ELSEVIER

Social Science & Medicine 62 (2006) 815–827

SOCIAL SCIENCE & MEDICINE

www.elsevier.com/locate/socscimed

Cross-national variation of gender differences in adolescent subjective health in Europe and North America

Torbjørn Torsheim[a,*], Ulrike Ravens-Sieberer[b], Jorn Hetland[a], Raili Välimaa[c], Mia Danielson[d], Mary Overpeck[e]

[a] Research Centre for Health Promotion, University of Bergen, Christiesgt. 13, N-5015 Bergen, Norway
[b] Robert Koch Institute,Child and Adolescent Health,Stresemannstr 90, D 10963 BERLIN, Germany
[c] University of Jyväskylä, Department of Health Sciences, PO Box 35,FIN 40351 Jyväskylä, Finland
[d] National Institute of Public Health,103 52 Stockholm, Sweden
[e] Health Resources and Services Administration, Maternal and Child Health Bureau, 5600 Fishers Lane, Room 18-41, Rockville, MD 20857, USA

Available online 11 August 2005

Abstract

The cross-national consistency and variation of gender differences in subjective health complaints was examined in a sample of 125732 11- to 15-year-olds from 29 European and North American countries, participating in the WHO collaborative study 'Health behaviour in school-aged children (HBSC) 1997/98'. Health complaints were measured with the Health Behaviour in School-aged Children Symptom Checklist. Gender differences in health complaints were analysed through multilevel logistic regression analysis. The results indicated a very robust pattern of increasing gender differences across age, with 15-year-old girls as a group at increased risk for health complaints across all countries. The magnitude of gender differences varied across countries, with some countries showing a consistently strong gender difference across age group and different health complaints, and other countries showing a consistently weak gender difference. The gender difference in health complaints was stronger in countries with a low gender development index score. The findings underscore the need to incorporate socio-contextual factors in the study of gender health inequalities during adolescence.

Keywords: Somatic-complaints; Adolescence; Gender; Gender development index; Gender empowerment measure

Introduction

One of the most well-established findings in health-related research is that women tend to report higher levels of health complaints than men do (Eriksen, Svendsrod, Ursin, & Ursin, 1998; Van-Wijk & Kolk, 1997; Verbrugge, 1985), but the magnitude of differences varies considerably across study populations (Lahelma, Martikainen, Rahkonen, & Silventoinen, 1999; MacIntyre, Hunt, & Sweeting, 1996; Sweeting, 1995; Sweeting & West, 2003). Comprehensive literature reviews have made a strong case for the notion that gender differences in health complaints can be observed for some groups,

*Corresponding author. Tel.: +4755583301; fax: +4755589887.
E-mail address: torbjoern.torsheim@psyhp.uib.no (T. Torsheim).

0277-9536/$ - see front matter
doi:10.1016/j.socscimed.2005.06.047

but not in other groups (MacIntyre et al., 1996; Sweeting, 1995). The observed variability of gender differences has stimulated a renewed interest in factors that moderate such differences. Rather than viewing gender differences as fixed, the question becomes: under what conditions are gender differences strong, and under what conditions are gender differences weak? In this paper we highlight conditions at the societal level as a strong source of variability in gender differences.

Attempts to explain gender differences have included factors at various levels of explanation, including genetic factors (Silberg et al., 1999), perceptual factors (Van-Wijk & Kolk, 1997), as well as structural factors related to the distribution of work and political power (Kawachi, Kennedy, Gupta, & Prothrow-Stith, 1999), and gender role socialisation. These factors offer scope for influences at several levels, including the societal level.

Although the notion of a 'society and health' perspective on gender differences is not new, relatively little is known about the degree to which national social, and cultural characteristics influence the magnitude of gender health inequalities. Work by the United Nations Development Program (UNDP, 1995) has revealed profound national differences in women's achievements and life opportunities as compared to men. Such differences indicate that the national level is a potentially important level for examining social moderating influences on gender differences in health.

The few studies to address cross-national variation in gender differences have revealed a complex pattern. In a large multisite cross-national sample of primary care attendees (Piccinelli & Simon, 1997), women reported a statistically significant higher level of somatic complaints than men in 11 of the 15 centres. However, in two of the remaining four centres, men showed the highest mean score, although not significantly higher than women.

Cross-national variability of gender differences has also been suggested in adolescent community samples. A cross-cultural study (Crijnen, Achenbach, & Verhulst, 1999) of the Child Behaviour Checklist Syndromes (Achenbach & McConaughy, 1987) involving adolescents from 12 nations, reported that gender differences in somatic complaints varied as a function of culture and age, with increasing cross-national variability in gender across age. For US adolescents there was virtually no gender difference. Similarly, a gender by country interaction was reported in recent study of adolescents from four European countries (Haugland, Wold, Stevenson, Aaroe, & Woynarowska, 2001), with the strongest gender difference shown by Polish adolescents, and the weakest gender difference shown by Scottish adolescents.

Previous reports of cross-national variation in gender differences are open to several interpretations. From a "society and health" perspective, the cross-national variation could reflect a moderating influence from national differences in the gender distribution of power and life opportunities. As a general hypothesis one would expect the gender differences to be smaller in countries with equal life opportunities for men and women compared to countries with unequal opportunities. From recent work it is well established that modern industrialised countries differ considerably in their distribution of power and life opportunities between men and women. However, none of the above-cited studies addressed the influences of social factors.

A second interpretation to be considered is that previous reports of cross-national variation reflect a so called 'compositional' effect (e.g. Diez-Roux, 2002). From pure chance one would expect gender differences to vary across samples. Such compositional effects could erroneously lead to conclusions about contextual variation in gender differences, when in fact the differences are due to individual level factors in several life domains, such as body image, school pressures, material living conditions, and social network variables. Previous studies have not been able to rule out such factors as likely explanations for the variability of gender differences.

The present study examined compositional and contextual sources of cross-national variation of gender differences in health complaints in young people. The study was undertaken as part of the collaborative study Health Behaviour in School-Aged Children (HBSC) 1997/98. In the present study our investigation was structured according to three main objectives: the first part address the amount of gender differences across countries and age groups. In the second part of the paper we examine compositional effects of individual level factors on cross-national variation in gender differences. In the last part of the article we present relationships between macro-level factors and gender differences.

