman et al. (2000) reported high sensitivity (>80%) of three SDQ subscale scores (Conduct, Emotional, Hyperactivity) in identifying children who were clinically diagnosed with a disorder. However, this study was carried out with children referred to a multidisciplinary child mental health clinic rather than a general population.

Findings from the above studies suggest the SDQ is missing a significant proportion of children in the general population who should receive onward referral. In addition, one study analysed population SDQ data from seven countries and compared SDQ "caseness" (prevalence based on the mean total difficulty scores, adjusted for the population's age and sex composition) against the measured prevalence of disorder using the

Development and Well-being Assessment (DAWBA) tool (A. Goodman et al., 2012). They reported average coefficients of determination of $R^2=0.29$ and $R^2=0.56$ for the parent and teacher versions, respectively, concluding that SDQ scores cannot be compared cross-nationally without population-specific norms. This could have consequences for countries like Aotearoa New Zealand where the SDQ is used routinely with pre-schoolers' (Ministry of Health, 2008). In addition, for the New Zealand context it is important to consider the sizeable population that are Māori (14.9%) (Statistics New Zealand, 2013). Māori are the indigenous people of Aotearoa New Zealand, descended from the ocean-going navigators of Te Moana nui a Kiwa (the Pacific Ocean). Māori continue to be a young population with approximately 33% under 15 years old (Statistics New Zealand, 2013). Māori tamariki (children), similar to indigenous children from around the world, are recognised as facing health disparities that frequently begin in pre-school years (Bramley, Hebert, Tuzzio, & Chassin, 2005; King, Smith, & Gracey, 2009; Robson & Harris, 2007). We, and others, have shown previously that Māori generally prefer face to face interactions rather than questionnaires, a non-deficit approach in assessment, and the consideration of the historical context of colonisation (Elder & Kersten, 2015; Kersten et al., 2016b; Smith, 1999). It is therefore important to evaluate if the SDQ is valid for use with Māori.

This study aimed to examine the concurrent validity of the SDQ in tamariki (Māori word for children and young people) of pre-school age compared to the gold standard of clinical evaluation carried out by nurses with mental health and / or child experience.

METHOD

PARTICIPANTS

Family, whānau (extended family groups including parents) and caregivers (hereafter referred to as whānau) of 4- and 5- year old tamariki, for whom a recent SDQ questionnaire was available from the New Zealand Before School Check (B4SC) database held by the New Zealand Ministry of Health, were invited to take part in the study. Invitations were via letters and postcards introducing the study and research team, sent out by B4SC coordinators. Family and whānau were included if they had recently completed the SDQ for their tamariki; identified the child as Māori; lived in one of the five most northern District Health Board regions of New Zealand (from the Waikato to Northland) and provided informed consent. They were excluded if they were unavailable for assessment or had moved out of the region since the B4SC.

PROCEDURE

Tamariki were assessed by a nurse who had experience of mental health and / or child nursing. Twelve nurse assessors took part in a one-day training course in groups of 3-5. This included instruction on how to conduct a comprehensive assessment (clinical interview and observation of tamariki) by our study child and adolescent psychiatrist (HE) (Elder et al., 2016). The sessions included for example, common behavioural problems in pre-schoolers, Māori concepts of development, where children fit within the wider whānau, and how to conduct the semi-structured interview using questions adapted from the Ages and Stages Questionnaire (Squires et al., 2002). Training also included sessions on cultural safety, being a clinical researcher, the study methods, safety, ethics, care and protection issues and practice of case scenarios. Nurses were trained to follow Child Youth and Family guidelines around care and protection concerns (Child Youth and Family, 2011). Nurse assessors also took part in monthly debriefs by telephone, or more frequent when requested, both in small groups and individually. These were carried out by the principal investigator (PK) and HE. During these debriefs nurses and the research team could raise any concerns about tamariki, whānau, the assessment procedures and how decisions had been reached in terms of recommended referrals. This approach ensured they had appropriate supervision and that the quality of assessments was maintained. A small number of children in the Northland District Health Board were visited by two nurses, due to the nurses considering the locations too remote to travel alone. Nurse reliability checks were not carried out because we wished to mimic clinical practice. Nurses remained blind to the results of the SDQ.

