The Nordic advantage in child mental health: Separating health differences from reporting style in a cross-cultural comparison of psychopathology

Einar Heiervang[1], Anna Goodman[2], Robert Goodman[3]

[1]Centre for Child and Adolescent Mental Health, Bergen Norway
[2]London School of Hygiene & Tropical Medicine, London UK
[3]King's College London Institute of Psychiatry, London UK

Corresponding author: Einar Heiervang, Centre for Child and Adolescent Mental Health, PO Box 7800, N-5020 Bergen, Norway Phone: +4799568583 Fax: +4755588379 E-mail: einar.heiervang@rbup.uib.no

Abstract

Background: The use of similar standardized measures of psychopathology for population surveys permits cross-cultural comparisons. However, interpretation of findings can be challenging because rating thresholds may differ across cultures. By combining questionnaire and interview data, we explore whether lower questionnaire scores in Norway as compared to Britain reflect genuine differences in child mental health, or simply different reporting thresholds.

Methods: Information from the Strengths and Difficulties Questionnaire (SDQ) and the Development and Well-Being Assessment (DAWBA) interview were compared across recent population surveys in Norway and Britain. The Norwegian study (2002-2003) had questionnaire data for 6,658 and interview data for 1,024 8-10 year-old children. The British dataset included questionnaire and interview data for 4,898 children of the same age range from two independent surveys (1999 and 2004).

Results: Norwegian children had lower SDQ scores on all problem scales (emotional, behavioral, hyperactive and peer relationship) according to parents as well as teachers. DAWBA information showed that the Norwegian prevalence of externalizing disorders (behavioral and hyperactivity) was about half that found in Britain, whereas rates of emotional disorders were similar. Norwegian and British children with non-emotional disorders had similar questionnaire scores and rates of problem-recognition by parents and teachers. By contrast, questionnaire scores and problem-recognition were all lower in Norwegian children with emotional disorders.

Conclusions: Lower Norwegian questionnaire scores for externalizing problems appear to reflect real and substantial differences between the two countries. By contrast, lower questionnaire scores for emotional problems seem to reflect under-reporting/under-recognition by Norwegian adults, and not a genuinely lower prevalence of emotional disorders. This illustrates that cross-cultural differences in psychopathology based only on questionnaire data may be misleading. Nevertheless, careful use of questionnaire and interview data can focus research in mental health on cross-cultural variations likely to reflect genuine health differences.

Keywords: Cross-cultural, comparison, child, psychopathology, informant.

Abbreviations: SDQ: Strengths and Difficulties Questionnaire; DAWBA: Development and Well-Being Assessment; ADHD: Attention deficit hyperactivity disorder.

Introduction

Standardized measures of psychopathology are increasingly used to compare rates of mental health problems across space, time and culture. For child mental health such comparisons have mostly been limited to questionnaires, and have not included interview data (Achenbach et al., 1990; Weine, Phillips, & Achenbach, 1995; Crijnen, Achenbach, & Verhulst, 1997; Goodman, Renfrew, & Mullick, 2000; Obel et al., 2004; Becker et al., 2006). One particularly well-replicated finding is the lower problem scores on questionnaires completed by parents and teachers for children from the Nordic countries (Denmark, Finland, Iceland, Norway and Sweden) compared to other developed countries (Nøvik, 1999; Larsson & Frisk, 1999; Bilenberg, 1999; Koskelainen, Sourander, & Kaljonen, 2000; Broberg et al., 2001)

Lower mean scores do not, however, necessarily prove that Nordic children are particularly well adjusted – it could equally be true that Nordic adults require a higher intensity or frequency of any given symptom before they will endorse the corresponding questionnaire item. It is important to distinguish between genuine health differences and relative underreporting which, from a public health perspective, could be considered an artefact. Genuine health differences may hold important clues for aetiology and prevention; reporting biases have different implications, mainly for detection and referral.

In this paper we compare 8-10 year olds from Norway and Britain in terms of differences in SDQ scores and DAWBA-based prevalence rates, and also investigate the relationship between questionnaire scores and investigator-based diagnoses in the two countries. Our primary purpose is to investigate the extent to which the apparent Nordic advantage in child mental health reflects a genuine health difference or a bias from underreporting.

Methods

British sample (England, Scotland And Wales)

Two nationally-representative cross-sectional surveys of the mental health of children and adolescents in Great Britain were carried out in 1999 and 2004. Both surveys were population-based, using the British Child Benefit Register as a sampling frame; in Britain child benefit is a universal state benefit payable for each child in the family with extremely high uptake. The surveys sampled children using a clustered design by postal sector - full details of the survey design have been published elsewhere (Meltzer, Gatward, Goodman, & Ford, 2000; Green, McGinnity, Meltzer, Ford, & Goodman, 2004). Between the two British surveys 26,544 children and adolescents were selected and their principal caregivers (referred to as 'parents') approached. Of these 18,415 (69%) participated. In order to match the age range of the Norwegian sample, we only included 8-10 year old children in these analyses. This left a final British sample of 4,898 children, all of whom had SDQ and DAWBA information. The mean age was 9.0 years, and 50.6% were male. When participating parents gave their permission, teachers were also approached, resulting in teacher SDQ and DAWBA information for 3,836 (78%).

