

Measuring depressive symptoms during adolescence: what is the role of gender?

C Bulhões^{1*}, E Ramos^{1,2}, M Severo^{1,2}, S Dias³ and H Barros^{1,2}

¹ EPIUnit, Instituto de Saúde Pública, Universidade do Porto, Porto, Portugal

² Department of Public Health and Forensic Sciences, and Medical Education, Faculdade de Medicina, Universidade do Porto, Porto, Portugal

³ Global Health and Tropical Medicine, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Lisboa, Portugal

Aims. The higher prevalence of depressive symptoms among females is well recognised but the reasons for this gender difference are not fully understood. There is growing evidence that current diagnostic criteria and instruments used to assess depression are less sensitive to depression in men, but studies about this issue among adolescents are scarce, especially in Latin countries. Our aim was to assess sex differences in the intensity of depressive symptoms, measured using the Beck Depression Inventory second edition (BDI-II), among Portuguese adolescents, at 13 and 17 years of age.

Methods. Urban adolescents born in 1990 and enrolled in schools of Porto, Portugal, in 2003–2004 (EPITeen study) completed the BDI-II at 13 and 17 years of age. The final sample included 1988 (52.2% girls) and 2131 (53.0% girls) adolescents at 13 and 17 years, respectively. Sex differences in the frequency of endorsing the statements on the 21 items of the BDI-II were examined using the χ^2 test and effect sizes were estimated (Cohen's *w*). To examine whether responses were linked systematically to sex, we used a differential item functioning (DIF), based on the logistic regression approach. Option characteristic curves were estimated for items with differential endorsement and a new BDI-II score was computed excluding those items.

Results. Girls and boys at the same level of depression expressed similar severity ratings for most of the depressive symptoms. We had four items with DIF at 13 and 17 years of age. At 13 years, two items provided lower scores (*sadness* and *crying* items) and two higher scores (punishment feelings and loss of interest in sex items) among boys, comparing with equally depressed girls. At 17 years, the four items with DIF provided lower scores among boys (*sadness*, *crying*, *self-dislike* and *tiredness or fatigue* items). After excluding these items the prevalence of depression remained higher among girls but at 17 years the difference between sexes was attenuated.

Conclusions. Sex differences were found in the functioning of the BDI-II, more relevant at 17 years of age, which may lead to an overestimation of symptoms among girls as well as to lower reported rates of depression among boys. For a higher diagnostic accuracy it is important that the criteria and instruments used to assess depression adequately reflect female and male common symptoms and experiences of depression.

Received 6 December 2016; Accepted 4 June 2017; First published online 28 June 2017

Key words: Adolescents, depression, gender differences, psychological assessment.

Introduction

Depressive symptoms in adolescents are a rising concern (Dick and Ferguson, 2015). Depression is more prevalent in females than males, a finding that begins in adolescence and persists throughout adulthood (Angold *et al.* 1998). This gender imbalance in depression is described as one of the most robust findings in epidemiological research (Kuehner, 2003), nevertheless it has also been called as one of the major unresolved

problems in psychiatric epidemiology (Bebbington *et al.* 1998). The reasons for this post-pubertal-onset gender difference are not fully understood (Kaltiala-Heino *et al.* 2003; Angold and Costello, 2006; Thapar *et al.* 2012) and a range of hypotheses have been advanced. Differences in help-seeking behaviour are unlikely to explain the higher rates of depression in girls since this is seen both in nonclinical and clinical samples, and is robust across different methods of assessment (Kessler *et al.* 1993; Thapar *et al.* 2012). Biological and psychosocial factors are also proposed to justify the more common onset of depression in females. The effect of sex steroids on the maturing hypothalamic-pituitary-adrenal axis might increase female sensitivity to stress, whereas androgens appear to play a

* Address for correspondence: C. Bulhões, Instituto de Saúde Pública, Universidade do Porto, Rua das Taipas, no. 135, 4050-600 Porto, Portugal.

(Email: claudiapbulhoes@gmail.com)

protective role in males (Chaplin *et al.* 2009; Thapar *et al.* 2012). Psychosocial factors include gender differences in stress coping and/or coping techniques, gender-specific expectations and differences in social cognitive function, with females presenting a greater sensitivity to rejection (Angold *et al.* 1998; Hankin, 2006; Chaplin *et al.* 2009; Naninck *et al.* 2011).

Another possibility for explaining this gender difference found in epidemiological research can be related to the diagnostic criteria or instruments used to assess depression (Santor *et al.* 1994; Bennett *et al.* 2005). It is assumed that symptoms or questions are equally effective in assessing depression in both genders (Santor *et al.* 1994; Salokangas *et al.* 2002) although it is well recognised that females and males experience and express their symptoms in a different way (Brooks, 2001; Brownhill *et al.* 2005; Weller *et al.* 2006; Martin *et al.* 2013; Rice *et al.* 2013). There is a growing suggestion that the current diagnostic criteria and instruments used to assess depression include symptoms that represent a feminine-gender pattern of the disorder, being less sensitive to depression in men (Rochlen *et al.* 2010; Martin *et al.* 2013; Castro *et al.* 2015), and male-specific depression rating scales have also been proposed (Zierau *et al.* 2002; Martin *et al.* 2013; Rice *et al.* 2013).

