Gender differences in changes in and stability of conduct problems from early adolescence to early adulthood

Elisabet E. Storvolla,*, Lars Wichstrømb

a NOVA-Norwegian Social Research, P.O. Box 3223, Elisenberg, Oslo 0208, Norway
b Norwegian University of Science and Technology, Norway

Abstract

The present paper examines whether there are gender differences in changes in the aggregate level and individual stability of conduct problems from early adolescence to early adulthood. We analysed self-reported data collected as part of a longitudinal general population study of Norwegian adolescents (n = 1234). The analyses showed that involvement in conduct problems for both genders peaked in mid-adolescence. The extent to which this peak reflected changes in the number of participants or differences in the frequency of problems among the participants varied according to gender. In most cases, the individual stability of involvement in conduct problems was similar for girls and boys. However, in the long term, relative involvement was more stable for boys than for girls. Moreover, the proportion of boys who showed a high level of involvement both during adolescence and early adulthood was greater than for girls.

Keywords: Conduct problems; Antisocial behavior; Stability; Change; Gender differences; Adolescence

1. Introduction

Several researchers have noted that there is a lack of knowledge about conduct problems among girls (e.g. Zoccolillo, 1993; Loeb & Stouthamer-Loeber, 1998; Maughan & Rutter, 2001). The present paper aims to provide such knowledge by comparing girls and boys with regard to changes in and stability of conduct problems from early adolescence to early adulthood. Conduct problems were investigated at both the aggregate (group) level and the individual level.
Aggregate measures of change and individual stability are complementary aspects in the description of the developmental course and prognostic significance of such problems (Gersten, Langner, Eisenberg, Simcha-Fagan, & McCarthy, 1976). Aggregate measures of change provide information about how the level of conduct problems diminishes or increases according to age. In other words, do adolescents tend to grow out of such problems or do more problems develop? The stability of relative involvement indicates the extent to which individuals preserve their rank ordering within a group, whereas the stability of absolute involvement indicates the extent to which individuals who have a high level of conduct problems at one time also have a high level later. Both types of stability provide information about the predictability of later problems from current problems.

1.1. Gender differences in changes in the aggregate level of conduct problems

Previous studies have shown that the level of conduct problems increases to a peak somewhere between the ages of 14 and 17 and then decreases (e.g. Farrington, 1986; Moffitt, 1993; Smith, 1995; Thornberry, 1997). Moreover, there is general agreement between researchers that boys have a higher level of conduct problems than girls have (e.g. Viale-Val & Sylvester, 1993; Smith, 1995; Rutter, Giller, & Hagell, 1998; Lahey et al., 2000). However, there has been little interest in whether girls and boys follow the same pattern of involvement over time, or whether their conduct problems peak at different ages. A longitudinal study addressing this question found that self-reported delinquency peaked at the same age for boys and girls (Windle, 2000). However, a cross-sectional study found that self-reported delinquency peaked at age 16 for girls and at age 17 for boys (Pedersen & Wichstrøm, 1995). A study of registered criminality also found that such problems peaked somewhat earlier for girls than for boys (Rutter et al., 1998).

Researchers have argued that the peak of conduct problems in mid-adolescence reflects a temporary increase in participation, i.e. the number of people involved in conduct problems, rather than a temporary increase in frequency, i.e. the frequency of such behaviour among the involved individuals (Farrington, 1986; Moffitt, 1993). However, there is reason to believe that this finding is culture-specific. A comparative study of English and Swedish boys concluded that participation accounted for most of the age differences in aggregate crime rates in London, while age differences in Stockholm largely reflected differences in frequency (Farrington & Wikström, 1994). A recent study of Norwegian girls and boys found that the age-related increase in antisocial behaviour in early adolescence (age 11–14) reflected both participation and frequency (Bendixen & Olweus, 2001).