Assessments followed a standardised but flexible approach and were carried out within whānau homes, to facilitate a more accurate and appropriate assessment of the behaviour and strengths and difficulties of the tamariki. At the start of the visit, nurses answered any questions from whānau and obtained written consent. They then carried out a clinical interview with whānau and observation of the tamariki. Tamariki were playing in the room during the visit and the nurses engaged the children during this time. The interview began with gaining an understanding of who's who in the whānau and how whānau members relate (whakawhanaungatanga), followed by a general discussion of the birth, any subsequent illness or injury and hospitalisations. This was followed by the aforementioned semi-structured interview, always using the context of a Te Ao Māori framework of concepts of development in their conclusions of the interview and observations (Ministry of Education, 2007). At the end of the interview, nurses discussed the findings of their assessment with the whānau, including whether they considered a referral for further assessment was warranted, any concerns they might have had about the tamariki and if they were concerned, any care and protection issues. Letters were sent to every participant's general practitioner, summarising the nurses' findings. Thus, if any

referral was needed but had not been picked up by routine services they could address this. In addition, nurses left information about community resources with whanau. Nurses carried out on average 19 assessments each (SD 13.6). Visits lasted between one and three hours (mean 79 minutes, SD 25).

Bias was managed from the point of view of the study population by inviting probabilistically selected eligible tamariki into the study and post-stratifying on gender-specific SDQ scores and gender itself. The selection was effected over time between 2012 and 2013, to ensure a reasonably short delay between SDQ administration and assessment. Given the relative homogeneity of our study population (single ethnicity and narrow age-band), we assumed that any non-response bias would be sufficiently mitigated by matching the sample to the target population on gender and SDQ score distribution, through post-stratification. Bias was also managed from the point of view of assessment by training and regular debriefs of nurse assessors, to ensure they had appropriate supervision and that the quality of assessments was maintained. In addition, nurses were blind to the SDQ scores previously carried out as part of the B4SC.

Nurses determined during the interview whether further referral was needed. Three referral outcomes were considered as gold-standard diagnoses. The primary outcome was **child referral**, consisting of an assessment-based referral to either: 1) Child and Adolescent Mental Health Services, or 2) to a Paediatric clinic, physician or service. Two secondary outcomes were identified: i) **parental referral** to a parenting programme and ii) **combined referral**, consisting of both a child and parental referral.

MEASURES

The Strengths and Difficulties Questionnaire obtained from the New Zealand B4SC database.

DATA ANALYSES

We post-stratified sampled children on distinct values of the SDQ Difficulty score (parent or teacher, depending upon the analysis) separately for each gender, and by gender also for the overall analysis. Nurse assessments of the need for further referral were used to calculate the sensitivity, specificity, positive predictive value and negative predictive values of the SDQ-P and the SDQ-T for all tamariki (see Table 1 for definitions of these terms). In addition, separate analyses were carried out for boys and girls. Logistic regression was used (using a quasibinomial model) to assess sensitivity and specificity at all observed thresholds in the sample. The approach was to regress the high (vs. low) Difficulty Score status determined for a given threshold (obtained from SDQ data available from the B4SC) on the referral status (obtained from the prospective assessments), considering every possible threshold in turn. Appropriately transformed estimates from these regressions yielded sensitivity and specificity estimates for each threshold, as well as weighted maximum likelihood-based standard

errors. The same approach was used for the estimation of positive and negative predictive values by reversing the role of score status and referral status. Thresholds that maximised the Youden index (sensitivity+specificity-1) were deemed optimal (Youden, 1950). Areas Under the Curve (AUC) of the Receiver Operating Characteristic curves were computed and reported. The survey package in R version 3.0.0 was used (Lumley, 2012; R Core Team, 2013). Parental and teacher questionnaire domain subscore results were available for all children included in this study. We considered all zero scores for the 5 subscores as spurious since the Prosocial scale is scored in the opposite direction from the four difficulty scales. This spurious pattern was not found in any of the parent questionnaires. However, the spurious pattern was observed in 150 teacher questionnaires, which were excluded from the teacher SDQ analyses.

Insert table 1 about here

SAMPLE SIZE

We assumed working values of 10% positive SDQ screening, 63.3% sensitivity and 94.6% specificity (R. Goodman et al., 2004). We determined that 248 assessments were required to obtain a standard error of 10 percentage points on the sensitivity estimate. With this number the standard error about the specificity was expected to be 0.7 percent.

RESULTS

A total of 3,009 written invitations to participate were sent, representing 54% of the eligible children over the period considered. 225 Tamariki were included in the study (46% female, 54% male; 20% aged 4, 80% aged 5). Interviewees were mostly parents (85% mothers, 5% fathers, 2% mothers and fathers interviewed together). In addition, 15 participants were grandparents, one a foster parent and one a great great aunt. Results are presented as two summary tables detailing the sensitivity, specificity, positive and negative predictive value for the SDQ-P (Table 2 and Table 3) and the SDQ-T (Table 4 and Table 5).