Norwegian sample (Bergen)

The Bergen Child Study has been described in detail elsewhere (Heiervang et al., 2007). Bergen is the second largest city of Norway, with a total population of around 235,000. All 79 primary schools in Bergen took part in the Bergen Child Study in 2002 and 2003. Between them, they had 9,430 children attending grades 2-4. In this paper we include only children with parent and teacher reports or parent-only reports, in order to achieve greater comparability with the British sample. The parents of 6,658 (71%) children completed a valid SDQ and gave permission for teachers to complete the SDQ. The mean age of these children was 9.0 years and 50.1% were male.

The Norwegian study was a 2-phase study, with information from the SDQ and some supplementary screening questions being used to divide the children into 1,530 who were screen positive and 5,128 who were screen negative. The parents of all screen positive children and a random 15% of screen negative children were invited to take part in the DAWBA phase. DAWBA information was obtained on 1,024 individuals: 655 who had been screen positive and 369 who had been screen negative, representing participation rates of 43% and 48% respectively. These low response rates are a cause for concern, since they might introduce bias in our estimate of rates of psychopathology in Norwegian children. Given that the SDQ score is a powerful predictor of DAWBA diagnosis (see below), one would expect a difference between participants and non-participants in likelihood of receiving a diagnosis to correspond to a difference in mean SDQ scores. Mean parent SDQ scores in participants and non-participants were, however, very similar in both screen-positive children (11.2 and 11.4 respectively, $p=0.47$) and screen-negative children (4.2 in both cases, $p=0.90$). Mean teacher SDQ scores of participants and non-participants were likewise very similar in screen-positive children (9.6 and 9.2 respectively, $p=0.22$) and screen-negative children (2.6 and 2.4 respectively, $p=0.29$). The lack of any differences in teacher SDQ scores is particularly reassuring as it represents an assessment of mental health independent of the parent who took the decision as to whether to participate in the second DAWBA phase or not. Participants and non-participants also showed only small and non-significant ($p>0.1$) differences in the proportion of males or mean age, these being the two other characteristics known about non-participants.

For all participating children a written informed consent was obtained from their carers.

Measures

The Strengths and Difficulties Questionnaire (SDQ) is a brief behavioural screening questionnaire that asks about 25 attributes, some positive and some negative (Goodman, 1997; Goodman, 1999; Goodman, 2001). The 25 items are divided between five scales of five items each, generating scores for emotional symptoms, conduct problems, hyperactivity-inattention, peer problems and prosocial behaviour; all but the last one are summed to generate a total difficulties score. The extended version of the SDQ also includes an impact supplement that includes the question of whether the respondent

thinks that the child or teenager has a problem, and if so, enquires further about impact (distress or social impairment) and burden.

For the British dataset, the SDQ has been demonstrated to have the predicted five-factor internal structure, satisfactory internal consistency, and a cross-informant correlation which compares favourably with that of other questionnaire measures (Goodman, 2001). Multi-informant SDQs have also been shown to have high specificity and moderate sensitivity in predicting disorder, as judged against both diagnoses made in psychiatric clinics (Goodman et al., 2000a) and by using the DAWBA (Goodman et al., 2000b). In Norway there has thus far been relatively little formal assessment of reliability and validity of the SDQ, although the factor structure and screening properties reported so far are generally in line with the British findings (Ronning, Handegaard, Sourander, & Morch, 2004; Hysing, Elgen, Gillberg, Lie & Lundervold, 2007). Part of the purpose of this paper is to present evidence on the relationship between the Norwegian SDQ and the Norwegian DAWBA.

The DAWBA includes a detailed psychiatric interview for parents and a briefer questionnaire for teachers (Goodman, Ford, Richards, Gatward, & Meltzer, 2000). It can be administered by lay interviewers and includes a fully-structured section supplemented by open-ended prompts in which a verbatim record is made of a respondent's own description of problem areas. A computer algorithm uses responses to the fully structured questions to predict likely diagnoses. Experienced clinicians review the computer predictions in the light of both the closed and open responses from all informants before assigning diagnoses according to DSM-IV (American Psychiatric Association, 1994). In Britain, the DAWBA has been shown to be able to discriminate well between community and clinic samples and between different diagnoses (Goodman et al., 2000b). Both in Britain and Norway the DAWBA generated reasonable prevalence estimates for mental disorders, plausible comorbidity and risk factor associations, and was strongly predictive of mental health service contact (Meltzer, Gatward, Corbin, Goodman, & Ford, 2003; Heiervang et al.. 2007).

All the clinical raters of DAWBA diagnoses in both Britain and Norway were trained by and subsequently supervised by the author (RG) who developed the instrument. High inter-rater reliabilities have previously been reported for the British and Norwegian studies (Green et al., 2004; Heiervang et al., 2007).