1.2. Gender differences in individual stability of conduct problems

Researchers agree that there is high individual stability of involvement in conduct problems over time (e.g. Farrington, 1995; Smith, 1995; Loeber & Coie, 2001; Maughan & Rutter, 2001). However, there has been little interest in whether the stability of such behaviour is different for girls and boys. On the one hand, one could expect that conduct problems are less stable for girls than for boys. Conduct problems are considered to be more inappropriate for girls than for boys (Heimer, 1996). Thus socializing agents may react earlier to girls’ deviant behaviour than to boys deviant behaviour and may control girls more than their male counterparts. This makes it more
likely for girls to reduce their level of involvement or desist from such behaviour. On the other hand, one could expect that conduct problems are more stable for girls than for boys. The inappropriateness of girls’ deviant behaviour may make socializing agents more likely to react to girls’ deviant behaviour in more extreme and stigmatizing ways. The result could be a stronger selection into norm breaking and possibly also more male-dominated peer networks. This could increase the risk for girls to continue to be involved in deviant behaviour. Moreover, researchers have argued that girls who have conduct problems have faced more risk factors than their male counterparts (Robins, 1966), and thus they are a more selected group. This may also put girls at greater risk of continuing their involvement.

Most studies of the stability of relative involvement (i.e. correlation coefficients) in conduct problems have been conducted among boys (e.g. Simons, Johnson, Conger, & Elder, 1998; Stattin & Magnusson, 1991). The few studies that have considered gender differences in this respect have found the stability of relative involvement to be equal for boys and girls (Ferdinand & Verhulst, 1995; Brook, Whitman, Finch, & Cohen, 1996). In addition, some researchers have considered gender differences in the relative stability of the aggressive aspects of conduct problems. Whereas some of these studies have found that the stability of aggressive behaviour is similar for girls and boys (Moskowitz, Schwartzman, & Ledingham, 1985; Pulkki, 1992), others have found that the stability is lower for girls than for boys (Olweus, 1981; Huesmann, Eron, Lefkowitz, & Walder, 1984). However, since the overt and aggressive aspects of conduct problems are probably seen as more inappropriate for girls than the more covert and less aggressive aspects, it is difficult to compare the results from the latter studies with results from studies that have used more general definitions of conduct problems.

According to Loeber and Stouthamer-Loeber (1998), correlation coefficients are poor indicators of the actual degree of stability. They only reflect the degree to which the rank ordering of individuals on the basis of their antisocial behaviour at one assessment is similar to that at another assessment. The fact that correlations may be high, even though the absolute level of involvement changes dramatically over time, is often overlooked. Loeber and Stouthamer-Loeber argue that it is more useful to know what the probability is that highly antisocial individuals will persist or desist in their behaviour, and what the probability is that less antisocial individuals will persist in their behaviour or that the level of their behaviour will escalate. Thus, stability can be best understood by using measurements of levels of antisocial behaviour that are comparable over time. Several studies have examined persistence of conduct problems among boys and their desistance from conduct problems (e.g. Loeber, Stouthamer-Loeber, Van Kammen, & Farrington, 1991; Mischkowitz, 1994). As far as we know no researchers have considered such questions in studies of conduct problems among girls.

Previous researchers have found antisocial behaviour to be highly stable in a relatively small number of individuals whose behavioural problems are quite severe (e.g. Loeber, 1982; Stattin & Magnusson, 1991; Moffitt, Caspi, Dickson, Silva, & Stanton, 1996). Moffitt (1993) distinguished this group, which she called the life-course-persistent (LCP) group, from the adolescence-limited (AL) group. The first group engages in antisocial behaviour of one sort or another at every life stage, whereas the second is antisocial only during adolescence. The male-to-female ratio is expected to be more pronounced for the LCP group than for the AL group (Moffitt, 1994; Moffitt & Caspi, 2001). This has been confirmed in the few empirical studies that have addressed this question (Kratter & Hogan, 1999; Moffitt & Caspi, 2001).
1.3. Aims of the study

The aim of the study was to examine gender differences in changes in the aggregate level of conduct problems over time, using the same sample that Pedersen and Wichstrøm (1995) conducted cross-sectional analyses on. Thus, we were able to examine to what extent changes within subjects follow the same pattern as changes between subjects. Moreover, we investigated whether changes in involvement over time reflect changes in participation, frequency, or both.

Three strategies were used to determine whether there are gender differences in the individual stability of conduct problems: first, comparison of girls and boys with respect to the stability of relative involvement over time: second, comparison of girls and boys with respect to the probability of showing a high level of involvement in mid-adolescence and early adulthood, if they showed a high level of involvement rather than a low level or no involvement at earlier assessments: third, comparison of the proportion of girls and boys who showed a high level of involvement both during adolescence and early adulthood with the proportion of girls and boys who showed a high level of involvement only in adolescence.