Concurrent validity SDQ-P for child referral: The optimal New Zealand threshold values for Māori child referral for the SDQ-P was 13 (all children, Table 2), significantly lower than the published threshold score of 17 and used as part of the New Zealand B4SC (R. Goodman, 1997; Ministry of Health, 2008); this optimal threshold was higher for girls than boys (14 vs 10).

The SDQ-P *sensitivity* for child referral was 62% (Table 2). In other words, out of all children who should be referred for further assessment (as determined by the gold standard assessment carried out by the nurses), 62% were correctly classified by the SDQ-P (i.e. falling on or above our optimal SDQ-P threshold value of 13); and 38% of children who should be referred were being missed by the SDQ-P (false negatives). SDQ-P sensitivity was better for boys than for girls (85% versus 61%).

Insert table 2 about here

The SDQ-P *specificity* for child referral was 83% (Table 2); out of all children who should not be referred for further assessment (as determined by the gold standard assessment), 83% were correctly classified by the SDQ-P as not requiring referral (i.e. falling below our optimal SDQ-P threshold value of 13) and 17% of children who should not be referred were falsely classified as needing referrals (false positives, i.e. falling on or above our optimal SDQ-P threshold value of 13). SDQ-P specificity was better for girls than boys (88% versus 66%).

Using our optimal threshold of 13 resulted in a *positive predictive value* estimate for child referral of 0.35 (Table 3); i.e. 35% of children with SDQ-P scores on or above the threshold were correctly classified as needing a referral, and 65% were false positives and classified as needing referral but did not have problematic behaviours (as determined by nurse assessments). The *negative predictive value* was 0.94, meaning 94% of children classified by the SDQ-P as not needing referral were correctly classified as such (Table 3).

Insert table 3 about here

When Goodman's (2007) published threshold for the SDQ-P (i.e. 17) was applied to our data, specificity for child referral was improved (94%) but sensitivity was much lower (28%).

Concurrent validity SDQ-P for parent referral: The optimal threshold value for parental referral for the SDQ-P was 10 (Table 2). For girls the optimal threshold was 14, for boys 10. The SDQ-P sensitivity for parental referral was 77% and specificity was 67% (Table 2); the positive predictive value estimate was 0.29 and the negative predictive value estimate was 0.94 (Table 3).

Concurrent validity SDQ-P for combined (child and parent) referral: We identified the optimal threshold value for combined referral for the SDQ-P to be 10 (14 for girls and 11 for boys). The SDQ-P

sensitivity for combined referral was 76% and specificity was 71% (Table 2). The positive predictive value estimate for combined referral was 0.41 and the negative predictive value estimate was 0.92 (Table 3).

Concurrent validity SDQ-T for child referral: The optimal threshold values for child referral for the SDQ-T was 7 (10 for girls and 7 for boys, Table 4), lower than the published threshold score of 16 (R. Goodman, 1997). The SDQ-T sensitivity for child referral was 77% and specificity was 78% (Table 4). Both sensitivity and specificity of the SDQ-T were better for girls than boys. The positive predictive value estimate for child referral was 0.31 and the negative predictive value estimate was 0.96 (Table 5). When the published threshold for the SDQ-T (i.e. 16) was applied to our data, we noted an increase in specificity for child referral (94%) but sensitivity dropped down to 0% (R. Goodman, 1997).

Insert table 4 about here

Concurrent validity SDQ-T for parent referral: The optimal threshold value for parental referral for the SDQ-T was 10 (Table 4, 10 for girls and 12 for boys). The SDQ-T sensitivity for parental referral was 60% (38% for girls and 73% for boys) and specificity for parental referral was 87% (87% for girls and 88% for boys) (Table 4). The SDQ-T positive predictive value estimate for parental referral was 0.24 and the SDQ-T negative predictive value estimate for parental referral was 0.97 (Table 5).

Insert table 5 about here

Concurrent validity SDQ-T for combined referral: The optimal threshold value for combined referral for the SDQ-T was 7 (10 for girls, 7 for boys). The SDQ-T sensitivity for combined referral was 75% (65% for girls and 80% for boys) and specificity was 80% (91% for girls and 75% for boys) (Table 4). The SDQ-T positive predictive value estimate for combined referral was 0.41 and the SDQ-T negative predictive value estimate for combined referral was 0.95 (Table 5).