According to the studies available, adolescent females may more likely disclose their feelings (Breslau *et al.* 2008). Data from adult populations show that depressed women often experience changes in appetite, weight gain, carbohydrate craving, sleep disturbances, increased sense of failure or guilt, somatisation, crying and increased anger (Bennett *et al.* 2005; Weller *et al.* 2006; Romans *et al.* 2007). Men may experience alternative depressive symptoms. Traditional depressive symptoms (e.g., sadness, crying) are at odds with societal ideals of masculinity and men may be reluctant to report experiencing those symptoms (Castro *et al.* 2015). Men's depression commonly manifests itself as anger, impulse control difficulties, anxiety and gender-role discord, irritability, aggression, substance abuse, risk-taking and escaping behaviours, emotional numbness, inability or unwillingness to express emotion, impoverished relationships and suicide (Rutz *et al.* 1997; Bennett *et al.* 2005; Brownhill *et al.* 2005; Weller *et al.* 2006; Martin *et al.* 2013; Rice *et al.* 2013). Although it is believed that depressed adolescents share gender similarities with depressed adults (Bennett *et al.* 2005; Weller *et al.* 2006), studies about this issue among adolescents are scarce, especially in Latin countries, since gender patterns and masculinity are expected to be culturally defined. Furthermore, it is still unclear how early such gender specific depressive symptom patterns emerge.

The Beck Depression Inventory (BDI) is widely used among adults and adolescents to measure depressive

symptoms and has a strong track record in depression research (Roberts *et al.* 1991; Myers and Winters, 2002; Carnevale, 2011; Stockings *et al.* 2015). It assesses cognitive, behavioural, affective and somatic dimensions of depression (Beck *et al.* 1961; Ambrosini *et al.* 1991; Roberts *et al.* 1991; Bennett *et al.* 2005), being described as a relevant instrument with robust psychometric properties (Ambrosini *et al.* 1991; Roberts *et al.* 1991; Osman *et al.* 2008; Wang and Gorenstein, 2013; Stockings *et al.* 2015).

Our aim was to assess sex differences in the intensity of depressive symptoms, measured using the BDI, among Portuguese adolescents, to understand whether the gender difference in the prevalence of depression could be partially explained by the characteristics of the measurement instrument.

Method

This study is part of the Epidemiological Health Investigation of Teenagers in Porto (EPITeen). This population-based cohort was assembled during 2003–2004, when adolescents born in 1990, enrolled at public and private schools of Porto, Portugal, were recruited, as previously described (Ramos and Barros, 2007). A second evaluation took place when participants were on average 17 years of age (2007–2008). Participants were evaluated at schools and information was collected using self-administered questionnaires and performing a physical examination. In both study waves the procedures were standardised and completed by a team of trained health professionals. Written informed consent was obtained both from the adolescents and their parents or legal guardians. The study was approved by the Ethics Committee of Hospital S. João.

At the recruitment, 2786 eligible participants were identified and 2159 (77.5%) participated (77.9% in public and 77.0% in private schools; $p=0.71$). We excluded 171 (7.9%) adolescents with missing information on BDI second edition (BDI-II). In the second wave (at 17 years), 1716 participants (79.5%) were re-evaluated. Further 783 adolescents born in 1990 who moved to the schools in Porto during this time frame were evaluated for the first time at 17 years. We excluded 368 (14.7%) adolescents because of missing information on the BDI-II. The final sample comprised 1988 (52.2% girls) and 2131 (53.0% girls) adolescents at 13 and 17 years, respectively.

Depressive symptomatology measurement

Depressive symptoms at 13 and 17 years of age were assessed using the BDI-II (Beck *et al.* 1996). This scale

consists of 21 items corresponding to 21 different symptoms, with four (two items with seven) statement responses representing how the respondent has been feeling during the previous 2 weeks and current day (Beck et al. 1996). The four-response statements are presented in order of increasing severity and are scored from 0 to 3 (two items are provided with alternative statements sharing the same score) (Beck et al. 1996). Responses are summed, yielding scores from 0 to 63, with higher scores indicating greater depressive symptoms (Beck et al. 1996).

The BDI-II was previously validated in Portuguese adolescents and scores over 13 are considered to represent significant depressive symptoms (Coelho et al. 2002).

Data analysis

Separate analyses were performed at 13 and 17 years of age. Sex differences in the frequency of endorsing the statements on the 21 items of BDI-II were examined using the χ^2 test. As higher scores were less frequent (the relative frequency of item scores of 3 fell below 5%), item scores 2 and 3 were combined and three levels of intensity were considered: score 0, score 1 and scores 2–3. We used Cohen's w effect sizes to indicate the magnitude of differences between females and males removing the dependence on sample size, and values of 0.10, 0.30 and 0.50 or greater were considered small, moderate and large effect sizes, respectively (Cohen, 1992).

BDI-II has items, which may work in different ways concerning gender, representing a severe threat to the validity of this measure of intensity of depressive symptoms. To examine whether responses were linked systematically to sex, we used a differential item functioning (DIF) analysis, based on the logistic regression approach (Swaminathan and Rogers, 1990). The DIF is the unexpected difference in response to a test item between two populations, once the attribute that the test is measuring is controlled (Bares et al. 2012). Since the possible answers correspond to a Likert scale (score 0 to 3) we used the proportional logistic regression to estimate proportional odds ratios (POR) with respective 95% confidence intervals (95% CI). For each of the 21-items we compared the probability of selecting a statement between females and males (females as reference group), adjusting for the total BDI-II score. An item was considered to have DIF when the proportional logistic regression showed a significant association between sex and the probability of selecting a statement (score), after adjusting for BDI-II score. We considered $\text{POR} > 1.5$ and $\text{POR} \leq 1/1.5$ as clinically relevant.

Option characteristic curves were computed for boys and girls separately for the items with DIF, and were plotted as a function of expected total score. This technique computes an individual's response pattern at each evaluation point with respect to the option characteristic curves estimated from the entire sample (Santor et al. 1994).

For each sex and age (13 and 17 years) a new BDI global score was computed excluding items showing evidence of DIF. We use the mean values of the new score multiplied by the original number of items (21 items), to make both scores (complete and without items with DIF) equivalent and comparable. Prevalence of depressive symptoms was calculated with both BDI scores and compared with the McNemar's χ^2 test.

All the analysis was performed including only adolescents assessed both at 13 and 17 years. Secondly, in order to evaluate if the similarities between 13 and 17 years could be a consequence of having the same sample evaluated at both moments we analysed, at 17 years, only the adolescents who newly joined the study for the first time at this age (data not shown). As we obtained similar results we used the larger sample.