2. Method

2.1. Participants and procedure

Since the purpose of the present study was to examine involvement in conduct problems from early adolescence to early adulthood, we examined adolescents who were 13 or 14 years old at the first assessment (1992) in a longitudinal study of Norwegian adolescents. These adolescents were 15 or 16 years old at the second assessment (1994) and 20 or 21 years old at the third assessment (1999) (n = 1378). Only adolescents who had answered the questions about conduct problems at all three assessments were examined (n = 1234). There were 655 (53.1%) girls and 579 (46.9%) boys in this sub-sample.

These adolescents constituted part of the Young in Norway Study. The initial sample of this study comprised 12287 students from 67 schools in grades 7–12 (12–20-year-olds). At t1 the students filled out a self-administered questionnaire at school. The response rate was 97.0%. The selection of the sample and the procedure at t1 have been described in more detail by Wichstrøm, Skogen, and Øia (1996). Three of the schools that participated in the study at t1 were included in connection with an evaluation study, and were not a part of the follow-up. In addition, one of the schools included in the sample at t1 was excluded from the longitudinal study because of a burglary in their archives, which resulted in the loss of the project’s ID-record. Thus, students from 63 schools were followed up at t2 (n = 9679). At t2 a sizeable proportion of the adolescents had completed the third grade at junior or senior high school, and had left the school they had been attending at t1. These students received a questionnaire by mail. Students who were still attending their original school filled out the questionnaire at school. The response rate in the latter group was 91.8%. Only adolescents who completed the questionnaires at school and consented to be followed up at later assessments (91.2%) were followed up at t3 (n = 3507), due to the comparatively lower response rate among those who had received the questionnaire by mail.
all the adolescents received the questionnaire by mail. Data were obtained from 83.8% \((n = 2924)\). Thus, the overall response rate was 68.1%.

Logistic regression analysis identified the following predictors of \(t2\) attrition: gender (boy), conduct problems, low parental SES, poor grades, few hours spent on homework, low parental monitoring, urbanity, and vocational training. By entering these variables, 65% of the attrition group was classified correctly (Wichström & Pedersen, 2001). The following variables measured at \(t1\) were identified as predictors of \(t3\) attrition: gender (boy), age, conduct problems, poor grades, few hours spent on homework, urbanity, and attrition at \(t2\). By entering these variables, 91% of the attrition group was classified correctly.

2.2. Instruments

The questions about conduct problems constituted part of a large questionnaire that covered several aspects of the adolescents’ lives. Conduct problems were measured using 11 items, ranging from behaviours that are most likely not problematic when low in frequency to behaviours that are serious even when infrequent. These items are presented in Appendix A. The behaviours were selected from Olweus’ scale of antisocial behaviour (Olweus, 1989) and the National Youth Longitudinal Study in the USA (Windle, 1990). Subjects reported the frequency of their involvement in these behaviours during the previous 12 months. Response options were: never (0), once (1), 2–5 times (2), 6–10 times (3), 10–50 times (4), and more than 50 times (5). A general measure of conduct problems with a possible range from 0 to 55 was computed from these questions. Those who reported that they had been involved in at least one of these behaviours were defined as participants. The cut-off for a high level of involvement in conduct problems was a score of seven or more on this scale. The rationale for selecting this cut-off point was that it constituted the 90th percentile at the first assessment. This cut-off point was reached by 11.1% of the respondents at \(t1\), 19.9% at \(t2\), and 10.9% at \(t3\). Those who were defined as participants, but who did not reach the cut-off for a high level of involvement, were described as having a low level of involvement.

In connection with analyses of relative stability, conduct problems were divided into three sub-dimensions: non-destructive covert, destructive covert, and overt. The first dimension reflects a tendency to avoid arenas under adult control (items 1 and 2 in Appendix A). The second dimension includes different kinds of theft and vandalism (items 3–8). The third dimension includes fighting with a weapon and school-related conduct problems of an overt aggressive kind (items 9–11). This subdivision of items was based on earlier analyses of the dimensionality of conduct problems in the same sample. First, these three dimensions were revealed with exploratory factor analyses (Pedersen & Wichström, 1995). Second, confirmatory factor analyses found these three dimensions to describe the data better than either a two-factor model or a single-syndrome model (Storvoll, Wichström, Kolstad, & Pape, 2002). The three sub-dimensions fitted well for both girls and boys and students at both junior and senior high schools (Storvoll et al., 2002).