DISCUSSION

Our findings showed the optimal thresholds for Māori child referral were lower than the published threshold of 17 for the SDQ-P, which are being used in New Zealand (R. Goodman, 1997). Consequently, using the published SDQ thresholds leads to unacceptably low sensitivity values for child referral (28% for SDQ-P

and 0% for SDQ-T) and children who may benefit from support or intervention failing to be identified. This is especially important when considering the proportion of children with conduct problems is greater in Māori (15-20%) than non-Māori (5-10%) in New Zealand (The Advisory Group on Conduct Problems, 2009). In addition, there was a marked difference in optimal threshold values for girls and boys. Although sensitivity was low in our study, it was higher than that found in a number of studies. For example, a UK based study reported sensitivity of 47% and a Spanish study 29.9%, in both cases at threshold value 17 (Ezpeleta et al., 2012; R. Goodman, 2001). When the latter study lowered the threshold value to 13, sensitivity improved to 52.1%. In a Norwegian study sensitivity was lower and specificity higher than that found in our study but with a lower cut off value (≥ 10) (Sveen, Berg-Nielsen, Lydersen, & Wichstrøm, 2013).

Low sensitivity rates lead to higher than appropriate false negatives. If a screening tool is used as part of a rolling surveillance programme it is possible to pick up false negatives in subsequent screening rounds, as is the case in some countries. However, in New Zealand the SDQ is only used as part of the B4SC, the last routine Tamariki Ora (the New Zealand Well Child Programme) assessment carried out by nurses. A recent qualitative study indicated that for Māori parents, cultural values and perspectives might impact on people's willingness or ability to be open when answering the questions, and that this may lead to the possibility that people do not complete the questionnaire honestly (Kersten et al., 2016b). This could be one reason for the low threshold values.

Prevalence rates of conduct disorders have previously been reported as much higher in Māori children than the wider population. For example, at age 8, 15% of Māori compared to 10% of non-Māori were estimated to have significant conduct problems, an odds ratio of 1.5 when compared with non-Māori (The Advisory Group on Conduct Problems, 2009). Similarly, a recent audit of referrals based on SDQ data from Hawke's Bay (New Zealand) found Māori children were disproportionately represented in those referred (47% compared to 24.5% being Māori in the region) (Hedley et al., 2012). Thus, with the relatively poor sensitivity of the SDQ in this population and higher prevalence of conduct disorders, a surveillance approach to screening children's psychosocial problems should be considered.

Specificity values were low, more so for boys than for girls (66% versus 88%). Similarly, positive predictive values were low for both boys and girls. This results in a large number of false positives and an unacceptable burden on families and whānau, and indeed services if only SDQ scores were used to decide whether referral is warranted. By contrast, Goodman and colleagues (2001) presented a specificity of 94% with a threshold of 17. At that threshold, our specificity was identical but sensitivity dropped down to 28%,

unacceptably low. However, New Zealand DHB's Advisory Boards or triage teams, which include nurses, do not only rely on SDQ scores; they discuss children with high SDQ scores before referrals are made (Hedley et al., 2012). This practice should potentially avoid some unnecessary referrals. However, given the low sensitivity found in our study the discussion should also include children with lower scores and for whom concerns are raised by the assessing nurse.

Our study found an even lower optimal threshold for the SDQ-T (i.e. 10 for girls, 7 for boys) as opposed to 16 as recommended by Goodman (1997). Yet, the sensitivity of the SDQ-T was extremely high for girls in our study (100%) with 91% specificity and for boys respective values were 72% and 70%. This compares to sensitivity values of 43% and specificity values of 95% in the British national cohort (threshold value 16) (R. Goodman, 2001). If the New Zealand threshold was raised accordingly, sensitivity would drop to 0% both for boys and girls. In other words, the recommended threshold value would not pick up any cases at all. Our findings are similar to the study conducted in Spain, which found that lowering SDQ-T threshold values to 6 resulted in marked improvements in sensitivity (from 18.5% to 72.1%) but at the expense of specificity (from 89.2% to 44.3%) (Ezpeleta et al., 2012).

Optimal threshold values for parental referral and combined child-parental referral were almost identical for boys and girls as for child-referrals, suggesting that nurses should consider both referral for a child and a parent if difficult behaviour or emotional problems are identified. Parental referral and combined referral thresholds have not previously been reported in the literature, therefore, we cannot compare our findings with that from others. A retrospective audit in Hawke's Bay (New Zealand) showed that 71% of parents and whānau of children whose SDQ-P scores were greater than 15 were referred to Non-Government Organisations providing parenting services for families, with a further 8% as either enrolled in or on waiting lists for the Incredible Years programmes (programmes for parents, children, and teachers, aiming to prevent and treat young children's behaviour problems and promote their social, emotional, and academic competence) (Hedley et al., 2012). This group suggested that the child's behaviour was perceived to mainly stem from child management or family relationship issues for which support for the parents from such programmes would best benefit the child. However, these programmes lack a review of the particular issues for indigenous people and how they may or may not be efficacious. It is important, therefore, that such programmes and the impact upon the behavioural and emotional wellbeing of tamariki are robustly evaluated.