Statistical analyses

We carried out all analyses using the svyset options in STATA 9.2, to permit appropriate adjustment for the clustered sampling design of the British surveys. The svyset command was also used to weight the Norwegian data to allow for the two-phase design, in which 100% of screen positive children but only 15% of screen negative children were approached for interview. We additionally modified these weights to adjust for the small differences in second-phase participation rate between the screen-positive and screen-negative children, calculating separate weights for boys and girls. We did not use techniques such as multiple imputations to perform analyses taking into account baseline

characteristics of non-participants because non-participation was not predicted by any observed characteristic (see below). The p-values presented in these results are not adjusted for multiple testing by informant or subscale because almost all p-values are either non-significant (p>0.05) or highly significant (p<0.001) – that is, sufficiently large or sufficiently small to be outside the ambiguous range in which protected p-values are useful in guiding interpretation.

Results

SDQ scale and item scores

The parent SDQ was highly predictive of DAWBA diagnosis in both Norway and Britain and showed similar effect sizes, with an odds ratio (OR) for diagnosis per one-point increase on the 40-point SDQ scale of 1.30 (95% CI 1.27 – 1.32) in Britain and 1.25 (1.20 - 1.31) in Norway. For teachers the corresponding ORs were 1.22 (1.20-1.25) and 1.24 (1.19 – 1.29) respectively. Graphical examination of SDQ scores showed that the Norwegian and British samples had similarly shaped distributions but, as shown in Table 1, there were substantial Norwegian-British differences for mean scores on all SDQ scales. In all 14 'problem' scores (7 scores times two informants), the British mean was between 0.2 and 0.5 of a standard deviation higher than the Norwegian one.

Table 1. Mean SDQ scores and standard deviations in Norway and Britain, according to parents and teachers

SDQ score	Parent report			Teacher report		
	Britain Mean (SD) N=4898	Norway Mean (SD) N=6658	Difference in mean (Britain - Norway) (95%CI)	Britain Mean (SD) N=3836	Norway Mean (SD) N=6184	Difference in mean (Britain - Norway) (95%CI)
Total symptoms	8.37 (6.39)	5.87 (4.87)	2.50 (2.29,2.72)	6.57 (6.07)	4.12 (4.79)	2.46 (2.23,2.68)
Emotional	1.92 (2.09)	1.30 (1.71)	0.62 (0.55,0.69)	1.52 (1.97)	0.64 (1.32)	0.88 (0.81,0.96)
Conduct	1.53 (1.84)	0.96 (1.28)	0.58 (0.52,0.64)	0.90 (1.66)	0.61 (1.26)	0.29 (0.23,0.35)
Hyperactivity	3.47 (2.82)	2.65 (2.15)	0.82 (0.72,0.91)	2.85 (2.79)	2.12 (2.36)	0.73 (0.62,0.84)
Peer	1.45 (1.81)	0.96 (1.54)	0.49 (0.42,0.55)	1.30 (1.75)	0.74 (1.46)	0.55 (0.49,0.62)
Prosocial (higher more desirable)	8.81 (1.62)	8.52 (1.53)	0.30 (0.24,0.36)	7.59 (2.48)	8.44 (2.01)	-0.85 (-0.94,-0.76)
Impact	0.46 (1.33)	0.16 (0.94)	0.30 (0.25,0.34)	0.43 (0.94)	0.20 (0.83)	0.22 (0.18,0.26)
Burden	0.37 (0.78)	0.19 (0.55)	0.17 (0.15,0.20)	0.37 (0.071)	0.23 (0.60)	0.15 (0.12,0.18)

All Norway-Britain differences in means significant at p<0.001, adjusted for clustered sample design.

The differences between parent and teacher reports were explored further by looking individually at the 25 core SDQ items. In the original description of the SDQ (Goodman, 1997), 10 items are classified as strengths; 14 items as difficulties; and one item (getting along better with adults than with other children) is neutral. Norwegian teachers reported lower mean levels of all 14 difficulties than British teachers and higher mean levels of all 10 strengths (nearly all at p<0.001). Norwegian as opposed to British parents reported lower mean levels of all 14 difficulties (nearly all at p<0.001), but the picture for strengths was mixed: 2 were reported more commonly by Norwegian parents, 4 were reported less commonly, while for 4 strengths there was no difference.

DAWBA diagnoses

The prevalences of the main groups of DSM-IV diagnoses are shown in Table 2. Externalizing disorders (behavioural and hyperactivity disorders) were almost twice as common in Britain as in Norway, whereas the rates of emotional disorders were very similar in the two countries.