2.3. Statistical analyses

Independent sample \(t\)-tests were used to test for group differences in involvement in the continuous measures of conduct problems. Paired-sample \(t\)-tests were used to test for a group’s
change in involvement between two assessments. Pearson’s \( \chi^2 \) test of independence was used to test for group differences in the categorical measures of conduct problems. The McNemar test, which is a non-parametric test for two related dichotomous variables, was used to test whether the proportions of respondents who were involved at two different times were different or not.

Structural equation models in LISREL 8.30 (Jöreskog & Sörbom, 1993) were used to estimate the stability of relative involvement in conduct problems over time, and to test whether there were gender differences in this respect. The covariance matrix used as input in these analyses was obtained using the computer program PRELIS 2 (Jöreskog & Sörbom, 1996), and the models were estimated using the maximum likelihood method. Conduct problems were seen as a latent construct measured by the three sum-indexes non-destructive covert, destructive covert, and overt. Use of sum-indexes rather than individual items is recommended in such analyses (Kishton & Widaman, 1994). Since different aspects of conduct problems change in importance over time, we allowed the factor loadings to vary over time. In addition, the error terms of the observed variables were allowed to covary over time. By doing this, the contribution of shared method variance could be removed from the estimates of stability. Even though the observed variables were log transformed to reduce the skewness and kurtosis of the distributions, the distributions still deviated from the normal distribution. However, visual inspection of the plots of the standardized residuals for the models showed that they did not deviate much from normality. Several tests are available to evaluate the fit of estimated models in LISREL. The most common test is the minimum fit function chi-square (\( \chi^2 \)). However, this test of significance has several problems when used for large samples, where it is likely to deem the fit unacceptable even though the residual matrix is small (Van de Vijver & Leung, 1997). Thus, in addition to the minimum fit function chi-square, we used the root mean square error of approximation (RMSEA), which is a measure of discrepancy per degree of freedom. The critical level for RMSEA is 0.050 for a good fit and 0.080 for an acceptable fit.

Logistic regression analyses were used to examine the probability of having a high level of conduct problems in mid-adolescence and early adulthood for those who had a high level rather than a low level or no involvement at earlier assessments. Girls and boys were analyzed separately. To test whether there were gender differences in the association between conduct problems at different assessments, we used a \( t \)-test developed to test for differences in unstandardized regression coefficients in two independent samples (Paternoster, Brame, Mazerolle, & Piquero, 1998).\(^1\) In order to correct for a high number of comparisons, a one per cent level of significance was applied in all the analyses.

3. Results

3.1. Changes in the aggregate level of conduct problems over time

Both for boys and for girls, involvement in conduct problems was greater at \( t_2 \) (age 15/16) than at \( t_1 \) (age 13/14) and \( t_3 \) (age 20/21) (Table 1). For girls, but not for boys, involvement at \( t_3 \) was

\[ t = \frac{b_1 - b_j}{\sqrt{SE_{b_1}^2 + SE_{b_j}^2}}. \]
significantly greater than involvement at t1. Boys had a greater involvement than girls at all three assessments.

Further, we examined whether changes in involvement over time reflected participation or frequency. For both genders the increase from t1 to t2 reflected an increase in participation as well as frequency. For boys the reduction from t2 to t3 also reflected a reduction in both. For girls, however, the reduction in this period reflected only a reduction in frequency. Moreover, the increase in girls’ involvement from t1 to t3 reflected a higher level of participation at t3 than at t1. Thus, the higher number of girls involved in mid-adolescence seems to be maintained into early adulthood.

Moreover, we determined the proportion of adolescents who showed a high level of involvement in conduct problems. For both genders the proportion of respondents with a high level of involvement was higher at t2 than at t1 and t3. A greater proportion of boys than girls showed a high level of involvement at all three assessments.