Method

Sample

Data were obtained from the large cross-national WHO collaborative study 'Health Behaviour in School-aged Children 1997/98' (Currie, 1998). The target populations for the HBSC study are 11-, 13-, and 15-year-old students. In the HBSC 1997/98 study, a total of 125 732 students from 29 countries and regions participated (see Table 1). As shown in Table 1, the sample consisted of adolescents from a culturally and economically diverse collection of countries.

Table 1
Sample details for the 1997/98 HBSC survey

Survey	Principal investigator	*N*
Austria	Wolfgang Dur	4316
Belgium-Flemish	Lea Maes	4824
Belgium-French	Danielle Piette	2505
Canada	Alan King	6567
Czech Republic	Ladislav Czemy	3703
Denmark	Pernille Due	5066
England	Mary Hickman	6373
Estonia	Mai Maser	1897
Finland	Jorma Tynjäla	4864
France[a]	Cristiane Dressen	4133
Germany[a]	Klaus Hurrelmann	4792
Greece	Anna Kokkevi	4299
Greenland	Michael Pedersen	1648
Hungary	Anna Azmann	3609
Israel	Yossi Harel	5054
Latvia	Ieva Ranka	3775
Lithuania	Apolinaras Zaborskis	4513
Northern Ireland	Grace McGuiness	3346
Norway	Bente Wold	5026
Poland	Barbara Woynarowska	4861
Portugal	Margarida Gaspar de Matos	3721
Ireland	Saoirse Nic Gabhainn	4394
Russia[a]	Aleksander Komkov	3997
Scotland	Candace Currie	5632
Slovak Republic	Miro Bronis	3789
Sweden	Ulla Marklund	3802
Switzerland	Beatrice Janin Jaquat	5520
USA	Mary Overpeck	5169
Wales	Chris Roberts	4537
Total		125 732

[a]Regional samples.

The sample was obtained through a complex multi-stage sampling procedure. The primary sampling unit was school-class, with self-selection of students. However, due to differences in the school-systems across countries, national adaptations had to be made. For the majority of countries the desired age group coincided with school entry, resulting in a homogeneous age composition of school classes. The strategy followed in these countries was to select one school-class per school within each age group to be covered. In a small number of samples (Flemish-speaking Belgium, French-Speaking Belgium, England, Republic of Ireland, Switzerland, and the USA), the age composition of school classes were more heterogeneous, due to students repeating grades, and different age of school entry. In this group of countries, a larger sample of school classes had to be sampled, and students matching the desired age-range within these classes would be selected. More detailed information about the sample and the sampling frame can be obtained elsewhere (Currie, Hurrelmann, Settertobulte, Smith, & Todd, 2000; NSD, 2000).

Given the multistage sampling procedure, nonresponse may occur at several levels, including school, school classes, and student. The available documentation (NSD, 2000) provide detailed information on nonresponse at the level of school and student. The response rate (RR) at the level of school was in general high, with a majority of countries above 80%.

Procedure

Data were collected in accordance with a standardised protocol (Currie, 1998). Teachers received instructions on how to administer the survey. Questionnaires were distributed during ordinary class-hours. Pupils were informed that participation was voluntary, and that responses were treated as anonymous. Each student had 45 min to complete the survey.

Measurement

Individual level factors. Subjective health complaints were measured through the HBSC symptom checklist (Haugland & Wold, 2001). This scale includes a list of eight common health complaints (headache, stomachache, backache, depressed mood, irritable, nervousness, sleeping difficulties, dizziness). Students are asked: 'In the last 6 months: how often have you had the following?' Each health complaint is rated on a five-point frequency scale: 'about every day'; 'more than once a week'; 'about every week'; 'about every month'; 'rarely or never'. A qualitative validation study of the scale suggested that each complaint were interpreted as negative states with consequences for daily life and well-being (Haugland & Wold, 2001). In quantitative analysis the HBSC symptom checklist has revealed a satisfactory reliability (Haugland & Wold, 2001), with test–retest reliabilities ranging from 0.70 to 0.80.

Alcohol use was measured by the item: 'At present, how often do you drink anything alcoholic such as beer, wine or spirits?' For each of the exemplars the response options were: 'Every day' (1), 'Every week' (2), 'Every month' (3), 'Rarely' (4), 'Never' (5). Smoking was measured by the item: 'How often do you smoke at present?' 'Every day' (1); 'At least once a week, but not every day' (2), 'Less than once a week' (3), 'I do not smoke' (4). Material living conditions was measured through the family affluence scale (Currie et al., 2000). The family affluence scale is a linear composite of three indictors, the first being: 'Does your family have a car or a van?'. The response options were: 'No' (0), 'Yes' (1), 'Yes, two or more' (2). The second indicator was: 'Do you have your own bedroom?' The response options were: 'Yes' (1), 'No' (0). The third indicator was: 'During the past year, how many times did you travel away on holiday (vacation) with your family?' The response options were: 'Not at all' (0), 'Once' (1), 'Twice' (2), and 'More than twice' (3). When the three indicators are combined to produce a linear composite score, the family affluence scale ranges from 0 (lowest affluence) to 6 (highest affluence). Social support was measured with six items from the teacher and classmate support scale (Torsheim, Wold, & Samdal, 2000). Body satisfaction was measured through the single item: 'Is there anything about your body you would like to change?' Response options were 'Yes ' (1); 'No ' (2).

Macro-level factors. National figures on all macro-level indicators factors were obtained from the UNDP database (UNDP, 1999). The gender empowerment measure (GEM) focuses on gender differences in life opportunities in three areas: economic participation, political participation, and power over economic resources. The GEM is computed through a linear combination of women's and men's percentage shares of parliamentary seats, women's and men's percentage shares of positions as legislators, senior officials and managers and women's and men's percentage shares of professional and technical positions, and share of economic resources, as measured by women's and men's estimated earned income (PPP US$). The gender development index (GDI) focuses on achievements in areas of education, health, and income. The computation of the GDI is based on female life expectancy at birth (years), male life expectancy at birth (years), combined gross enrolment ratio (%), adult literacy rate, and estimated earned income.

To examine some of the specific subcomponents in the GEM and GDI we also included two other macro indicators. The female/male life expectancy ratio was calculated by taking the ratio of female and male life expectancy at birth. Male and female life expectancy at birth for 1997 was obtained from the UNDP Database. To include a general measure of wealth, gross domestic product (GDP) at purchase power parities for the year 1997 was included.