Statistical analyses were performed using R 2012 (R Foundation for Statistical Computing, Vienna, Austria). We assume a significance level of 0.05.

Results

At 13 years of age, the overall prevalence of depressive symptoms (BDI-II score > 13) was 13.4% (18.8% in girls and 7.6% in boys, $p < 0.001$). The median score (P25–P75) of the depressive symptoms was 6.01 (3.00–11.00) among females and 3.00 (1.01–6.99) among males ($p < 0.001$). At 17 years of age, the prevalence of depressive symptoms was similar, 12.7%, also with a higher prevalence among females (17.9% *v.* 6.8%, $p < 0.001$). The median score (P25–P75) of depressive symptoms was 6.00 (2.00–11.00) and 3.00 (1.00–6.00) among females and males, respectively ($p < 0.001$).

Differential item functioning

BDI-II showed good reliability at both assessments (at 13 years the Cronbach alpha was 0.88 and 0.86 among females and males, respectively; at 17 years it was 0.87 and 0.84 among females and males, respectively). Table 1 shows the distribution of answers on each of the 21-items of the BDI-II, by sex. Globally, females chose statements with significantly higher scores both at 13 and 17 years, but effect sizes were considered small. Only items 1 and 10 presented moderate

Table 1. Sex differences in score endorsement on the 21 items of Beck Depression Inventory II (BDI-II), at 13 and 17 years of age

BDI-II items	13 years of age (Girls: 1037; Boys: 951)					17 years of age (Girls: 1130; Boys: 1001)				
	Score 0 <i>n</i> (%)	Score 1 <i>n</i> (%)	Score 2–3 <i>n</i> (%)	Effect size (Cohen's <i>w</i>)	<i>p</i> -value ^a	Score 0 <i>n</i> (%)	Score 1 <i>n</i> (%)	Score 2–3 <i>n</i> (%)	Effect size (Cohen's <i>w</i>)	<i>p</i> -value ^a
1. Sadness										
Girls	648 (62.5)	364 (35.1)	24 (2.3)	0.23	<0.001	718 (63.7)	389 (34.5)	21 (1.9)	0.23	<0.001
Boys	787 (83.1)	148 (15.6)	12 (1.3)			838 (83.9)	151 (15.1)	10 (1.0)		
2. Pessimism										
Girls	718 (70.0)	221 (21.6)	86 (8.4)	0.15	<0.001	689 (61.4)	387 (34.5)	47 (4.2)	0.19	<0.001
Boys	774 (82.6)	129 (13.8)	34 (3.6)			780 (78.3)	186 (18.7)	30 (3.0)		
3. Past failure										
Girls	749 (72.9)	220 (21.4)	59 (5.7)	0.12	<0.001	809 (72.0)	258 (23.0)	56 (5.0)	0.13	<0.001
Boys	777 (82.5)	140 (14.9)	25 (2.7)			818 (82.0)	163 (16.3)	16 (1.6)		
4. Loss of pleasure										
Girls	711 (69.3)	246 (24.0)	69 (6.7)	0.10	<0.001	849 (75.7)	211 (18.8)	61 (5.4)	0.12	<0.001
Boys	731 (78.1)	160 (17.1)	45 (4.8)			846 (85.2)	114 (11.5)	33 (3.3)		
5. Guilty feelings										
Girls	627 (60.9)	367 (35.6)	36 (3.5)	0.09	<0.001	695 (61.6)	397 (35.2)	36 (3.2)	0.13	<0.001
Boys	655 (69.5)	265 (28.1)	22 (2.3)			736 (73.8)	244 (24.5)	17 (1.7)		
6. Punishment feelings										
Girls	701 (68.3)	257 (25.0)	69 (6.7)	0.05	0.076	879 (78.1)	189 (16.8)	57 (5.1)	0.11	<0.001
Boys	676 (72.1)	218 (23.2)	44 (4.7)			858 (86.1)	116 (11.6)	23 (2.3)		
7. Self-dislike										
Girls	829 (80.6)	122 (11.9)	78 (7.6)	0.14	<0.001	869 (77.6)	158 (14.1)	93 (8.3)	0.18	<0.001
Boys	853 (90.5)	56 (5.9)	34 (3.6)			898 (90.5)	71 (7.2)	23 (2.3)		
8. Self-criticalness										
Girls	667 (65.1)	235 (22.9)	122 (11.9)	0.16	<0.001	755 (67.4)	247 (22.1)	118 (10.5)	0.10	<0.001
Boys	747 (79.6)	129 (13.7)	63 (6.7)			752 (76.3)	160 (16.2)	74 (7.5)		
9. Suicidal thoughts or wishes										
Girls	920 (89.9)	80 (7.8)	23 (2.2)	0.07	0.005	1028 (91.6)	83 (7.4)	11 (1.0)	0.09	<0.001
Boys	882 (93.9)	44 (4.7)	13 (1.4)			947 (95.9)	32 (3.2)	9 (0.9)		
10. Crying										