Our previous qualitative work suggested that Māori parents considered the process of administering the SDQ and the tool itself as a Pākehā (New Zealander of European descent) approach to labelling their children,

that they preferred a non-deficit approach, and that they were concerned about providing affirmative responses to negative questions (Kersten et al., 2016b). This may impact on the cultural and face validity of the SDQ for Māori and inadvertently lead to them reporting fewer difficulties on the tool. This would result in problems going undetected and children needing further support slipping through the net. This underscores the importance of culturally appropriate ways of working and referral pathways with Māori families, as we did in this study (Elder et al., 2016).

STRENGTHS AND LIMITATIONS

The strengths of our study include the training of our nurse assessors with ongoing debriefs, which ensured a standardised approach to assessments (Elder et al., 2016). In addition, the study included tamariki from different areas of the country, both from urban and rural settings. It also included tamariki who scored across the full range of the SDQ, which is important as research has shown that even with low scores false negative cases arise (A. Goodman & Goodman, 2009, 2011; A. Goodman et al., 2012).

A limitation was that the SDQ was not completed at the exact same time as the nurse assessment. This was a decision made in consultation with our stakeholders, which suggested that a questionnaire-based assessment could undermine the relationship the nurse would be required to develop in a relatively short assessment period (1 to 2 hours). In addition, a large number of teacher questionnaires (150) were excluded from the analysis as their data patterns were spurious (e.g. all zero scores). This could be a result of data entry systems, although we were unable to assess this retrospectively. This study focused on tamariki only and we can therefore not make generalisations from our findings to other ethnic New Zealand groups. The appropriateness of the published thresholds for these groups is therefore warranted.

All study participants were sent a lay summary and the results were reported to the Ministry of Health in New Zealand who are considering these with their clinical advisors. We recommend that the SDQ-P and SDQ-T Total Difficulties scale thresholds be lowered for Māori pre-schoolers New Zealand and that on-going evaluation of routinely available data be carried out to examine if such thresholds result in more efficient identification of children needing onward referral or treatment. In addition, we recommend that the study population is followed longitudinally and that the concurrent validity of other tools be evaluated for the New Zealand pre-school population.

COMPLIANCE WITH ETHICAL STANDARDS

The study was carried out in accordance with the 1964 Declaration of Helsinki and its later amendments. Ethics approval was gained from the Health and Disability Ethics Committee (ref: NTY/12/04/028) and University's Ethics Committee (ref: 12/63). All study participants provided informed consent.

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

Conflict of Interest: The authors declare that they have no conflict of interest.

AUTHOR CONTRIBUTIONS

PK: designed and executed the study, assisted with the data analyses, and wrote the paper. AV: analysed the data and wrote part of the results. HE: collaborated with the design and writing of the paper, and assisted with the nurse training and support. RT: assisted with recruitment and organising clinical visits. KM: collaborated with the design and writing of the paper and assisted with the nurse training and support.