Statistical analysis

Differential item functioning (DIF) was assessed with SPSS 11.0 through a method developed by Zumbo (1999). Using this method, uniform and nonuniform DIF across multiple groups can be tested in a three-block logistic regression model. In the first block, the relevant item was regressed on the total symptom score. In the second block, main effects of countries were entered, whereas in the third block, interaction terms of country by total symptom score were entered. By convention, DIF is indicated when the two following criteria are met: (a) blocks 2 and 3 are statistically significant, and (b) the added R^2 for blocks 2 and 3 is higher than .12 (For further details, see Zumbo (1999)).

A multilevel modelling approach was followed in the data analysis. Multilevel modelling originates from educational research (Bryk & Raudenbush, 1992; Goldstein, 1995), but has become a valuable approach also in public health research (Diez-Roux, 2000; Leyland & Goldstein, 2001). By performing a multilevel analysis, the clustered structure of the data is taken into account, and accurate estimates of individual standard errors are obtained. The multistage sampling indicates that a hierarchical three-level model including separate variances for students, schools, and countries may adequately reflect the data structure. However, for five of the countries, proper identification for school memberships was not available. To carry out a three-level model, the samples without school identification codes would have to be excluded from the analysis. Rather than excluding valuable data from these countries, a decision was made to model a two-level variance structure, including students within countries. Not modelling a random school-level intercept variance introduces a potential bias. However, previous research suggests that this bias is relatively modest, given that health complaints differ little across schools (Torsheim & Wold, 2001).

The cross-national variation of gender differences was modelled in a series of logistic random intercept and coefficient models. The *random coefficient* model enables for simultaneous estimation of the fixed average effect of gender and the cross-national variation around the fixed effect. The fixed effect reflects the overall grand mean effect of gender across countries. Cross-national variation across the grand mean effect of gender is modelled as a separate variance component. The random cross-national variance component is a population estimate, to be generalised to the population of countries that are sampled for the study. Under the assumptions of normal-distributed residuals at the higher level, the random coefficient variance has a standard error and can be subject to statistical tests of significance. Analysis was carried out for each specific health complaints. In addition, to assess gender differences in overall symptom level, 'Any recurrent health complaints' was defined as the group of adolescents reporting at least one recurrent health complaint. The statistical software MLwiN 1.10 (Rasbash et al., 2000) was used for the estimation of multilevel models. Penalised second-order quasi-likelihood estimation was used as this method provides more accurate estimates when the higher level units (i.e. countries) are few relative to the lower level units (Goldstein & Rasbash, 1996), as compared to marginalised first-order estimation. Depending on the research question, analyses were performed pooled for the total sample, pooled within age groups, and pooled within gender.

The results to be presented are nonweighted. The decision to not weight the data was based on the relatively homogenous sample size per country unit. In line with the suggestions of other authors (Frohlich, Carriere, Potvin, & Black, 2001) we did not weight the samples according to population size, as the population sizes of the sampled countries are highly heterogeneous. For example, in the US the eligible population surpassed 9 million students. In contrast, the comparable population in Norway was 150 000 students. Weighting by country population would thus allow very little influence to data from countries with small populations.

Results

Differential item functioning

Table 2 shows the results of a test for DIF by country, using a hierarchical blockwise logistic regression analysis with single items as the dependent variables. The first block included the main effect of total scale score (trait). In block 2, the main effect of country membership was entered. For each item the entry of country membership in block 2 accounted for items score beyond the scale score. This means that young people with the same scale score would respond somewhat differently depending on their country membership. However, the change in R^2 when entering country member ship was clearly smaller than the critical criterion suggested by Zumbo (1999), and also smaller than the slightly more conservative criterion recommended by Jodoin and Gierl (2001). This indicates that the level of uniform DIF was small across all items. In block 3, the score by country interaction term accounted for a very small proportion of variance, indicating that the impact of country membership did not differ across scale scores. Overall, the results of the analysis indicate that DIF is not a threat to the validity of findings in the present study.

Health complaints by age and gender

Point estimates for each health complaints with 95% confidence intervals (CI) for the point estimates are shown in Table 3. It can be seen from Table 3 that a sizeable percentage of boys and girls reported recurrent health complaints several times weekly. Irrespective of gender, irritability was the most common recurrent complaint, but also recurrent episodes of feeling low, headache, and sleeping difficulties were reported by a high percentage of adolescents. In all age groups, and across gender, recurrent back pain and dizziness was reported by a relatively small group of adolescents.

Cross-national variation of gender differences

Table 4 shows the results of a series of random coefficient models, with estimates of fixed gender differences and cross-national variation around the fixed effect. For illustrative purposes, the regression coefficients of the logistic model were transformed to odds ratios (OR) of recurrent subjective health complaints, with boys serving as the reference group. For each health complaint, and within all age groups, girls had statistically significant higher odds. For all complaints, the gender differences grew larger across age groups. In the group of 11-year-olds, the OR for gender ranged between 1.11 (sleeping difficulties) to 1.86 (abdominal pain), whereas for the 15-year-olds, the OR ranged between 1.34 (Back pain) to 2.90 (Headache).

The cross-national variation in gender differences became more pronounced with increasing age. It can be seen from Table 4 that among the 11-year-olds, the gender differences showed statistically significant cross-national variance in only one out of eight complaints. For 13-year-olds the corresponding number was four

Table 2
Test of differential item functioning by country for the HBSC symptom checklist

Health complaint	Block1		Block 2		Block 3	
	HBSC scl		Uniform dif		Nonuniform dif	
	R^2	χ^2	ΔR^2	$\Delta\chi^2$	ΔR^2	$\Delta\chi^2$
Headache	0.438	72765.90	0.007	709.29	0.001	76.27
Abdominal pain	0.415	51535.02	0.006	494.44	0.002	113.97
Back pain	0.333	59627.87	0.021	1498.64	0.001	82.90
Feeling low	0.512	57221.80	0.017	1594.06	0.003	247.60
Irritable	0.527	80924.98	0.011	1482.18	0.004	528.23
Nervous	0.487	76961.90	0.034	4030.99	0.008	875.03
Sleeping difficulties	0.391	83230.79	0.029	3042.59	0.003	348.06
Dizziness	0.423	49966.31	0.021	1519.61	0.002	105.55