Continued

Table 1. Continued

BDI-II items	13 years of age (Girls: 1037; Boys: 951)					17 years of age (Girls: 1130; Boys: 1001)				
	Score 0 <i>n</i> (%)	Score 1 <i>n</i> (%)	Score 2–3 <i>n</i> (%)	Effect size (Cohen's <i>w</i>)	<i>p</i> -value ^a	Score 0 <i>n</i> (%)	Score 1 <i>n</i> (%)	Score 2–3 <i>n</i> (%)	Effect size (Cohen's <i>w</i>)	<i>p</i> -value ^a
Girls	660 (63.8)	254 (24.6)	120 (11.6)	0.34	<0.001	726 (64.9)	256 (22.9)	137 (12.2)	0.31	<0.001
Boys	867 (91.9)	40 (4.2)	37 (3.9)			898 (90.7)	67 (6.8)	25 (2.5)		
11. Agitation										
Girls	673 (65.6)	291 (28.4)	62 (6.0)	0.15	<0.001	742 (66.3)	319 (28.5)	58 (5.2)	0.15	<0.001
Boys	741 (78.7)	165 (17.5)	36 (3.8)			777 (78.7)	191 (19.4)	19 (1.9)		
12. Loss of interest										
Girls	835 (81.2)	137 (13.3)	56 (5.4)	0.11	<0.001	868 (77.5)	211 (18.8)	41 (3.7)	0.10	<0.001
Boys	831 (88.1)	93 (9.9)	19 (2.0)			843 (85.2)	123 (12.4)	23 (2.3)		
13. Indecisiveness										
Girls	689 (67.0)	289 (28.1)	51 (5.0)	0.17	<0.001	772 (68.9)	292 (26.1)	56 (5.0)	0.12	<0.001
Boys	772 (82.0)	143 (15.2)	26 (2.8)			786 (79.7)	170 (17.2)	30 (3.0)		
14. Worthlessness										
Girls	851 (82.9)	116 (11.3)	59 (5.8)	0.10	<0.001	941 (84.6)	111 (10.0)	60 (5.4)	0.11	<0.001
Boys	834 (89.6)	67 (7.2)	30 (3.2)			899 (91.6)	63 (6.4)	19 (1.9)		
15. Loss of energy										
Girls	774 (74.9)	221 (21.4)	39 (3.8)	0.12	<0.001	753 (66.7)	304 (26.9)	72 (6.4)	0.18	<0.001
Boys	797 (84.1)	124 (13.1)	27 (2.8)			818 (82.2)	152 (15.3)	25 (2.5)		
16. Changes in sleeping pattern										
Girls	571 (55.4)	357 (34.6)	103 (10.0)	0.10	<0.001	518 (46.1)	488 (43.4)	118 (10.5)	0.13	<0.001
Boys	608 (64.4)	280 (29.7)	56 (5.9)			570 (57.5)	367 (37.0)	55 (5.5)		
17. Irritability										
Girls	739 (71.9)	249 (24.2)	40 (3.9)	0.14	<0.001	754 (66.8)	312 (27.7)	62 (5.5)	0.18	<0.001
Boys	792 (83.7)	138 (14.6)	16 (1.7)			815 (82.0)	159 (16.0)	20 (2.0)		
18. Changes in appetite										
Girls	557 (53.8)	410 (39.6)	68 (6.6)	0.16	<0.001	628 (55.7)	430 (38.2)	69 (6.1)	0.19	<0.001
Boys	655 (69.2)	240 (25.4)	51 (5.4)			730 (73.6)	233 (23.5)	29 (2.9)		
19. Concentration difficulty										
Girls	499 (48.5)	456 (44.4)	73 (7.1)	0.20	<0.001	586 (52.3)	421 (37.6)	114 (10.2)	0.16	<0.001
Boys	644 (68.2)	253 (26.8)	47 (5.0)			678 (68.3)	254 (25.6)	60 (6.0)		
20. Tiredness or fatigue										
Girls	722 (70.2)	276 (26.8)	30 (2.9)	0.16	<0.001	706 (62.6)	368 (32.7)	53 (4.7)	0.20	<0.001
Boys	784 (82.8)	134 (14.1)	29 (3.1)			804 (89.0)	171 (17.2)	19 (1.9)		

effect sizes, both at 13 and 17 years. Item 10 (*Crying*) revealed the highest differences between sexes, at both ages, with boys selecting the highest category 3-times less frequently than girls. Concerning item 7 (*Self-dislike*), item 18 (*Changes in appetite*), item 19 (*Concentration difficulty*), item 20 (*Tiredness or fatigue*) and item 21 (*Lost of interest in sex*), the differences between sexes increased from 13 to 17 years of age, with boys choosing the highest category for the last four items approximately 2-times less frequently than girls at 17 years.

After adjustment for global BDI-II score, the differences between sexes were attenuated (Table 2). Item 10 continued to have the highest sex differences at both ages, with boys presenting a lower probability of choosing statements with higher scores (POR = 0.21, 95% CI: 0.15–0.28 at 13 years; POR = 0.27, 95% CI: 0.20–0.35 at 17 years). Item 1 (*Sadness*) also presented similar results at both ages, with boys having a lower probability of choosing options with higher scores (POR = 0.50, 95% CI: 0.39–0.63 at 13 years; POR = 0.59, 95% CI: 0.46–0.76 at 17 years). At 13

years of age, we also found a significant difference for item 6 (*Punishment feelings*) and item 21 (*Loss of interest in sex*); for both items boys presented a higher probability of choosing a higher score. At 17 years old, the differences were for item 7 (*Self-dislike*) and item 20 (*Tiredness or fatigue*); in these items, boys presented a lower probability of choosing a higher score. We analysed the data for a significant interaction between sex and BDI-II score and there were no items with non-uniform DIF (data not shown).

Option characteristic curves for items with DIF

The option characteristic curves for items with DIF are represented in Fig. 1. As we have few adolescents with higher scores of depression, the analysis at higher ranges of depression must be made with caution.

Concerning item 1, the statements with higher scores were rarely endorsed and the differences between sexes are related with the probability of choosing the score 1 (*I feel sad much of time*). This statement was less frequently endorsed by boys, both at 13 and 17 years, independently of the expected total score. Regarding item 10, at both ages, boys tended to more frequently endorse the score 0 (*I don't cry anymore than I used to*) and girls tended to more frequently choose the statements with scores 2 and 3 than boys. For this item the difference between sexes was higher in the higher scores of depression.

Items 6 and 21 only had sex different performances at 13 years of age. For item 6, few adolescents chose statements with higher scores, so the sex differences are mostly due to the probability of endorsing score 0 (*I don't feel I am being punished*) and score 1 (*I feel I may be punished*). For this item the probability of choosing score 1 or higher was greater among boys. In item 21, girls were more likely to choose Option 0 (*I have not noticed any recent change in my interest in sex*) and the sex difference was higher for higher scores of depression.