Table 2. Weighted prevalence of DSM-IV disorders in 8-10 year olds in Norway and Britain

	Britain (percent and 95% CI)	Norway (percent and 95% CI)	Difference (Britain - Norway), and 95%CI	p-value of difference
Emotional disorder (anxiety/ depression)	3.0 (2.5 - 3.5)	3.2 (2.0 - 4.4)	0.3 (-1.5 - 1.0)	0.699
Behavioural disorder (oppositional/conduct)	4.8 (4.2 - 5.4)	2.5 (1.9 - 3.0)	2.3 (1.5 – 3.2)	<0.001
ADHD[1]	2.5 (2.1 - 2.9)	1.3 (0.8 - 1.7)	1.2 (0.6 – 1.8)	<0.001
ASD[2]	0.7 (0.5 - 1.0)	0.4 (0.2 - 0.6)	0.3 (0.00 – 0.07)	0.048
Any disorder	8.7 (7.9 - 9.5)	6.1 (4.7 - 7.4)	2.6 (1.0 – 4.2)	0.001

[1] ADHD=Attention-Deficit/Hyperactivity Disorder; [2] ASD=Autistic Spectrum Disorder. All 4898 British children had detailed psychiatric diagnostic measures. Only 1024 of the 6658 Norwegian children had detailed psychiatric diagnostic measures – prevalences weighted back to the 6658 adjusting for selective sampling and participation. Confidence intervals and p-values adjust for clustered sample design in Britain and weight for selective sampling in Norway.

The diagnoses presented in Table 2 were made by clinicians. We repeated these analyses drawing on computer-generated predictions of diagnosis, which are based solely upon the structured section of the interview, and so do not require any rater input. We used a computer prediction based just on the parent DAWBA, because some children had parent data but not teacher data. This produced results very similar to those shown in Table 2. Emotional disorders were diagnosed by computer algorithm in 2.9% of the British sample and 3.0% of the Norwegian (p=0.93), behavioural disorders in 4.1% of the British sample and 2.9% of the Norwegian (p=0.02) and Attention-Deficit/Hyperactivity Disorder (ADHD) in 2.0% of the British and 1.3% of the Norwegian sample (p=0.01).

Questionnaire scores for diagnostic groups

Because some children have diagnoses from more than one diagnostic grouping, there are several possible ways of examining questionnaire scores by diagnostic grouping. In Table 3 we present the results in the most transparent way, including all the children with a relevant diagnosis in the corresponding row of the table, even though this means that some children contribute to more than one row. Here, and in all subsequent analyses, the pattern of results is closely similar if the analyses are repeated after excluding individuals with diagnoses from two or more diagnostic groups (i.e. including 'pure' disorder only), or after assigning such individuals to just one group according to a diagnostic hierarchy (tables available on request).

Table 3. Mean questionnaire scores for children from Norway and Britain by diagnostic group

Diagnosis	Mean parent-rated SDQ score						Mean teacher-rated SDQ score					
	Total symptoms		Impact		Burden		Total symptoms		Impact		Burden	
	Britain	Norway	Britain	Norway	Britain	Norway	Britain	Norway	Britain	Norway	Britain	Norway
None	7.49 (7.33 - 7.65)**	5.32 (5.05 - 5.59)	0.24 (0.22 - 0.27) **	0.07 (0.05 - 0.10)	0.25 (0.24 - 0.27) **	0.17 (0.14 - 0.19)	5.73 (5.56 - 5.90) **	3.47 (3.21 - 3.74)	0.28 (0.25 - 0.30) **	0.09 (0.07 - 0.11)	0.27 (0.25 - 0.29) **	0.15 (0.13 - 0.18)
Emot-ional[1]	18.05 (16.86 - 19.25)**	10.47 (8.18 - 12.77)	2.80 (2.32 - 3.28)**	1.04 (0.48 - 1.60)	1.53 (1.35 - 1.71)**	0.71 (0.43 - 1.00)	12.11 (10.66 - 13.55)**	6.93 (4.83 - 9.03)	1.38 (1.08 - 1.68)*	0.65 (0.27 - 1.02)	0.94 (0.76 - 1.12)*	0.52 (0.29 - 0.75)
Behav-ioural[1]	19.2 (18.3 - 20.1)	18.15 (16.38 - 19.92)	2.97 (2.60 - 3.35)	3.48 (2.71 - 4.26)	1.80 (1.66 - 1.93)	1.84 (1.63 - 2.06)	17.31 (16.29 - 18.33)	16.82 (15.03 - 18.62)	2.19 (1.94 - 2.43)**	3.24 (2.74 - 3.73)	1.68 (1.53 - 1.83)	1.98 (1.72 - 2.24)
ADHD[1],[2]	19.89 (18.71 - 21.06)	20.89 (19.15 - 22.62)	3.63 (3.12 - 4.14)	4.88 (3.78 - 5.97)	1.89 (1.71 - 2.07)	2.03 (1.73 - 2.32)	18.10 (16.81 - 19.39)	19.09 (17.27 - 20.92)	2.62 (2.28 - 2.96)**	3.89 (3.32 - 4.46)	1.85 (1.68 - 2.02)	2.05 (1.74 - 2.36)
ASD[1,3]	21.57 (19.24 - 23.90)	20.71 (17.19 - 24.23)	5.20 (4.07 - 6.33)*	7.75 (6.65 - 8.85)	2.17 (1.86 - 2.48)	2.63 (2.35 - 2.92)	15.92 (13.22 - 18.62)	16.41 (13.17 - 19.65)	2.73 (2.03 - 3.43)	3.59 (2.46 - 4.72)	1.69 (1.33 - 2.06)	1.39 (0.82 - 1.97)

[1] Some children in this group also have additional co-morbid diagnoses from other groups; [2] ADHD=Attention-Deficit/Hyperactivity Disorder; [3] ASD=Autistic Spectrum Disorder; Asterisked scores significantly higher/lower in Britain than Norway: *=p<0.01; **=p<0.001 (thresholds higher than the conventional p<0.05 used to reflect multiple testing).