Table 3
Subjective health complaints several times a week, by age and gender

	11-year-olds		13-year-olds		15- year-olds	
	Point estimate	95% CI	Point estimate	95% CI	Point estimate	95% CI
Headache						
Girls	18.0	(16.0–20.1)	19.3	(17.4–21.5)	23.6	(21.3–26.1)
Boys	12.1	(10.6–13.8)	10.7	(9.4–12.1)	9.6	(8.3–11.2)
Abdominal pains						
Girls	13.5	(11.7–15.4)	11.6	(10.3–13.0)	11.1	(9.6–12.8)
Boys	7.8	(6.7–9.2)	5.5	(4.5–6.6)	4.4	(3.5–5.5)
Back pain						
Girls	8.8	(7.5–10.3)	11.2	(9.8–12.8)	14.1	(12.6–15.8)
Boys	7.1	(6.1–8.3)	8.8	(7.8–9.9)	10.9	(9.8–12.1)
Feeling low						
Girls	13.0	(11.3–15.0)	16.8	(14.6–19.2)	20.4	(17.7–23.5)
Boys	10.0	(8.5–11.7)	9.1	(7.5–10.9)	9.3	(7.5–11.5)
Irritable or bad mood						
Girls	22.7	(20.1–25.6)	27.2	(24.5–30.2)	29.6	(26.6–32.8)
Boys	19.6	(17.4–22.0)	21.0	(18.4–23.8)	21.3	(18.8–24.0)
Nervous						
Girls	19.8	(17.2–22.6)	24.0	(20.5–27.8)	26.5	(22.1–31.4)
Boys	16.9	(14.6–19.4)	16.7	(14.1–19.6)	18.1	(15.2–21.4)
Sleeping difficulties						
Girls	20.2	(18.0–22.6)	19.9	(17.6–22.5)	21.2	(19.1–23.4)
Boys	18.7	(16.6–21.0)	15.5	(13.6–17.7)	14.6	(12.7–16.9)
Dizziness						
Girls	9.1	(7.5–10.9)	11.5	(10.2–13.0)	13.2	(11.7–14.9)
Boys	6.6	(5.4–7.9)	6.7	(5.6–7.9)	7.9	(6.7–9.2)
Any complaint						
Girls	53.5	(50.4–56.5)	58.1	(55.6–60.6)	63.0	(60.1–65.8)
Boys	46.9	(44.1–49.7)	45.5	(42.5–48.5)	46.2	(43.1–49.4)

out of eight complaints. In the group of 15-year-olds the OR of gender differed cross nationally in five of the eight complaints, indicating that the gender differences were significantly stronger in some countries than in others. Notably, in the group of 15-year-olds the 95% estimate of the population variance of the fixed effect of gender did not include one for any complaint. This indicates that although countries differed in the magnitude of gender differences, all countries showed the same overall pattern of girls reporting more complaints. For the group of 15-year-olds the 95% CI for the country-level variance in gender differences suggests that girls would have higher prevalence of recurrent complaints in all countries within the population of countries that were sampled for the present study.

Fig. 1 illustrates the main findings of the random coefficient models. In the figure, countries are ranked according to the size of the OR of girls reporting any recurrent complaints. First, in all countries girls showed a higher OR for recurrent health complaints. Second, within all countries except Sweden, Ireland, and Norway there was a monotone increase in the OR of gender across age groups. ORs of gender were weakest among 11-year-olds, and strongest among 15-year-olds. Third, the figure shows that the magnitude of gender differences differed across countries. Finally, it can also be seen that countries showing a high level of gender differences at age 11 were also likely to show a high gender difference in the other age groups.

Controlling for compositional effects

Cross-national variation in gender differences could occur as a result of known individual risk factors for gender differences. To test for such compositional effects, gender differences were adjusted for social support from friends and teachers, alcohol use and smoking, material conditions, and body image. If the pattern of variation in gender differences reflected a compositional effect related to these life domains, we would expect a reduction in the variation after controlling for such individual level factors. Table 5 shows the OR of gender on health complaints, before and after adjustment for compositional effects. It can be seen from

Table 4
Logistic random coefficient models of recurrent health complaints regressed on gender

Recurrent health complaint	11-year-olds				13-year-olds				15-year-olds			
	Fixed effect		Random cross-national variation		Fixed effect		Random cross-national variation		Fixed effect		Random cross-national variation	
	OR	95% CI	95% CI	p^a	OR	95% CI	95% CI	p^a	OR	95% CI	95% CI	p^a
Headache	1.58	(1.46, 1.71)	—	ns	2.01	(1.84, 2.18)	—	ns	2.90	(2.65, 3.18)	—	ns
Abdominal pains	1.86	(1.69, 2.04)	—	ns	2.29	(2.03, 2.59)	(1.38, 3.82)	*	2.67	(2.34, 3.05)	(1.53, 4.66)	*
Back pain	1.25	(1.15, 1.37)	—	ns	1.29	(1.19, 1.39)	—	ns	1.34	(1.23, 1.47)	—	ns
Dizziness	1.41	(1.26, 1.58)	—	ns	1.83	(1.63, 2.06)	(1.11, 3.02)	*	1.79	(1.59, 2.0)	(1.10, 2.89)	*
Feeling low	1.34	(1.22, 1.47)	—	ns	2.03	(1.79, 2.29)	(1.14, 3.59)	*	2.49	(2.22, 2.8)	(1.46, 4.25)	*
Irritable	1.20	(1.10, 1.31)	(0.81, 1.79)	*	1.41	(1.27, 1.56)	(0.86, 2.31)	*	1.54	(1.41, 1.69)	(1.02, 2.34)	*
Nervous	1.22	(1.14, 1.30)	—	ns	1.59	(1.47, 1.71)	—	ns	1.66	(1.52, 1.81)	(1.12, 2.45)	*
Sleeping difficulties	1.11	(1.04, 1.18)	—	ns	1.36	(1.26, 1.47)	—	ns	1.57	(1.46, 1.68)	—	ns

[a]Wald test on df = 2. * $p < .05$; ** $p < 0.01$. ns = statistically non-significant at the .05 level of significance.

the table that the cross-national variation changed very little after adjusting for these potential confounders, which indicates that the variation in gender differences was not due to these individual level factors.

Ecological analysis of gender differences and macro-level factors

To examine the aggregate relationships between estimated gender differences in health complaints, we calculated country and age-specific estimates of (1) the percentage of girl with health complaints, (2) the percentage of boys with health complaints, and (3) the corresponding rate difference. These aggregate level estimates were correlated against relevant macro-level indicators, including the UNDP GEM, the GDI, the female/male life expectancy ratio, and GDP for 1997. Table 6 shows the bivariate Spearman-rank correlation. The table shows several consistent findings. Countries' gender differences in health complaints were inversely related to the GDI among 11-year-olds, but not in the older age groups. The GEM was inversely related to the level in of health complaints in boys and girls, but not related to the countries gender difference in health complaints. GDP was correlated with the level of health complaints in girls, but not in boys. The female/male life expectancy ratio was strongly correlated with the rate difference of gender on health complaints, meaning that in countries were females were expected to live comparatively longer than men, girls tended to have a comparatively higher level of health complaints than boys.