The performance of items 7 and 20 differed by sex only at 17 years of age. For item 7, boys were less likely to choose score 1 (*I have lost confidence in myself*) and score 2 (*I am disappointed in myself*). At higher levels of depression the probability of choosing score 1 was similar between boys and girls and increases the probability of choosing score 2 among girls. For item 20 the effect is similar to the reported for item 7.

Prevalence of depressive symptoms

The prevalence of depression after excluding the items with DIF (items 1, 6, 10 and 21 at 13 years of age; items 1, 7, 10 and 20 at 17 years of age) was higher in all groups when using the short version (Table 3) but remained higher among girls. Nevertheless, at 17

Table 2. Proportional Odds Ratio (POR) for endorsing a score on items of Beck Depression Inventory II (BDI-II), at 13 and 17 years of age, using girls as reference and adjusting for BDI-II global score

BDI-II items	POR (95% CI)	
	13 years of age	17 years of age
1. Sadness	0.50 (0.39–0.63)	0.59 (0.46–0.76)
2. Pessimism	0.80 (0.63–1.02)	0.80 (0.64–0.99)
3. Past failure	1.03 (0.80–1.32)	1.09 (0.85–1.39)
4. Loss of pleasure	1.13 (0.90–1.41)	1.10 (0.85–1.42)
5. Guilty feelings	1.27 (1.03–1.58)	1.11 (0.90–1.39)
6. Punishment feelings	1.50 (1.23–1.87)	1.13 (0.87–1.46)
7. Self-dislike	0.79 (0.58–1.09)	0.67 (0.50–0.90)
8. Self-criticalness	0.78 (0.62–0.99)	1.34 (1.08–1.68)
9. Suicidal thoughts or wishes	1.19 (0.80–1.77)	1.24 (0.79–1.93)
10. Crying	0.21 (0.15–0.28)	0.27 (0.20–0.35)
11. Agitation	0.82 (0.66–1.02)	0.89 (0.72–1.11)
12. Loss of interest	1.11 (0.82–1.50)	1.41 (1.07–1.84)
13. Indecisiveness	0.73 (0.57–0.92)	1.11 (0.88–1.39)
14. Worthlessness	1.23 (0.88–1.72)	1.36 (0.95–1.93)
15. Loss of energy	1.04 (0.81–1.35)	0.77 (0.61–0.98)
16. Changes in sleeping pattern	1.02 (0.84–1.23)	1.00 (0.83–1.20)
17. Irritability	0.83 (0.65–1.07)	0.80 (0.63–1.01)
18. Changes in appetite	0.82 (0.67–0.99)	0.70 (0.58–0.86)
19. Concentration difficulty	0.69 (0.56–0.84)	0.90 (0.74–1.09)
20. Tiredness or fatigue	0.89 (0.70–1.13)	0.64 (0.51–0.80)
21. Loss of interest in sex	1.86 (1.24–2.82)	1.23 (0.83–1.82)