As shown in Table 3, British and Norwegian children with comparable non-emotional disorders also had similar questionnaire scores. The few significant differences are, moreover, in the opposite direction to what would be predicted from underreporting by Norwegian parents and teachers. By contrast, there were much stronger and more consistent effects in the six comparisons (three scores times two informants) relating to emotional disorders. In all cases questionnaire scores were substantially lower in Norway, as expected if low rates of questionnaire-reported psychopathology in Nordic countries were an artefact due to underreporting.

Exploring gender, comorbidity or specific diagnosis as possible explanations of the difference

Among children with emotional disorders, the gender ratio and co-occurence of multiple comorbid emotional disorders was very similar between the Norwegian and British samples, and adjusting for these two factors in linear regression models left the relative difference between Norwegian and British informants virtually unchanged.

There was, however, evidence of an important difference between the two countries in the prevalence of other comorbid disorders (behavioural disorders, hyperactivity or autism) among children with emotional disorders. Such comorbidity occurred in 34% and 16% of those with emotional disorders in Britain and Norway respectively (p=0.003). There were also differences in the prevalence of some emotional disorders between the two countries, with specific phobias more common in Norway and generalised anxiety and anxiety-NOS ('not otherwise specified') more common in Britain. We therefore repeated the analyses reported in Table 3 a) after excluding children with comorbid

disorders and b) after excluding children with comorbid non-emotional disorders, specific phobias, generalised anxiety disorders and anxiety disorder-NOS. In both cases the mean questionnaire scores for symptoms and impact remained substantially lower among Norwegian children according to both parents and teachers (details available on request).

Recognition of problems

Having compared symptom, impact and burden scores between British and Norwegian children with mental health diagnoses, we then explored rates of problem recognition. After completing the symptom section of the SDQ, both parents and teachers were asked whether they thought that the child had a problem with "emotions, concentration, behaviour or being able to get along with other people". We then compared the frequency with which Norwegian and British children with DAWBA diagnoses were identified as having 'definite' or 'severe' difficulties by their parents and teachers.

Figure 1. Proportion of Norwegian and British children with different groups of disorders who are reported to have definite or severe problems by parents and teachers.
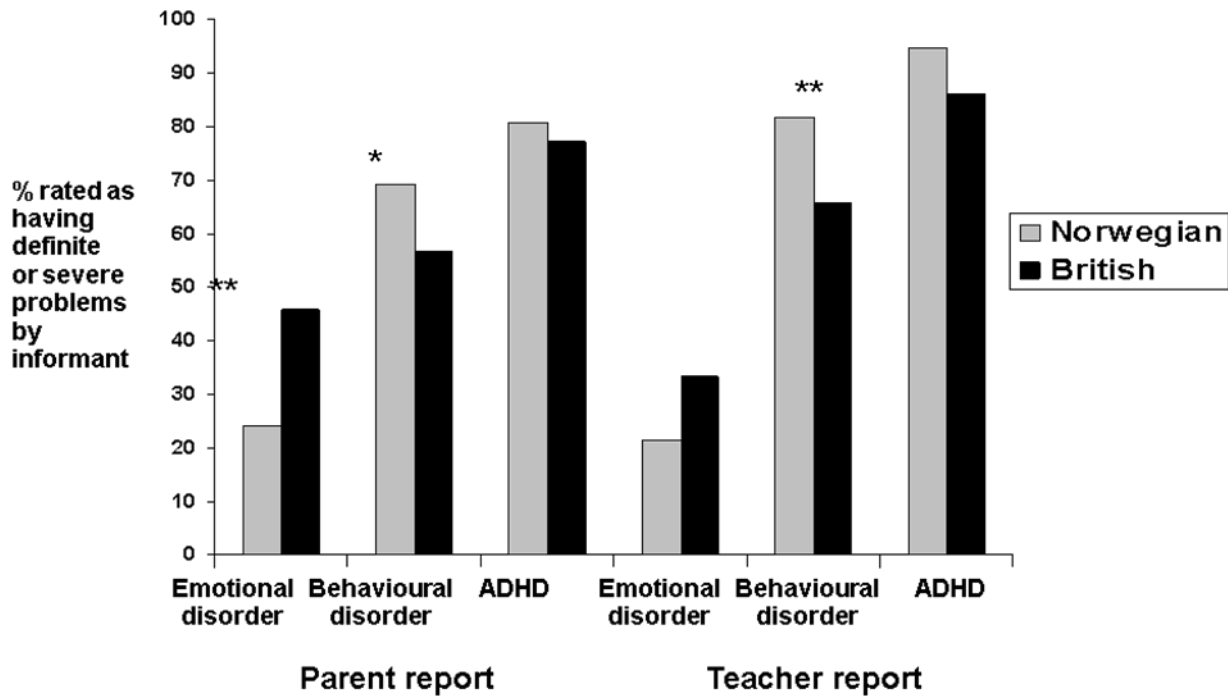
The results, displayed in Figure 1, indicate that for children with emotional disorders a lower proportion of Norwegian parents recognised definite/severe problems than did British parents (24.1% vs. 45.9%, p=0.002), and a similar trend was also observed for teachers (21.5% vs. 33.3%, p=0.09). This pattern persisted even when considering only 'pure' emotional disorders, which were recognised by 16.0% of Norwegian parents as