Fig. 2 shows a scatter plot of the association between the GDI (GDI) score and the girls/boys OR of health complaints (y-axis) among 11-year-olds. It can be seen from the figure that eastern European countries tended to have a low GDI score and a high girls/boys OR of health complaints, whereas western European countries tended to show the inverse pattern.

Cross-level associations between macro-level factors and individual health complaints

To examine the impact of national level characteristic on individual health complaints, a series of cross-level interaction models were tested. In these models students were grouped in low, medium, and high based on the tertile scores of on relevant macro-level indicators. To test the cross-level interaction between country subgroup and gender, interaction terms between gender and country subgroups were computed and tested against a main effects model. The Wald test revealed a statistically significant gender by subgroup interaction for all four macro-level criteria. Since logistic interaction terms are difficult to interpret on a logit scale, we computed the posterior gender OR of health complaints through linear

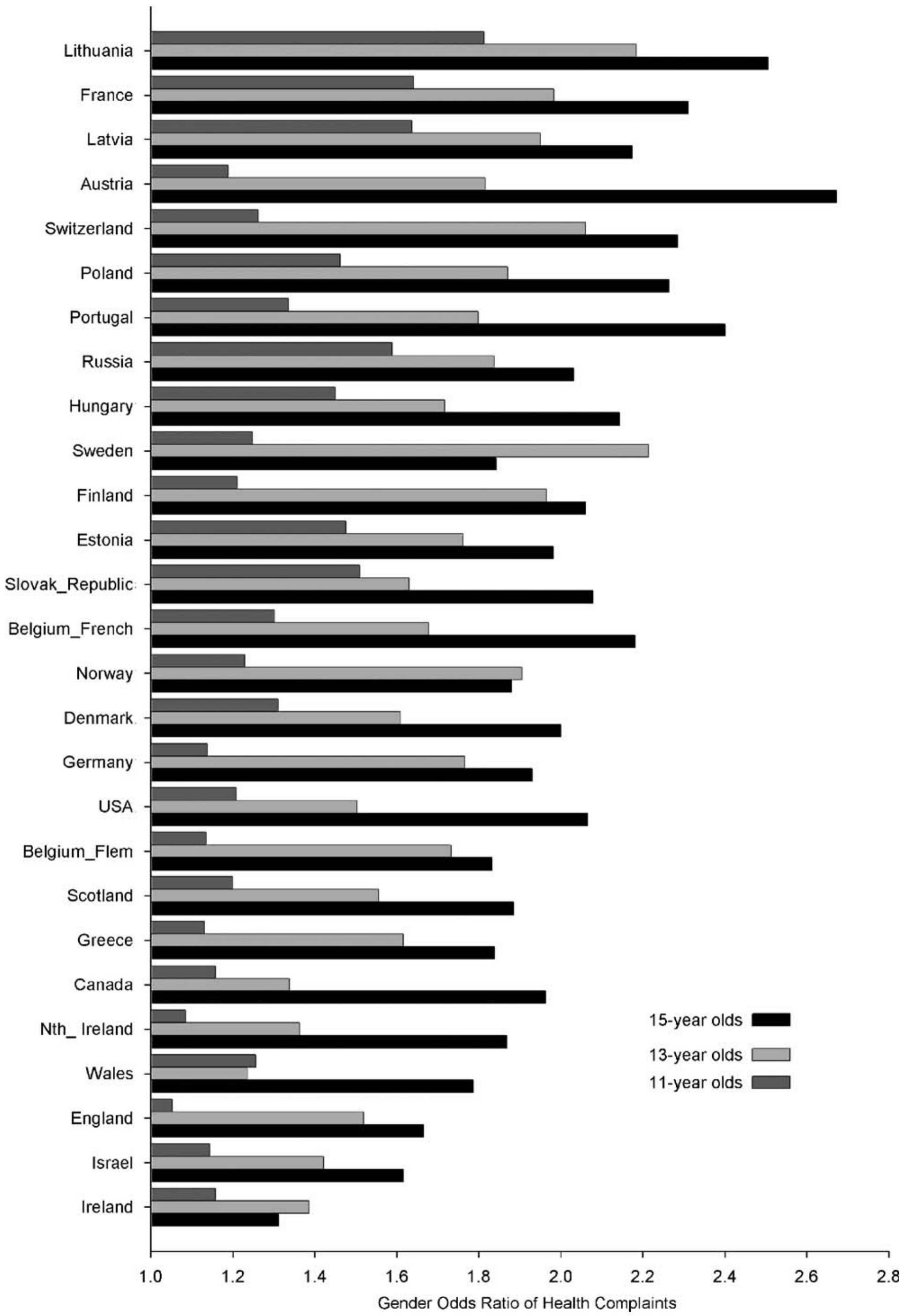

Fig. 1. Country- and age-specific odds ratios of girls reporting any recurrent complaints.

combinations of main effect terms and interaction terms, and logistic transformation of that linear combination.

It can be seen from Table 7 that the OR of gender on health complaints, varied significantly across country subgroups. In countries with a low score on the GDI, the OR of gender were higher than in countries with a high score on the GDI, and this pattern was evident for all age groups. Among the 11-year-olds, girls in countries with a low score on the GDI had an OR of 1.56 of health complaints, compared to boys. In

Table 5
Odds ratio of gender for reporting any health complaints, before and after adjustment for compositional effects

	11-year-olds			13-year-olds			15-year-olds		
	OR	95% CI	Cross national variation	OR	95% CI	Cross national variation	OR	95% CI	Cross national variation
Unadjusted	1.30	(1.20–1.40)	(0.93–1.81)	1.70	(1.57–1.84)	(1.20–2.41)	2.00	(1.85–2.17)	(1.42–2.83)
Adjusting for									
Social support in school	1.33	(1.23–1.43)	(0.97–1.82)	1.75	(1.62–1.88)	(1.28–2.38)	2.03	(1.90–2.18)	(1.54–2.69)
Alcohol use and smoking	1.42	(1.31–1.53)	(1.02–1.96)	1.84	(1.69–1.99)	(1.27–2.64)	2.16	(1.98–2.36)	(1.48–3.17)
Family affluence	1.29	(1.19–1.39)	(0.92–1.79)	1.69	(1.56–1.82)	(1.19–2.39)	1.98	(1.83–2.14)	(1.41–2.78)
Satisfaction with Body	1.19	(1.10–1.29)	(0.85–1.68)	1.48	(1.36–1.60)	(1.02–2.13)	1.74	(1.60–1.89)	(1.22–2.49)
Adjusting for all	1.32	(1.23–1.43)	(0.97–1.81)	1.64	(1.52–1.78)	(1.17–2.30)	1.92	(1.77–2.07)	(1.38–2.65)

contrast, in countries with a high score on GDI, the OR was only 1.18. For the GEM, and the GDP, there was a curvilinear, rather than a monotone increase of gender differences. Gender differences were stronger in countries with a high or a low score, as compared to girls in countries with a medium score.