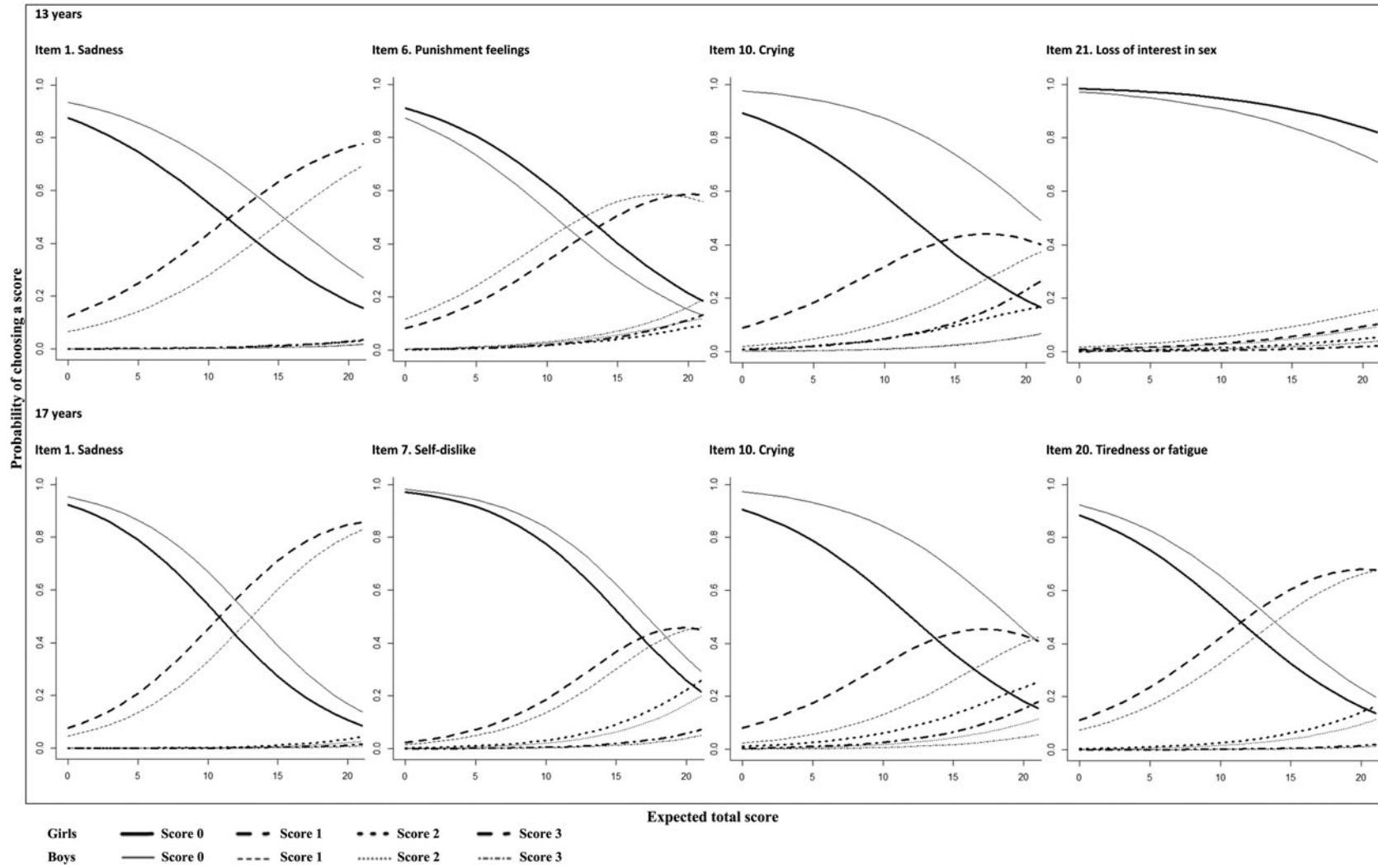


Fig. 1. Option characteristic curves for Beck Depression Inventory II items with differential item functioning for 13 and 17-year-old adolescents.

Table 3. Prevalence of Depressive Symptoms (Beck Depression Inventory II total score > 13) including and excluding the items with differential item functioning

	BDI-II > 13 (21 items) n (%)	BDI-II > 13 (17 items) n (%)	p-value ^a
13 years of age			
Girls	195 (18.8)	213 (20.5)	0.004
Boys	72 (7.6)	84 (8.8)	0.014
17 years of age			
Girls	202 (17.9)	208 (18.4)	0.429
Boys	68 (6.8)	84 (8.4)	<0.001

^aMcNemar's χ^2 test with continuity correction.

years the short version attenuated the difference: using the usual 21 items scale the prevalence in girls was 2.6 times higher than in boys, but using the short version the ratio decreased to 2.2 times.

Discussion

According to our results, females chose statements with higher scores at 13 and 17 years, with item 10 (*Crying*) revealing the highest differences between sexes, at both ages. However, we found that girls and boys at the same level of depression expressed similar severity ratings for most of the depressive symptoms evaluated, both at 13 and 17 years. Previous studies of youth (Roberts *et al.* 1995; Kovacs, 2001; Masi *et al.* 2001; Bennett *et al.* 2005) and adult (Young *et al.* 1990) samples made similar observations. However, potential differences emerged and there was evidence for DIF in 4 of the 21 items of the BDI-II at 13 and 17 years of age. At 13 years of age, two items with DIF provided lower scores (*sadness* and *crying* items) and the other two higher scores (*punishment feelings* and *loss of interest in sex* items) among boys, compared with girls at similar overall levels of depressive symptoms. At 17 years, the four items with DIF provided lower scores among boys (*sadness*, *crying*, *self-dislike* and *tiredness or fatigue* items).

As at 13 years we have two items more related to the female sex and other two to male sex, this DIF may not be relevant at this age. On the other hand, at 17 years we have four items particularly related with female sex. It is possible that the prevalence of depression is being underestimated among boys and overestimated among girls, as described in previous research, as those items are more related to the female gender socially or biologically than to depression *per se* (Salokangas *et al.* 2002; van Beek *et al.* 2012). When sex differences were accounted for (excluding the items with DIF), the gap regarding the prevalence of depressive symptoms among girls and boys at 17

years was reduced. However, since depression was also more than twice common in girls than in boys, at both ages, it is unlikely that this gap is entirely related with the screening instrument used at this assessment. Finding similar results in the group of adolescents who newly joined the study for the first time at 17 years when compared with the adolescents at 17 years who were also evaluated at 13, minimises the limitation related to the dependence of samples between 13 and 17 years.

The lower scores presented by boys concerning *sadness* and *crying* items, comparing with equally depressed girls, are consistent with studies of depressed adolescents (Kovacs, 2001; Bares *et al.* 2012; van Beek *et al.* 2012) and adults (Young *et al.* 1990; Castro *et al.* 2015). *Sadness* seems to be a gender-bound emotional reaction not necessarily related to depression (Newmann, 1986; George *et al.* 1996), being less commonly identified and expressed by males due to different socialisation processes (Bennett *et al.* 2005). The same seems to be true for *crying*. Unlike men, females express their distress more often by *crying* and use this as a coping mechanism (Williams and Morris, 1996), even during adolescence (Bennett *et al.* 2005; Bares *et al.* 2012). Prior research also found that boys value less the self-concept and give less importance to fatigue when compared with girls (Stehouwer *et al.* 1985; Siegel *et al.* 1999; Khan *et al.* 2002; Bennett *et al.* 2005; van Beek *et al.* 2012). Finding lower scores among boys on the items *self-dislike* and *tiredness or fatigue* reinforces those results. A possible explanation for having significant results only at 17 years of age relies on the fact that we had more girls choosing the statements with the highest scores at this age.

We did not find described in literature higher scores among boys on *punishment feelings* and *interest in sex*, at 13 years of age. According to adult data, it is possible that for cultural or biological reasons, both depressive and non-depressive girls are less interested in sex than boys (Salokangas *et al.* 2002). Thus, including questions concerning interest in sex can also give gender-bound biased results concerning depressive symptoms as depressive boys but not depressed girls can express changes in this issue more commonly. The small number of boys at 17 years that choose the statement with the highest score might explain not having significant results at this age.