compared to 35.1% of British parents (p=0.01), and by 11.2% of Norwegian teachers as compared to 19.4% of British teachers (p=0.18).

By contrast, rates of problem recognition were at least as high, if not higher, for externalizing problems in Norwegian as compared to British informants. For behavioural disorders, 69.1% of Norwegian and 56.6% of British parents reported definite/severe problems (p=0.05), with the corresponding figures for teachers being 81.1% vs. 65.8% (p=0.007). For ADHD, 80.7% of Norwegian and 77.2% of British parents reported definite/severe problems (p=0.65), with the corresponding figures for teachers being 94.6% vs. 86.0% (p=0.08).

Discussion

Epidemiology often involves the search for reasons behind observed differences across space, time or culture in rates of disease. When objective and comparable measures are lacking, as in most of psychiatric epidemiology, we first need to take a step back and establish whether apparent differences could simply reflect reporting biases. In this paper, we illustrate how questionnaires and interviews can be used for this purpose, investigating the specific case of the apparent Nordic advantage in child mental health.

We compared representative, population-based surveys of 8-10 year olds in Norway and Britain which assessed mental health using identical measures. Our findings on parent and teacher questionnaire reports of psychological difficulties were very clear and in line with the previous literature. Problem levels were consistently and substantially lower in Norwegian than in British children, whether judged by individual items or composite scores, and whether judged by parent or teacher report.

We then compared the prevalence of psychiatric diagnoses in Norway and Britain, as judged using a detailed investigator-based measure of psychiatric disorder. The overall rate of psychiatric disorder is almost a third lower in the Norwegian sample, reflecting lower rates of behavioral disorders and ADHD. In marked contrast to our questionnaire findings, however, Britain and Norway showed no difference in rates of emotional disorders. Our confidence in these findings is supported by the several tactics we sought to minimise the potential for cross-national diagnostic bias. Chief among these tactics were: identical diagnostic measures; the availability of open-ended descriptions of problems in making diagnoses; a shared training programme and shared supervision for Norwegian and British raters; and a closely similar pattern of results when we used computer-generated diagnoses based only upon information in the fully structured sections of the interview.

On this basis, we used the presence or absence of psychiatric disorders as an external standard for judging cross-national biases in questionnaire scores. For non-emotional disorders, we found no convincing evidence for any such biases. Children with equivalent non-emotional disorders in Norway and Britain generally had equivalent questionnaire scores: parents and teachers in both countries described similar levels of symptoms, impact and burden. These findings therefore do not support the notion that

Norwegian parents and teachers are less sensitive than British parents and teachers to the manifestations of child mental health problems in general. Neither do these findings suggest that Norwegian adults are simply more polite about troubled children than their British counterparts – they were just as likely, for instance, to describe children with externalizing disorders as being a "burden" to them at home or at school.

This makes it all the more striking to find such marked national differences in the questionnaire scores assigned by parents and teachers to British and Norwegian children with comparable emotional disorders. It is not that the Norwegian parents and teachers could not describe the cardinal symptoms of emotional disorder: their closed and open-ended answers to the detailed interview measure were the basis for making the diagnosis of an emotional disorder in the first place. Nevertheless, on the brief screening questionnaires, both parents and teachers in Norway rated substantially fewer symptoms and less impact and burden than their British counterparts when describing apparently similar children.

A Norwegian "blind spot" for emotional difficulties is also evident from Figure 1, which displays how frequently Norwegian and British parents recognised a mental health problem in children with different categories of DAWBA diagnosis. As Figure 1 indicates, all mental health problems suffer from some degree of underrecognition, but this is particularly the case for emotional disorders and particularly in Norway as compared to Britain. By contrast, Norwegian parents and teachers recognise externalizing problems at least as frequently as British parents and teachers, again demonstrating that Norwegian adults are not simply less sensitive or more polite about troubled children generally. It is the relative specificity of this apparent blind spot, combined with the particularly low rates of service use for emotional disorders (see below), that has led us to refer to 'Norwegian relative underreporting' rather than 'British relative over-reporting' throughout this paper.