Gender differences varied consistently as a function of the female/male life expectancy ratio, irrespective of age. In countries with a low female/male life expectancy ratio, 11-year-old girls were only slightly more likely than boys to report health complaints. In countries with a high female/male life expectancy ratio, the OR of gender on health complaints was 1.65. It can be seen from the table that the subgroup variation in gender differences was highly consistent across age groups.

Discussion

The present study extends previous research on gender differences in several ways. First, by studying samples of young people from all over Europe and North America, the observed gender differences findings can with some confidence be generalised to adolescents from a large proportion of the worlds industrialised countries. Second, the comprehensive set of data enabled for a more powerful statistical approach to the issue of cross-national variation than has previously been possible with data from studies with a smaller number of countries. Third, by providing empirical description and prediction of cross-national variation in gender differences, the results provide strong relevance to formulation of hypothesis regarding the role of socio-contextual factors in gender differences.

The present study confirms the findings from previous studies that gender differences are pervasive across a wide range of health complaints, and that gender differences increase from early to mid-adolescence (Sweeting & West, 2003). For all complaints covered in the present study, girls showed a higher prevalence of health complaints. The most ‘gendered’ health complaints were headache and abdominal pains, in which mid-adolescent girls had more than 2.5 higher odds than boys of reporting recurrent patterns. An inspection of the gender-specific point estimates reveals that the patterns leading to these gender differences were quite different. For headache, the gender differences in 15-year-olds resulted from a strong increase in headache for girls across age groups. For abdominal pain, however, the increasing gender differences were due to a strong *decrease* in the level of abdominal pains in boys.

The presence of specific patterns does not change the overall impression that gender differences were relatively homogenous across health complaints. The typical pattern for the majority of complaints was that the percentage of girls reporting recurrent health complaints

Table 6
Ecological analysis of associations (Spearman-Brown Rank correlation) between aggregate estimates of subjective health complaints and relevant macro-level indicators ($n = 27$)

	M	SD	Macro indicator			
			GEM	GDI	GDP	F/M LE ratio
11-year-olds						
Percentage of girls with health complaints	53	8	−0.52**	−0.23ns	−0.50**	0.25ns
Percentage of boys with health complaints	47	7	−0.34ns	0.09ns	−0.33ns	−0.17ns
Gender rate difference	6	4	−0.30ns	−0.53**	−0.40*	0.75***
13-year-olds						
Percentage of girls with health complaints	58	7	−0.56**	−0.29ns	−0.46*	0.16ns
Percentage of boys with health complaints	45	8	−0.50**	−0.08ns	−0.39*	−0.04ns
Gender rate difference	13	4	0.05ns	−0.32ns	0.01ns	0.52**
15-year-olds						
Percentage of girls with health complaints	63	7	−0.60**	−0.19ns	−0.43*	0.14ns
Percentage of boys with health complaints	46	8	−0.54**	−0.06ns	−0.27ns	−0.10ns
Gender rate difference	17	3	−0.08ns	−0.40*	−0.17ns	0.71**

* $p<0.05$, ** $p<0.01$, ns = non-significant at the 0.05 level of significance.

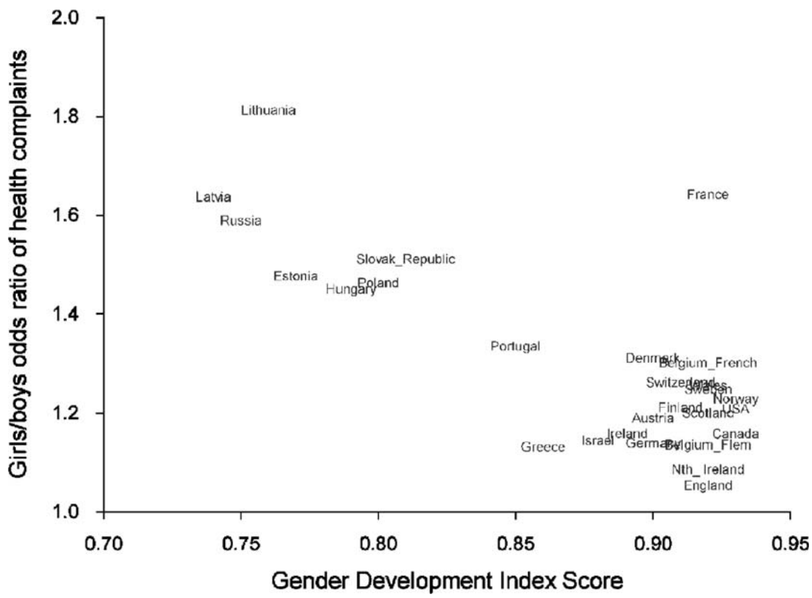

Fig. 2. Scatter plot of gender development index score (x-axis) and girls/boys odds ratio of health complaints among 11-year-olds (y-axis).

increased significantly across age groups, whereas boys showed no difference or an only small upwards or downwards difference. This overall pattern favour the view that gender differences in health complaints may reflect a relatively general processes, which create gender differences across a wide range of complaints.