REFERENCES

- Bekker, J., Bruck, D., & Sciberras, E. (2013). Congruent validity of the strength and difficulties questionnaire to screen for comorbidities in children with ADHD. *Journal of attention disorders*, 1-11.
- Bierman, K. L., Coie, J., Dodge, K., Greenberg, M., Lochman, J., McMohan, R., Pinderhughes, E., Coie, J. D., Dodge, K. A., Greenberg, M. T., Lochman, J. E., & McMahon, R. J. (2013). School Outcomes of aggressive-disruptive children: prediction from kindergarten risk factors and impact of the fast track prevention program. *Aggressive behavior*, 39(2), 114-130.
- Blair, M., & Hall, D. (2006). From health surveillance to health promotion: the changing focus in preventive children's services. *Archives of disease in childhood*, 91(9), 730-735.
- Bramley, D., Hebert, P., Tuzzio, L., & Chassin, M. (2005). Disparities in indigenous health: a cross-country comparison between New Zealand and the United States. *American journal of public health*, 95(5), 844-850.
- Child Youth and Family. (2011). *An interagency guide. Working together to keep children and young people safe*.
- Doughty, C. (2005). The effectiveness of mental health promotion, prevention and early intervention in children, adolescents and adults. NZHTA Report ; 8(2).
- Eivers, A. R., Brendgen, M., & Borge, A. I. H. (2010). Stability and change in prosocial and antisocial behavior across the transition to school: Teacher and peer perspectives. *Early education and development*, 21(6), 843-864.
- Elder, H., & Kersten, P. (2015). Whakawhiti kōrero, a method for the development of a cultural assessment tool, Te Waka Kuaka, in Māori traumatic brain injury. *Behavioural neurology*, Article ID 137402.
- Elder, H., Kersten, P., McPherson, K., Dawson, L., Harrison, H., Harrison, J., & Poulter, A. (2016). Making time: deeper connection, fuller stories, best practice. *Annals of psychiatry and mental health*, 4(4).
- Ezpeleta, L., Granero, R., la Osa, N. d., Penelo, E., & Domènech, J. M. (2012). Psychometric properties of the strengths and difficulties questionnaire (3-4) in 3-year-old preschoolers. *Comprehensive psychiatry*, 54(3), 282-291.
- Fujiwara, T., Kato, N., & Sanders, M. R. (2011). Effectiveness of group positive parenting program (Triple P) in changing child behavior, parenting style, and parental adjustment: an intervention study in Japan. *Journal of child and family studies*, 20(6), 804-813.
- Goodman, A., & Goodman, R. (2009). Strengths and difficulties questionnaire as a dimensional measure of child mental health. *Journal of the American academy of child and adolescent psychiatry*, 48(4), 400-403.
- Goodman, A., & Goodman, R. (2011). Population mean scores predict child mental disorder rates: Validating SDQ prevalence estimators in Britain. *Journal of child psychology and psychiatry and allied disciplines*, 52(1), 100-108.
- Goodman, A., Heiervang, E., Fleitlich-Bilyk, B., Alyahri, A., Patel, V., Mullick, M. S. I., Slobodskaya, H., Dos Santos, D. N., & Goodman, R. (2012). Cross-national differences in questionnaires do not necessarily reflect comparable differences in disorder prevalence. *Social psychiatry and psychiatric epidemiology*, 47(8), 1321-1331.
- Goodman, R. (1997). The strengths and difficulties questionnaire: a research note. *Journal of child psychology and psychiatry and allied disciplines*, 38(5), 581-586.
- Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American academy of child and adolescent psychiatry*, 40(11), 1337-1345.
- Goodman, R., Ford, T., Corbin, T., & Meltzer, H. (2004). Using the strengths and difficulties questionnaire (SDQ) multi-informant algorithm to screen looked-after children for psychiatric disorders. *European child and adolescent psychiatry, Supplement*, 13(2), II/25-II/31.
- Goodman, R., Renfrew, D., & Mullick, M. (2000). Predicting type of psychiatric disorder from strengths and difficulties questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *European child and adolescent psychiatry*, 9(2), 129-134.