Finally, the analyses presented in Table 1 were weighed for attrition at t2 and t3, respectively (not shown). In most cases these analyses indicated that the level of conduct problems was somewhat higher among those who dropped out during the study than among those who participated at all three assessments. However, the effect of attrition on the results was very small. The largest difference in mean values between the weighted and the unweighted data was less than 0.07 standard deviations and the largest difference in proportions of respondents was 3.1%. Moreover, the conclusions drawn about gender differences in involvement were similar in the
unweighted and weighted data. There was however one exception from this finding: The gender differences in the mean score for conduct problems at t3 disappeared in the weighted data.

3.2. Stability of relative involvement in conduct problems over time

The relative stability of conduct problems was examined using structural equation models. Table 2 presents the matrix applied as input in these models. To make the table easier to follow, the covariances were transformed into correlation coefficients. The table shows that the strongest correspondences were obtained within each sub-dimension over time and between the three sub-dimensions at each assessment. Moreover, the correlations were stronger from t1 to t2 than from t2 to t3 and from t1 to t3. This pattern was true for both genders, but the correlations seem to be somewhat stronger for boys than for girls.

First, we examined the stability between each of the three time periods. All three models had an acceptable fit for both genders (Table 3). For both genders, conduct problems were more stable from t1 to t2 than from t2 to t3 and from t1 to t3. We conducted a series of multi-sample analyses to test whether there were any statistically significant gender differences in stability over these time periods. When the measurement models were allowed to vary across gender, there were no such gender differences. However, when the factor loadings were invariant across gender, boys’ conduct problems were more stable than girls’ conduct problems from t1 to t3 ($\chi^2(1) = 10.84$, $p < 0.001$).

Second, we examined a model that contained all three measurement points (see Fig. 1). This model had an acceptable fit for both boys and girls. For both genders the paths from t1 to t2 and from t2 to t3 were statistically significant, and the first path was stronger than the second. Boys,

<table>
<thead>
<tr>
<th></th>
<th>t1</th>
<th>t2</th>
<th>t3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NC</td>
<td>DC</td>
<td>O</td>
</tr>
<tr>
<td>t1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC</td>
<td>1.00</td>
<td>0.48</td>
<td>0.45</td>
</tr>
<tr>
<td>DC</td>
<td>0.35</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>O</td>
<td>0.42</td>
<td>0.35</td>
<td>1.00</td>
</tr>
<tr>
<td>t2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC</td>
<td>0.48</td>
<td>0.27</td>
<td>0.37</td>
</tr>
<tr>
<td>DC</td>
<td>0.24</td>
<td>0.26</td>
<td>0.22</td>
</tr>
<tr>
<td>O</td>
<td>0.35</td>
<td>0.25</td>
<td>0.60</td>
</tr>
<tr>
<td>t3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC</td>
<td>0.14</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>DC</td>
<td>0.07</td>
<td>0.22</td>
<td>0.11</td>
</tr>
<tr>
<td>O</td>
<td>0.07</td>
<td>0.05</td>
<td>0.17</td>
</tr>
<tr>
<td>S.D.</td>
<td>Boys</td>
<td>0.42</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>Girls</td>
<td>0.36</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Boys are presented in the upper diagonal ($n = 579$) and girls are presented in the lower diagonal ($n = 655$). Note: NC = non-destructive covert; DC = destructive covert; O = overt.
but not girls, showed a statistically significant path from $t_1$ to $t_3$. A multi-sample analysis showed that this path was significantly stronger for boys than for girls ($\chi^2(1) = 8.27, p < 0.01$). The same conclusion was drawn when the factor loadings were invariant across gender. Thus, boys more often than girls seemed to preserve their rank ordering within their group from early adolescence to early adulthood. There were, however, no gender differences in the relative stability from early adolescence to mid-adolescence and from mid-adolescence to early adulthood.