Although gender differences were present across all the sampled countries, the *magnitude* of such differences varied cross nationally. The observed cross-national differences add significantly to the findings from other studies on gender differences. A gender by nation interaction was also reported in a recent WHO study comparing representative sample of 11-, 13- and 15-year-olds from Finland, Poland, Norway, and Scotland, and in a recent study of cross-national differences (Crijnen et al., 1999). Compared to these studies, the major strength of the present study was that a much larger sample of countries was included, giving a much stronger basis for making judgements about cross-national variability. The cross-national variation of

Table 7
Odds ratio of recurrent health complaints by gender, for subgroups defined by national level indicators

Group criterion	11-year-olds		13-year-olds		15-year-olds	
	OR gender	95% CI	OR gender	95% CI	OR gender	95% CI
Gender development index						
Lower GDI	1.56	(1.44–1.68)	1.88	(1.75–2.03)	2.27	(2.09–2.46)
Medium GDI	1.17	(1.08–1.27)	1.70	(1.57–1.84)	1.91	(1.76–2.07)
High GDI	1.18	(1.10–1.25)	1.56	(1.46–1.66)	1.88	(1.76–2.01)
Gender empowerment measure						
Lower GEM	1.49	(1.39–1.61)	1.84	(1.71–1.99)	2.17	(2.00–2.36)
Medium GEM	1.13	(1.05–1.22)	1.45	(1.34–1.57)	1.73	(1.60–1.88)
High GEM	1.20	(1.11–1.29)	1.75	(1.64–1.87)	2.07	(1.93–2.22)
Gross domestic product						
Lower GDP	1.56	(1.44–1.68)	1.88	(1.75–2.03)	2.27	(2.09–2.46)
Medium GDP	1.11	(1.03–1.19)	1.37	(1.28–1.47)	1.68	(1.57–1.81)
High GDP	1.29	(1.19–1.40)	1.91	(1.77–2.05)	2.17	(2.01–2.35)
Female/male life expectancy ratio						
Lower LE ratio	1.12	(1.05–1.20)	1.45	(1.36–1.56)	1.78	(1.65–1.91)
Medium LE ratio	1.18	(1.10–1.28)	1.73	(1.61–1.86)	2.01	(1.86–2.16)
High LE ratio	1.65	(1.53–1.79)	1.97	(1.82–2.12)	2.29	(2.11–2.49)

gender differences was highly robust to controlling for compositional effects in several life domains, related to body satisfaction, social relations, substance use, or material living conditions. The modest impact of controlling for potential confounders is consistent with that found in other studies (e.g. Lahelma et al., 1999; McDonough & Walters, 2001) and adds further credibility to the conclusion that the cross-national variation is not due to compositional effects.

One of the key findings from the present study was that gender differences in health complaints varied as function of gender-related characteristic at the macro level. Gender differences in health complaints were stronger in countries with a low level of gender development, than in countries with a high level of gender development. The observed pattern indicates a nonlinear relationship, in that gender differences in health complaints were stronger in countries with a low level of gender development, but did not differ much between countries at medium or high-level gender development.

Gender differences in health complaints were positively associated with the gender life expectancy ratio. In countries where women tend to live much longer than men, girls tended to have more health complaints than did boys. The life female/male life expectancy ratio reflects a compound of genetic and social factors that we are not fully able to discern in the present study. It is, however, notable that in countries with a comparatively equal gender distribution of life years, the distribution of health complaints was also comparatively equal. In previous work, the declining gender ratio in mortality has been attributed to women taking up unhealthy male behaviour patterns, and as an indication of modern gender role identity. The association between life expectancy ratio and gender differences in health complaints could thus reflect the influence of gender role development as a common third factor. A modern gender role development would make girls less susceptible to report health complaints, and boys more prone to report health complaints.

The GEM was associated with both girls and boys level of health complaints, but not with the gender difference. In countries where the gender distribution of political power and work was comparatively egalitarian, not only girls, but also boys had a lower level of health complaints. The finding that gender empowerment is beneficial for boys too, is consistent with a recent ecological study (Kawachi et al., 1999) in the US. In that study, the male mortality was lower in states with high level of women economic and political participation.

Overall, the highlighted findings suggest that gender-related societal characteristics relate systematically to the gender differences in health complaints, as well as to the overall level of health complaints in boys and girls. The external validity of these findings may, however, be difficult to ascertain. The conceptual basis for the macro-level indicators that were included in the present study has been questioned by several authors on the grounds that they are composites of conceptually very different indicators (Bardhan & Klasen, 1999; Dijkstra, 2002). As a tool for testing hypotheses about specific mechanisms these indicators may be of limited heuristic value. However, in the present context, these indicators support the view that macro-level factors needs to be taken into account in the study of gender differences in health. A natural next step would be a more detailed analysis of subcomponents in gender development.

Cross-national comparison of data such as the above, involves several sources of bias, including construct bias, method bias, and item bias (van de Vijver & Leung, 1997). Construct bias would be present if recurrent health complaints were not equally valid phenomena across countries. However, we have no data to suggest that the health complaints covered in the present study were atypical for particular countries. If health complaints were atypical for any given country, that country would appear as an 'outlier' in the statistical models. Rather, all complaints showed a fairly normally distributed cross-national variance, with some countries having a low prevalence, and other countries having a high prevalence. Although inevitable variation in the data collection exists, the available documentation (NSD, 2000) does not suggest systematic method bias with respect to countries adherence to the protocol. However, in some of the surveys, sample sizes were significantly smaller than prescribed by the protocol (Estonia, French speaking Belgium). Although this may threaten the precision of estimates, it was decided to include these samples in the analysis. Analysis without these countries provided essentially identical results. DIF represents a third kind of bias to potentially influence the results of the present study. However, our results did not suggest that DIF is a serious threat to the validity of findings. These findings indicate that items functioned relatively uniform across cultures, and that the cross-national differences are not primarily the result of item bias.

To conclude, the main contribution from this study is to provide empirical evidence of cross-national variation in gender differences in adolescent health, at a developmental stage where boys and girls have not yet entered traditional social roles related to marriage, occupation, and care-taking. Although it may be premature to make specific interpretations about this pattern, the observed cross-national variation warrants further attention to cultural influences, including political, and economical factors. The cross-national variation highlights the need for future studies to incorporate national political, economical, and cultural factor, as moderators of gender differences in health.