- Hedley, C., Thompson, S., Morris Mathews, K., Pentecost, M., Wivell, J., Stockdale Frost, A., & Morris, H. (2012). The B4 School Check behaviour measures: findings from the Hawkes' bay evaluation. *Nursing praxis in New Zealand*, 28(3), 13-23.
- Kersten, P., Czuba, K., McPherson, K., Dudley, M., Elder, H., Tauroa, R., & Vandal, A. (2016a). A systematic review of evidence for the psychometric properties of the Strengths and Difficulties Questionnaire. *International journal of behavioral development*, 40(1), 64-75.
- Kersten, P., Dudley, M., Nayar, S., Elder, H., Robertson, H., Tauroa, R., & McPherson, K. M. (2016b). Cross-cultural acceptability and utility of the strengths and difficulties questionnaire: views of families. *BMC psychiatry*, 16(1), 347.
- Kim-Cohen, J., Arseneault, L., Newcombe, R., Adams, F., Bolton, H., Cant, L., Delgado, K., Freeman, J., Golaszewski, A., & Moffitt, T. E. (2009). Five-year predictive validity of DSM-IV conduct disorder research diagnosis in 41/2-5-year-old children. *European child and adolescent psychiatry*, 18(5), 284-291.
- Kim-Cohen, J., Caspi, A., Moffitt, T. E., Harrington, H., Milne, B. J., & Poulton, R. (2003). Prior juvenile diagnoses in adults with mental disorder: developmental follow-back of a prospective-longitudinal cohort. *Archives of general psychiatry*, 60(7), 709-717.
- King, M., Smith, A., & Gracey, M. (2009). Indigenous health part 2: the underlying causes of the health gap. *The lancet*, 374, 76-85.
- Lakes, K. D., Vargas, D., Riggs, M., Schmidt, J., & Baird, M. (2011). Parenting intervention to reduce attention and behavior difficulties in preschoolers: a CUIDAR evaluation study. *Journal of child and family studies*, 20(5), 648-659.
- Lumley, T. (2012). Survey: analysis of complex survey samples. R package version 3.28-2.
- Mathai, J., Anderson, P., & Bourne, A. (2004). Comparing psychiatric diagnoses generated by the strengths and difficulties questionnaire with diagnoses made by clinicians. *Australian and New Zealand journal of psychiatry*, 38(8), 639-643.
- Ministry of Education. (2007). Atawhaingia Te Pa Harakeke: training & professional support manual. Wellington
- Ministry of Health. (2008). The B4 School Check. A handbook for practitioners. Wellington.
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of clinical epidemiology*, 63(7), 737-745.
- Mullick, M. S., & Goodman, R. (2001). Questionnaire screening for mental health problems in Bangladeshi children: a preliminary study. *Social psychiatry and psychiatric epidemiology*, 36(2), 94-99.
- Poulou, M. S. (2015). Emotional and behavioural difficulties in preschool. *Journal of child and family studies*, 24(2), 225-236.
- R Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Robson, B., & Harris, R. (2007). *Hauora: Māori standards of health IV. A study of the years 2000-2005*. Wellington: Te Rōpū Rangahau Hauora a Eru Pomare.
- Smith, L. T. (1999). *Decolonizing methodologies* (9th impression ed.). Dunedin: University of Otago Press.
- Squires, J., Bricjer, D., Twombly, E., Yockelson, S., Schoen Davis, M., & Kim, Y. (2002). *Ages & Stages Questionnaires®: Social-Emotional*. Paul H. Brookes Publishing Co.
- Statistics New Zealand. (2013). 2013 Census.
- Sveen, T. H., Berg-Nielsen, T. S., Lydersen, S., & Wichstrøm, L. (2013). Detecting psychiatric disorders in preschoolers: screening with the strengths and difficulties questionnaire. *Journal of the american academy of child and adolescent psychiatry*, 52(7), 728-736.
- The Advisory Group on Conduct Problems. (2009). Conduct problems. Best practice report.
- White, J., Connelly, G., Thompson, L., & Wilson, P. (2013). Assessing wellbeing at school entry using the strengths and difficulties questionnaire: professional perspectives. *Educational research*, 55(1), 87-98.

Williamson, A., Redman, S., Dadds, M., Daniels, J., D'Este, C., Raphael, B., Eades, S., & Skinner, T. (2010). Acceptability of an emotional and behavioural screening tool for children in aboriginal community controlled health services in urban NSW. *Australian and New Zealand journal of psychiatry*, *44*(10), 894-900.

Wilson, P., Furnivall, J., Barbour, R. S., Connelly, G., Bryce, G., Phin, L., & Stallard, A. (2008). The work of health visitors and school nurses with children with psychological and behavioural problems. *Journal of advanced nursing*, *61*(4), 445-455 411p.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*, 32-35.

Table 1 Explanation of statistical terms used in the data analysis

Statistical term	Definition	Context
Sensitivity (true positive rate)	The proportion of positives that are correctly identified as such	The percentage of tamariki who are correctly identified by the SDQ* as requiring onward referral for their emotional and behavioural difficulties
Specificity (true negative rate)	The proportion of negatives that are correctly identified as such	The percentage of tamariki who are correctly identified by the SDQ as not requiring onward referral for their emotional and behavioural difficulties
Positive predictive values	The ratio of true positives to combined true and false positives	True positives: those tamariki deemed to require onward referral for their emotional and behavioural difficulties following clinical assessment (gold standard) Combined true and false positives: those tamariki identified by the published SDQ thresholds to require onward referral for their emotional and behavioural difficulties
Negative predictive value	The ratio of true negatives to combined true and false negatives	True negatives: those tamariki deemed not to require onward referral for their emotional and behavioural difficulties following clinical assessment (gold standard) Combined true and false negatives: those tamariki identified by the published SDQ thresholds to not require onward referral for their emotional and behavioural difficulties

* SDQ = Strengths and Difficulties Questionnaire

Table 2 Youden-optimal sensitivity, specificity: Parent SDQ Difficulty score, New Zealand Māori 4 and 5 year-old children 2012-2013