The strengths of this study consist in its large population-based sample at both ages of evaluation, early and late adolescence, comprising adolescents of a major Portuguese city. Our findings are specific for the BDI-II. Although the BDI is one of the most widely used self-reported measures of depression in research and clinical practice (Wang and Gorenstein, 2013; Stockings *et al.* 2015), other depression questionnaires

should be examined for similar sex differences. Given the overlap of the construct measured by BDI-II with that of other widely used scales to assess depression (Wang and Gorenstein, 2013) it is expected that instruments including questions related to these items with DIF can also give sex-bound biased results concerning depressive symptoms. However, the results should be viewed in the context of the following limitation. In this method of examining response characteristic curves as a function of depression, the expected total scores are themselves derived from the items we wish to evaluate. We had no formal diagnosis of depression and this self-report measure for screening depression does not necessarily refer to depression of clinical relevance.

In conclusion, sex differences were found in the functioning of the inventory, more relevant at 17 years of age. The use of BDI-II as an instrument for evaluating depressive symptoms may lead to an overestimation of symptoms among girls as well as to lower reported rates of depression among boys. Moreover, BDI-II closely parallels DSM-IV, both including symptoms that represent a feminine-genders pattern of depression. According to DSM-IV, depressed mood is one of two essential symptoms required to make a diagnosis of major depression, indicated by either subjective report (e.g., feels sad or empty) or observation made by others (e.g., appears tearful). As sadness and crying seems to be gender-bound emotional reactions, males with depression may not be identified in clinical practice and consequently untreated. For a higher diagnostic accuracy it is important that the criteria and instruments used to assess depression adequately reflect female and male common symptoms and experiences of depression.

Acknowledgements

The authors would like to thank the families enrolled in EPITeen for their kindness and all the staff involved in the evaluations for their help and support.

Financial Support

This study was supported through FEDER from the Operational Programme Factors of Competitiveness – COMPETE and through national funding from the Portuguese Foundation for Science and Technology – FCT (Portuguese Ministry of Education and Science) within the project PTDC/DTP-EPI/6506/2014, and by the Epidemiology Research Unit – Institute of Public Health, University of Porto (UID/DTP/047507/2013). Individual grant attributed to CB (SFRH/SINTD/60138/2009) was supported by the Portuguese Foundation for Science and Technology – FCT.

Conflict of Interest

None.

Availability of Data and Materials

Data supporting the findings of this study are available to other investigators by contacting the authors.

References

- Ambrosini PJ, Metz C, Bianchi MD, Rabinovich H & Undie A** (1991). Concurrent validity and psychometric properties of the Beck Depression Inventory in outpatient adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry* **30**, 51–57.
- Angold A & Costello EJ** (2006). Puberty and depression. *Child and Adolescent Psychiatric Clinics of North America* **15**, 919–937, ix.
- Angold A, Costello EJ, Worthman CM** (1998). Puberty and depression: the roles of age, pubertal status and pubertal timing. *Psychological Medicine* **28**, 51–61.
- Bares C, Andrade F, Delva J, Grogan-Kaylor A, Kamata A** (2012). Differential item functioning due to gender between depression and anxiety items among Chilean adolescents. *International Journal of Social Psychiatry* **58**, 386–392.
- Bebbington PE, Dunn G, Jenkins R, Lewis G, Brugha T, Farrell M, Meltzer H** (1998). The influence of age and sex on the prevalence of depressive conditions: report from the National Survey of Psychiatric Morbidity. *Psychological Medicine* **28**, 9–19.
- Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J** (1961). An inventory for measuring depression. *Archives of General Psychiatry* **4**, 561–571.
- Beck AT, Steer RA, Brown GK** (1996). *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corp.
- Bennett DS, Ambrosini PJ, Kudes D, Metz C, Rabinovich H** (2005). Gender differences in adolescent depression: do symptoms differ for boys and girls? *Journal of Affective Disorders* **89**, 35–44.
- Breslau J, Javaras KN, Blacker D, Murphy JM, Normand SL** (2008). Differential item functioning between ethnic groups in the epidemiological assessment of depression. *Journal of Nervous and Mental Disease* **196**, 297–306.
- Brooks GR** (2001). Masculinity and men's mental health. *Journal of American College Health* **49**, 285–297.
- Brownhill S, Wilhelm K, Barclay L, Schmied V** (2005). 'Big build': hidden depression in men. *Australian and New Zealand Journal of Psychiatry* **39**, 921–931.
- Carnevale T** (2011). An integrative review of adolescent depression screening instruments: applicability for use by school nurses. *Journal of Child and Adolescent Psychiatric Nursing* **24**, 51–57.
- Castro SM, Curi M, Torman VB, Riboldi J** (2015). Differential item functioning in the beck depression inventory. *Revista Brasileira de Epidemiologia* **18**, 54–67.
- Chaplin TM, Gillham JE, Seligman ME** (2009). Gender, anxiety, and depressive symptoms: a longitudinal study