Table 3
Boys' ($n = 579$) and girls' ($n = 655$) relative stability of conduct problems between three time periods

<table>
<thead>
<tr>
<th></th>
<th>Unstandardized estimate</th>
<th>Standardized estimate</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$p$</th>
<th>RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_1-t_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boys</td>
<td>0.84**</td>
<td>0.67</td>
<td>14.66</td>
<td>5</td>
<td>0.01</td>
<td>0.058</td>
</tr>
<tr>
<td>Girls</td>
<td>1.02**</td>
<td>0.79</td>
<td>8.09</td>
<td>5</td>
<td>0.15</td>
<td>0.031</td>
</tr>
<tr>
<td>$t_2-t_3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boys</td>
<td>0.38**</td>
<td>0.54</td>
<td>5.51</td>
<td>5</td>
<td>0.36</td>
<td>0.013</td>
</tr>
<tr>
<td>Girls</td>
<td>0.24**</td>
<td>0.37</td>
<td>4.21</td>
<td>5</td>
<td>0.52</td>
<td>0.000</td>
</tr>
<tr>
<td>$t_1-t_3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boys</td>
<td>0.46**</td>
<td>0.56</td>
<td>6.37</td>
<td>5</td>
<td>0.27</td>
<td>0.022</td>
</tr>
<tr>
<td>Girls</td>
<td>0.24*</td>
<td>0.26</td>
<td>1.57</td>
<td>5</td>
<td>0.90</td>
<td>0.000</td>
</tr>
</tbody>
</table>

*p < 0.01; **p < 0.001.

Fig. 1. Stability of conduct problems between three time periods. Girls’ values are presented in parentheses. Boys: $\chi^2(15, n = 579) = 26.38, p = 0.03; RMSEA = 0.036$. Girls: $\chi^2(15, n = 655) = 15.55, p = 0.41; RMSEA = 0.008$. 

Note: CP = conduct problems; NC = non-destructive covert; DC = destructive covert; O = overt.
*p < 0.01; **p < 0.001.
3.3. Stability of absolute involvement in conduct problems over time

Table 4 presents analyses of the probability of having a high level of conduct problems in mid-adolescence and early adulthood for those who showed a high level of involvement rather than a low level or no involvement at earlier assessments. It presents four different regression models. The first model indicates that the probability of having a high level of conduct problems at \( t_2 \) was almost similar for the boys (12.82) and the girls (12.93) who had a high level at \( t_1 \) compared to those who had a low level or no involvement. The second model indicates that the probability of having a high level of conduct problems at \( t_3 \) was somewhat greater for the boys (5.69) than for the girls (4.40) who had a high level at \( t_2 \) compared to those who had a low level or no involvement. However, this gender difference was not statistically significant. The third model indicates that the probability of having a high level of conduct problems at \( t_3 \) was almost similar for the boys (3.13) and the girls (3.38) who had a high level at \( t_1 \) compared to those who had a low level or no involvement.

In the fourth model, the possible combinations of involvement in conduct problems at the first two assessments predict the level of involvement at \( t_3 \). A low level or no involvement at both \( t_1 \) and \( t_2 \) was selected as a contrast. For both boys (6.69) and girls (5.76) the probability of showing a high level of involvement at \( t_3 \) was strongest for those who showed a high level at both \( t_1 \) and \( t_2 \). Boys (5.17) and girls (4.16) who showed a low level or no involvement at \( t_1 \) and a high level at \( t_2 \) showed the second highest probability. Even though the probability of showing a high level at \( t_3 \) seemed to be somewhat stronger for boys than for girls in both these groups, there were no statistically significant gender differences in this respect. In summary, there were no gender differences in the probability of having a high level of conduct problems in mid-adolescence and early adulthood for those who showed a high level of involvement rather than a low level or no involvement at earlier assessments.

Finally, we compared boys’ and girls’ patterns of involvement in conduct problems over time. Table 5 shows the five different patterns the respondents were allowed to follow. The first pattern includes respondents who showed a high level of involvement at all three assessments; the second

<table>
<thead>
<tr>
<th>Boys’ ( n = 579 ) and girls’ ( n = 655 ) absolute stability of conduct problems between three time periods</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td><strong>t1-t2</strong></td>
</tr>
<tr>
<td>2.55</td>
</tr>
<tr>
<td><strong>t2-t3</strong></td>
</tr>
<tr>
<td>1.74**</td>
</tr>
<tr>
<td><strong>t1-t3</strong></td>
</tr>
<tr>
<td>1.14**</td>
</tr>
<tr>
<td><strong>t1/t2-t3</strong></td>
</tr>
<tr>
<td>t1 = low/no, t2 = high</td>
</tr>
<tr>
<td>t1 = high, t2 = low/no</td>
</tr>
<tr>
<td>t1 = high, t2 = high</td>
</tr>
</tbody>
</table>