Referral type	Group	Optimal Threshold SDQ-P	AUC	Youden Index ^a	Sensitivity SDQ-P*			Specificity SDQ-P		
					Estimate	95% conf. Int.		Estimate	95% conf. Int.	
						lower	upper		lower	upper
Child	All	13	77	45	62	43	78	83	80	86
	Female	14	70	49	61	21	90	88	82	92
	Male	10	79	51	85	65	94	66	61	72
Parent	All	10	77	44	77	61	88	67	64	71
	Female	14	81	53	62	38	82	91	88	93
	Male	10	72	39	76	54	89	64	59	68
Combined	All	10	78	47	76	63	86	71	66	75
	Female	14	73	46	54	35	72	92	88	94
	Male	11	79	55	77	59	89	78	69	84

* *SDQ-P* = *Strengths and Difficulties Questionnaire – Parent completed*; ^a Sensitivity + Specificity - 1 (Youden, 1950)

Table 3 Positive and negative predictive values: Parent SDQ Difficulty score, New Zealand Māori 4 and 5 year-old children 2012-2013

Referral type	Group	Optimal Threshold SDQ-P	Positive predictive value SDQ-P				Negative predictive value SDQ-P			
			Estimate	Standard Error	95% conf. Int.		Estimate	Standard Error	95% conf. Int.	
					lower	upper			lower	upper
Child	All	13	0.348	0.071	0.223	0.497	0.938	0.0204	0.884	0.968
	Female	14	0.278	0.183	0.061	0.696	0.967	0.0197	0.897	0.990
	Male	10	0.346	0.081	0.208	0.516	0.955	0.0230	0.882	0.984
Parent	All	10	0.291	0.057	0.193	0.414	0.943	0.0190	0.892	0.971
	Female	14	0.492	0.087	0.329	0.657	0.945	0.0264	0.864	0.979
	Male	10	0.284	0.061	0.180	0.417	0.932	0.0308	0.841	0.973
Combined	All	10	0.413	0.063	0.297	0.539	0.916	0.0233	0.858	0.952
	Female	14	0.556	0.087	0.386	0.714	0.912	0.0328	0.823	0.959
	Male	11	0.537	0.099	0.347	0.717	0.912	0.0349	0.815	0.960

* SDQ-P = Strengths and Difficulties Questionnaire – Parent completed

Table 4 Youden-optimal sensitivity, specificity: Teacher SDQ Difficulty score, New Zealand Māori 4 and 5 year-old children 2012-2013

Referral type	Group	Optimal Threshold SDQ-T	AUC	Youden Index ^a	Sensitivity SDQ-T*			Specificity SDQ-T		
					Estimate	95% conf. Int.		Estimate	95% conf. Int.	
						lower	upper		lower	upper
Child	All	7	74	56	77	42	94	78	72	83
	Female	10	95	91	100	29	100	91	84	96
	Male	7	65	42	72	31	93	70	59	79
Parent	All	10	68	47	60	20	90	87	82	90
	Female	10	40	25	38	4	91	87	83	91
	Male	12	81	61	73	20	97	88	77	94
Combined	All	7	74	55	75	45	91	80	72	86
	Female	10	65	56	65	17	94	91	83	96
	Male	7	77	55	80	44	95	75	59	87

* *SDQ-T = Strengths and Difficulties Questionnaire – Teacher completed*; ^a Sensitivity + Specificity - 1 (Youden, 1950)

Table 5 Positive and negative predictive values: Teacher SDQ Difficulty score, New Zealand Māori 4 and 5 year-old children 2012-2013

Referral type	Group	Optimal threshold SDQ-T	Positive predictive value SDQ-T*				Negative predictive value SDQ-T			
			Estimate	Standard Error	95% conf. Int.		Estimate	Standard Error	95% conf. Int.	
					lower	upper			lower	upper
Child	All	7	0.318	0.118	0.138	0.576	0.963	0.0259	0.862	0.991
	Female	10	0.427	0.218	0.115	0.811	1.000	0.1178	0.692	1.000
	Male	7	0.301	0.152	0.094	0.639	0.933	0.0477	0.758	0.984
Parent	All	10	0.243	0.144	0.064	0.599	0.969	0.0223	0.880	0.992
	Female	10	0.141	0.141	0.016	0.617	0.962	0.0379	0.768	0.995
	Male	12	0.346	0.250	0.057	0.822	0.974	0.0260	0.833	0.996
Combined	All	7	0.413	0.133	0.193	0.674	0.945	0.0310	0.842	0.982
	Female	10	0.427	0.218	0.115	0.811	0.962	0.0379	0.768	0.995
	Male	7	0.463	0.192	0.159	0.797	0.933	0.0477	0.758	0.984

* SDQ-T = Strengths and Difficulties Questionnaire – Teacher complete