One interpretation of this under-reporting is that Norwegian adults in this sample take a more "normalizing" view of emotional (but not externalizing) symptoms when filling out screening questionnaires. A non-pathologising mindset towards worries and misery in childhood may often be advantageous, but could lead to under-referral to mental health services of children with severe, persistent and impairing anxiety or depression. This possibility is supported by previous findings relating to service use in the British and Norwegian surveys analysed in this paper. The lifetime rate of contact with specialist child and adolescent mental health services in the Norwegian sample was 13% for those with a pure emotional disorder, 41% for those with behavioral disorders and 75% of those with hyperactivity disorders (Heiervang et al., 2007). Green et al. (2004) reported that in the British sample contact with specialist mental health services in the previous year was 24% for those with emotional disorders, 28% for those with behavioural disorders, and 52% for those with hyperactivity disorders. While far from conclusive, this suggests the same pattern as that seen in Figure 1, namely that emotional disorders are relatively neglected in both countries, but particularly so in Norway.

Turning to non-emotional disorders, the rest of the world potentially has much to learn from Norway about promoting child mental health. Why does Norway have what seems to be a real advantage in terms of lower rates of externalizing disorders? Knowing this could inform preventative programs elsewhere in the world. There are many potentially plausible explanations including child-level factors such as a diet richer in omega-3 fatty acids, family-level factors such as higher maternal education, or societal differences such as low levels of income inequality. In this study we focus simply upon *describing* the mental health of children in Norway and Britain, and upon investigating the single specific potential explanation of differential reporting bias between the two countries. We hope, however, that the evidence we present here will stimulate further research into which particular aspects of Norwegian life protect children against behaviour and hyperactive problems.

The present study was limited by being unable to examine self-reported psychopathology, because the SDQ is only designed to be administered to children aged 11 or older. It also only includes children aged eight to ten years, making it possible that the results will not generalize to other age groups. Moreover, whereas the British sample was nationally representative, the Norwegian sample was drawn from just one city and its surrounding area. However, a comparison of our SDQ data with that from other regions of Norway and indeed other Nordic countries suggests similar levels of emotional and behavioral problems reported by teachers and parents (Obel et al., 2004; Heiervang et al., 2007). This, and the relative cultural, economic and political uniformity within and between the Nordic countries, suggests the findings may generalise to Nordic countries as a whole. Seeing if these findings replicate in comparisons with non-Nordic countries other than Britain will be important.

The methods we use in this paper also have limitations. Attempting to investigate reporting bias using measures administered in different languages potentially confounds differences in the measure (as a result of imperfect linguistic equivalence) with differences in respondents. Although the translation of both the SDQ and the DAWBA included careful back translations, we have previously shown that seemingly minor wording differences may have large impact on scores (Goodman, Iervolino, Collishaw, Pickles, & Maughan, 2007). It also needs to be remembered that even our more objective interviewer-based mental health measure ultimately relies on respondents' ability to recognise and describe symptoms so that clinical raters can make comparable ratings across linguistic and cultural boundaries.

Finally, while in this paper we have described a cross-national reporting bias, we have not provided an explanation for that bias. To do so would require an analysis of possible confounding and/or mediating factors which lead to parents of Norwegian nationality underreporting emotional symptoms, on average, relative to their British counterparts. Unfortunately this is not a question which we have been able to address in this paper, as we lack comparable measures for almost all plausible social and cultural factors. Whatever the explanation, however, the mere fact that there is an association between nationality and reporting patterns demonstrates the potential dangers of making cross-cultural comparisons without explicit evaluation of possible cross-national reporting bias.

Conclusion

Our comparison of Norway and Britain illustrates an inconsistent relationship between questionnaire and interview data, indicating why cross-cultural differences on questionnaires should not automatically be interpreted as reflecting true differences in psychopathology. Our exploration of the causes of these inconsistencies suggests that the Nordic advantage in child mental health may be both a fact and an artefact. The apparent 'fact' of a genuine health advantage for Norwegian children in terms of behavioral and attentional problems should stimulate research to identify which aspects of Norwegian life promote this good outcome. The apparent 'artefact' that Norwegian adults both underreport and under-recognize children's emotional symptoms when completing screening questionnaires points to the need to develop better approaches to screening so that fewer Norwegian children with emotional disorders go untreated.

Conflict of interest

EH is the director and owner of Careahead, which provides teaching and supervision to clinics on the use of the SDQ and the DAWBA. AG and RG are directors and part owners of Youthinmind, which provides no-cost and low-cost software and web sites related to the SDQ and the DAWBA.

Acknowledgements

References

Achenbach, T.M., Bird, H.R., Canino, G.J., Phares, V., Gould, M., Rubio-Stipec, M. (1990). Epidemiological comparisons of Puerto Rican and U.S. mainland children: Parent, teacher, and self reports. *Journal of the American Academy of Child and Adolescent Psychiatry*, 2, 84-93.

Becker, A., Steinhausen, H.C., Baldursson, G., Dalsgaard, S., Lorenzo, M.J., Ralston, S.J., Dopfner, M., Rothenberger, A.; ADORE Study Group. (2006). Psychopathological screening of children with ADHD: Strengths and Difficulties Questionnaire in a pan-European study. *European Journal of Child and Adolescent Psychiatry*, 15(Suppl), S56-S62.