Acknowledgments

The authors would like to thank the following Principal Investigators of the Health Behaviour in School-Aged Children: A WHO Cross national Study (HBSC) 1997/98 for their contribution in the production of the paper: Wolfgang Dür (Austria), Lea Maes (Belgium-Flemish speaking), Danielle Piette (Belgium-French speaking), William Boyce (Canada), Ladislav Czemy (Czech Republic), Pernille Due/Bjørn Holstein (Denmark), Mary Hickman (England), Mai Maser (Estonia), Lasse Kannas (Finland), Christiane Dressen (France), Klaus Hurrelmann (Germany), Anna Kokkevi (Greece), Michael Pedersen (Greenland), Anna Aszmann (Hungary), Jossi Harel (Israel), Ieva Ranka (Latvia), Apolinaras Zaborskis (Lithuania), Grace McGuiness (Northern Ireland), Margarida Gaspar De Matos (Portugal), Saoirse Nic Gabhainn (Republic of Ireland), Barbara Woynarowska (Poland), Alexander Komkov (Russian Federation), Oddrun Samdal, (Norway), Candace Currie (Scotland), Miro Bronis (Slovakia), Ulla Marklund (Sweden), Beatrice Janin Jacquat (Switzerland), Chris-Tudor Smith (Wales), Peter Scheidt (USA).

References

Achenbach, T. M., & McConaughy, S. H. (1987). *Empirically based assessment of child and adolescent psychopathology: Practical applications*. Newbury Park, CA: SAGE.

Bardhan, K., & Klasen, S. (1999). UNDP's gender-related indices: A critical review. *World Development*, *27*(6), 985–1010.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Crijnen, A. A. M., Achenbach, T. M., & Verhulst, F. C. (1999). Problems reported by parents of children in multiple cultures: The child behavior checklist syndrome constructs. *American Journal of Psychiatry*, *156*(4), 569–574.

Currie, C. (1998). *Health behaviour in school-aged children. A WHO cross-national survey (HBSC). Research protocol for the 1997–1998 study*. Edinburgh: University of Edinburgh.

Currie, C., Hurrelmann, K., Settertobulte, W., Smith, R., & Todd, J. (Eds.). (2000). *Health and health behaviour among young people*. Copenhagen.

Diez-Roux, A. V. (2000). Multilevel analysis in public health research. *Annual Review of Public Health*, *21*, 171–192.

Diez-Roux, A. V. (2002). A glossary for multilevel analysis. *Journal of Epidemiology and Community Health*, *56*(8), 588–594.

Dijkstra, A. G. (2002). Revisiting UNDP's GDI and GEM: Towards an alternative. *Social Indicators Research*, *57*(3), 301–338.

Eriksen, H. R., Svendsrod, R., Ursin, G., & Ursin, H. (1998). Prevalence of subjective health complaints in the Nordic European countries in 1993. *European Journal of Public Health*, *8*(4), 294–298.

Frohlich, N., Carriere, K. C., Potvin, L., & Black, C. (2001). Assessing socioeconomic effects on different sized populations: To weight or not to weight? *Journal of Epidemiology and Community Health*, *55*(12), 913–920.

Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold.

Goldstein, H., & Rasbash, J. (1996). Improved approximations for multilevel models with binary responses. *Journal of the*

Royal Statistical Society Series A–Statistics in Society, *159*, 505–513.

Haugland, S., & Wold, B. (2001). Subjective health complaints in adolescence–reliability and validity of survey methods. *Journal of Adolescence*, *24*(5), 611–624.

Haugland, S., Wold, B., Stevenson, J., Aaroe, L. E., & Woynarowska, B. (2001). Subjective health complaints in adolescence—a cross-national comparison of prevalence and dimensionality. *European Journal of Public Health*, *11*(1), 4–10.

Jodoin, M. G., & Gierl, M. J. (2001). Evaluating type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. *Applied Measurement in Education*, *14*(4), 329–349.

Kawachi, I., Kennedy, B. P., Gupta, V., & Prothrow-Stith, D. (1999). Women's status and the health of women and men: A view from the States. *Social Science & Medicine*, *48*(1), 21–32.

Lahelma, E., Martikainen, P., Rahkonen, O., & Silventoinen, K. (1999). Gender differences in illhealth in Finland: Patterns, magnitude and change. *Social Science & Medicine*, *48*(1), 7–19.

Leyland, A. H., & Goldstein, H. (Eds.). (2001). *Multilevel modelling of health statistics*. Chichester: Wiley.

MacIntyre, S., Hunt, K., & Sweeting, H. (1996). Gender differences in health: Are things really as simple as they seem? *Social Science & Medicine*, *42*(4), 617–624.

McDonough, P., & Walters, V. (2001). Gender and health: Reassessing patterns and explanations. *Social Science & Medicine*, *52*(4), 547–559.

NSD. (2000). *Health Behaviour in schoolaged children 1997/98. Technical Report*. Bergen: University of Bergen.

Piccinelli, M., & Simon, G. (1997). Gender and cross-cultural differences in somatic symptoms associated with emotional distress. An international study in primary care. *Psychological Medicine*, *27*(2), 433–444.

Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healey, M., et al. (2000). *A user's guide to MLwiN, Version, 2.1*. London: University of London.

Silberg, J., Pickles, A., Rutter, M., Hewitt, J., Simonoff, E., Maes, H., et al. (1999). The influence of genetic factors and life stress on depression among adolescent girls. *Archives of General Psychiatry*, *56*(3), 225–232.

SPSS (2001). SPSS Base 11.0. SPSS Inc. Chicago.

Sweeting, H. (1995). Reversals of fortune? Sex differences in health in childhood and adolescence. *Social Science and Medicine*, *40*(1), 77–90.

Sweeting, H., & West, P. (2003). Sex differences in health at ages 11, 13 and 15. *Social Science & Medicine*, *56*(1), 31–39.

Torsheim, T., & Wold, B. (2001). School-related stress, support, and subjective health complaints among early adolescents: A multilevel approach. *Journal of Adolescence*, *24*(6), 701–713.

Torsheim, T., Wold, B., & Samdal, O. (2000). The Teacher and Classmate Support scale—Factor structure, test-retest reliability and validity in samples of 13-and 15-year-old adolescents. *School Psychology International*, *21*(2), 195–212.

UNDP. (1995). *Human Development Report 1995*. New York: Oxford University Press.

UNDP. (1999). *The human development report CD-ROM: Human development statistical database*. New York: United Nations Development Programme.

van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage.

Van-Wijk, C. M., & Kolk, A. M. (1997). Sex differences in physical symptoms: The contribution of symptom perception theory. *Social Science and Medicine*, *45*(2), 231–246.

Verbrugge, L. M. (1985). Gender and Health—an Update on Hypotheses and Evidence. *Journal of Health and Social Behavior*, *26*(3), 156–182.

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and likert-type (Ordinal) item scores*. Ottawa: Directorate of Human Resources, Research and Evaluation, Department of National Defense.