- of early adolescents. *Journal of Early Adolescence* **29**, 307–327.
- Coelho R, Martins A, Barros H** (2002). Clinical profiles relating gender and depressive symptoms among adolescents ascertained by the Beck Depression Inventory II. *European Psychiatry* **17**, 222–226.
- Cohen J** (1992). A power primer. *Psychological Bulletin* **112**, 155–159.
- Dick B, Ferguson BJ** (2015). Health for the world's adolescents: a second chance in the second decade. *Journal of Adolescent Health* **56**, 3–6.
- George MS, Ketter TA, Parekh PI, Herscovitch P, Post RM** (1996). Gender differences in regional cerebral blood flow during transient self-induced sadness or happiness. *Biological Psychiatry* **40**, 859–871.
- Hankin BL** (2006). Adolescent depression: description, causes, and interventions. *Epilepsy & Behavior* **8**, 102–114.
- Kaltiala-Heino R, Kosunen E, Rimpela M** (2003). Pubertal timing, sexual behaviour and self-reported depression in middle adolescence. *Journal of Adolescence* **26**, 531–545.
- Kessler RC, McGonagle KA, Swartz M, Blazer DG, Nelson CB** (1993). Sex and depression in the National Comorbidity Survey. I: lifetime prevalence, chronicity and recurrence. *Journal of Affective Disorders* **29**, 85–96.
- Khan AA, Gardner CO, Prescott CA, Kendler KS** (2002). Gender differences in the symptoms of major depression in opposite-sex dizygotic twin pairs. *American Journal of Psychiatry* **159**, 1427–1429.
- Kovacs M** (2001). Gender and the course of major depressive disorder through adolescence in clinically referred youngsters. *Journal of the American Academy of Child and Adolescent Psychiatry* **40**, 1079–1085.
- Kuehner C** (2003). Gender differences in unipolar depression: an update of epidemiological findings and possible explanations. *Acta Psychiatrica Scandinavica* **108**, 163–174.
- Martin LA, Neighbors HW, Griffith DM** (2013). The experience of symptoms of depression in men vs women: analysis of the National Comorbidity Survey Replication. *JAMA Psychiatry* **70**, 1100–1106.
- Masi G, Favilla L, Mucci M, Poli P, Romano R** (2001). Depressive symptoms in children and adolescents with dysthymic disorder. *Psychopathology* **34**, 29–35.
- Myers K, Winters NC** (2002). Ten-year review of rating scales. II: scales for internalizing disorders. *Journal of the American Academy of Child and Adolescent Psychiatry* **41**, 634–659.
- Naninck EF, Lucassen PJ, Bakker J** (2011). Sex differences in adolescent depression: do sex hormones determine vulnerability? *Journal of Neuroendocrinology* **23**, 383–392.
- Newmann JP** (1986). Gender, life strains, and depression. *Journal of Health and Social Behavior* **27**, 161–178.
- Osman A, Barrios FX, Gutierrez PM, Williams JE, Bailey J** (2008). Psychometric properties of the Beck Depression Inventory-II in nonclinical adolescent samples. *Journal of Clinical Psychology* **64**, 83–102.
- Ramos E, Barros H** (2007). Family and school determinants of overweight in 13-year-old Portuguese adolescents. *Acta Paediatrica* **96**, 281–286.
- Rice SM, Fallon BJ, Aucote HM, Moller-Leimkuhler AM** (2013). Development and preliminary validation of the male depression risk scale: furthering the assessment of depression in men. *Journal of Affective Disorders* **151**, 950–958.
- Roberts RE, Lewinsohn PM, Seeley JR** (1991). Screening for adolescent depression: a comparison of depression scales. *Journal of the American Academy of Child and Adolescent Psychiatry* **30**, 58–66.
- Roberts RE, Lewinsohn PM, Seeley JR** (1995). Symptoms of DSM-III-R major depression in adolescence: evidence from an epidemiological survey. *Journal of the American Academy of Child and Adolescent Psychiatry* **34**, 1608–1617.
- Rochlen AB, Paterniti DA, Epstein RM, Duberstein P, Willeford L, Kravitz RL** (2010). Barriers in diagnosing and treating men with depression: a focus group report. *American Journal of Men's Health* **4**, 167–175.
- Romans SE, Tyas J, Cohen MM, Silverstone T** (2007). Gender differences in the symptoms of major depressive disorder. *Journal of Nervous and Mental Disease* **195**, 905–911.
- Rutz W, Walinder J, Von Knorring L, Rihmer Z, Pihlgren H** (1997). Prevention of depression and suicide by education and medication: impact on male suicidality. An update from the Gotland study. *International Journal of Psychiatry in Clinical Practice* **1**, 39–46.
- Salokangas RK, Vaahtera K, Pacriev S, Sohlman B, Lehtinen V** (2002). Gender differences in depressive symptoms. An artefact caused by measurement instruments? *Journal of Affective Disorders* **68**, 215–220.
- Santor DA, Ramsay JO, Zuroff DC** (1994). Nonparametric item analyses of the Beck Depression Inventory: evaluating gender item bias and response option weights. *Psychological Assessment* **6**, 255–270.
- Siegel JM, Yancey AK, Aneshensel CS, Schuler R** (1999). Body image, perceived pubertal timing, and adolescent mental health. *Journal of Adolescent Health* **25**, 155–165.
- Stehouwer RS, Bultsma CA, Blackford IT** (1985). Developmental differences in depression: cognitive-perceptual distortion in adolescent versus adult female depressives. *Adolescence* **20**, 291–299.
- Stockings E, Degenhardt L, Lee YY, Mihalopoulos C, Liu A, Hobbs M, Patton G** (2015). Symptom screening scales for detecting major depressive disorder in children and adolescents: a systematic review and meta-analysis of reliability, validity and diagnostic utility. *Journal of Affective Disorders* **174**, 447–463.
- Swaminathan H, Rogers HJ** (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement* **27**, 361–370.
- Thapar A, Collishaw S, Pine DS, Thapar AK** (2012). Depression in adolescence. *Lancet* **379**, 1056–1067.
- van Beek Y, Hessen DJ, Hutteman R, Verhulp EE, van Leuven M** (2012). Age and gender differences in depression across adolescence: real or 'bias'? *Journal of Child Psychology and Psychiatry and Allied Disciplines* **53**, 973–985.
- Wang YP, Gorenstein C** (2013). Psychometric properties of the Beck Depression Inventory-II: a comprehensive review. *Revista Brasileira de Psiquiatria* **35**, 416–431.
- Weller EB, Kloos A, Kang J, Weller RA** (2006). Depression in children and adolescents: does gender make a difference? *Current Psychiatry Reports* **8**, 108–114.

Williams DG, Morris GH (1996). Crying, weeping or tearfulness in British and Israeli adults. *British Journal of Psychology* **87**(Pt 3), 479–505.

Young MA, Scheftner WA, Fawcett J, Klerman GL (1990). Gender differences in the clinical features of unipolar major

depressive disorder. *Journal of Nervous and Mental Disease* **178**, 200–203.

Zierau F, Bille A, Rutz W, Bech P (2002). The Gotland Male Depression Scale: a validity study in patients with alcohol use disorder. *Nordic Journal of Psychiatry* **56**, 265–271.