Bilenberg, N. (1999). The Child Behavior Checklist (CBCL) and related material: standardization and validation in Danish population based and clinically based samples. *Acta Psychiatrica Scandinavica* 398(Suppl), S2-S52.

Broberg, A.G., Ekeroth, K., Gustafsson, P.A., Hansson, K., Hagglof, B., Ivarsson, T., & Larsson, B. (2001). Self-reported competencies and problems among Swedish adolescents: a normative study of the YSR. Youth Self Report. *European Journal of Child and Adolescent Psychiatry*, 10, 186-193.

Crijnen, A.A.M., Achenbach, T.M., & Verhulst, F.C. (1997). Comparisons of problems reported by parents of children in 12 cultures: Total Problems, Externalizing, and Internalizing. *Journal of the American Academy of Child and Adolescent Psychiatry*, 6, 1269-1277.

Ford, T., Goodman, R., & Meltzer, H. (2003). The British Child and Adolescent Mental Health Survey 1999: the prevalence of DSM-IV disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 42, 1203-1211.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry.* 38, 581-586.

Goodman, R. (1999). The extended version of the Strengths and Difficulties Questionnaire (SDQ) as a guide to child psychiatric caseness and consequent burden. *Journal of Child Psychology and Psychiatry*, 40, 791-801.

Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, 1337-1345.

Goodman, R., Ford, T., Richards, H., Gatward, R., & Meltzer, H. (2000b). The Development and Well-Being Assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *Journal of Child Psychology and Psychiatry*, 41, 645-655.

Goodman, R., Iervolino, A.C., Collishaw, S., Pickles, A., & Maughan, B. (2007). Seemingly minor changes to a questionnaire can make a big difference to mean scores: a cautionary tale. *Social Psychiatry and Psychiatric Epidemiology*, 42, 322-327.

Goodman, R., Renfrew, D., & Mullick, M. (2000a). Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *European Journal of Child and Adolescent Psychiatry*, 9, 129-134.

Green, H., McGinnity, A., Meltzer, H., Ford, T., & Goodman, R. (2004). *Mental Health of Children and Young People in Great Britain.* Basingstoke: Palgrave Macmillan.

Heiervang, E., Stormark, K.M., Lundervold, A.J., Heimann, M., Goodman, R., Posserud, M.B., Ullebø, A.K., Plessen, K.J., Bjelland, I., Lie, S.A., & Gillberg, C. (2007). Psychiatric disorders in Norwegian 8- to 10-year-olds: an epidemiological survey of prevalence, risk factors, and service use. *Journal of the American Academy of Child and Adolescent Psychiatry*, 46, 438-447.

Hysing, M., Elgen, I., Gillberg, C., Lie, S.A., & Lundervold, A.J. (2007). Chronic physical illness and mental health in children. Results from a large-scale population study. *Journal of Child Psychology and Psychiatry*, 48:785-92.

 Koskelainen, M., Sourander, A., & Kaljonen, A. (2000). The Strengths and Difficulties Questionnaire among Finnish school-aged children and adolescents. *European Journal of Child and Adolescent Psychiatry*, 9, 277-284.

Larsson, B., & Frisk, M. (1999). Social competence and emotional/behaviour problems in 6-16 year-old Swedish school children. *European Journal of Child and Adolescent Psychiatry*, 8, 24-33.

Meltzer, H., Gatward, R., Goodman, R., & Ford T. (2000). *Mental Health of Children and Adolescents in Great Britain.* London: The Stationery Office.

Meltzer, H., Gatward, R, Corbin, T., Goodman, R., & Ford, T. (2003). Persistence, onset, risk factors and outcomes of childhood mental disorders. London, The Stationery Office.

Nøvik, T.S. (1999). Validity of the Child Behaviour Checklist in a Norwegian sample. *European Journal of Child and Adolescent Psychiatry*, 8, 247-254.

Obel, C., Heiervang, E., Rodriguez, A., Heyerdahl, S., Smedje, H., Sourander, A., Guethmundsson, O.O., Clench-Aas, J., Christensen, E., Heian, F., Mathiesen, K.S., Magnusson, P., Njarethvik, U., Koskelainen, M., Ronning, J.A., Stormark, K.M., & Olsen, J. (2004). The Strengths and Difficulties Questionnaire in the Nordic countries. *European Journal of Child and Adolescent Psychiatry*, 13(Suppl), S32-S39.

Ronning, J.A., Handegaard, B.H., Sourander, A., & Morch, W.T. (2004). The Strengths and Difficulties Self-Report Questionnaire as a screening instrument in Norwegian community samples. *European Journal of Child and Adolescent Psychiatry*, 13:73-82. Weine, A.M., Phillips, J.S., & Achenbach, T.M. (1995). Behavioral and emotional problems among Chinese and American children: parent and teacher reports for ages 6 to 13. *European Journal of Child and Adolescent Psychiatry*, 4:237-248.