* *p < 0.01; **p < 0.001.
pattern includes respondents who showed a high level once in adolescence and in adulthood; the third pattern includes respondents who showed a high level one or two times in adolescence; the fourth pattern includes respondents who showed a high level only in adulthood; and the fifth pattern includes respondents who showed a low level or no involvement at all three assessments. As can be seen from the table, the majority of both genders showed a low level or no involvement at all three assessments (boys: 65.1%; girls 77.4%). The second most common pattern was a high level of involvement only in adolescence (boys: 20.4%; girls 14.8%). The least frequent pattern was a high level of involvement at all three assessments (boys: 4.1%; girls 1.1%).

A $\chi^2$ test showed that the proportion of respondents who followed the different patterns was significantly different for boys and girls ($\chi^2(4) = 29.15, p<0.001$). Further analyses indicated that this gender difference reflected the fact that a larger proportion of the girls than of the boys showed a low level or no involvement at all three assessments ($\chi^2(1) = 22.86, p<0.001$). In addition, a larger proportion of the boys than of the girls showed a high level of involvement at all three assessments ($\chi^2(1) = 11.88, p<0.01$). There were no gender differences in the proportion of those who showed a high level only in adolescence.

### Table 5
Comparison of boys' ($n = 579$) and girls' ($n = 655$) pattern of involvement in conduct problems at three time periods

<table>
<thead>
<tr>
<th></th>
<th>Boys</th>
<th>Girls</th>
<th>B:G ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Obt. %</td>
<td>Exp.</td>
<td>Obt. %</td>
</tr>
<tr>
<td>1. High at all three assessments</td>
<td>24  4.1</td>
<td>14.5</td>
<td>7  1.1</td>
</tr>
<tr>
<td>2. High once in adolescence and in adulthood</td>
<td>27  4.7</td>
<td>20.2</td>
<td>16  2.4</td>
</tr>
<tr>
<td>3. High only in adolescence</td>
<td>118 20.4</td>
<td>100.9</td>
<td>97 14.8</td>
</tr>
<tr>
<td>4. High only in adulthood</td>
<td>33  5.7</td>
<td>28.6</td>
<td>28  4.3</td>
</tr>
<tr>
<td>5. Low at all three assessments</td>
<td>377 65.1</td>
<td>414.8</td>
<td>507 77.4</td>
</tr>
</tbody>
</table>

Note: $\chi^2 (4) = 29.15, p<0.001$.

4. Discussion

The present paper examined gender differences in changes in the aggregate level and individual stability of conduct problems from early adolescence to early adulthood. Both girls and boys had a peak of involvement in conduct problems in mid-adolescence. To what extent this peak reflected participation, frequency, or both varied according to gender. In most cases, the individual stability of involvement in conduct problems was similar for girls and boys. However, the relative stability from early adolescence to early adulthood was stronger for boys than for girls. Moreover, a larger proportion of the boys than of the girls showed a high level of involvement both during adolescence and early adulthood.

4.1. Changes in the aggregate level of conduct problems

In line with cross-sectional analyses of the same sample as the one analysed in the present paper (Pedersen & Wichstrøm, 1995), both genders showed a peak of involvement in conduct problems
in mid-adolescence. However, the time spans between the data collections in the longitudinal study were too long to address whether there are gender differences in the exact peak of such behaviour. Pedersen and Wichstrøms' (1995) cross-sectional study indicates that girls’ conduct problems peak at age 16, whereas boys' conduct problems peak at age 17. This is in line with a study of registered criminality (Rutter et al., 1998), but not in line with a longitudinal self-report study that found that delinquency peaked at the same age for boys and girls (Windle, 2000). Thus, longitudinal studies are still needed in order to determine whether or not conduct problems among girls peak somewhat earlier than conduct problems among boys.

Another question for further studies is whether different sub-dimensions of conduct problems peak at different ages. The cross-sectional analyses presented by Pedersen and Wichstrøm (1995) indicate that this is the case. Whereas overt and destructive covert conduct problems peaked in mid-adolescence, non-destructive covert behaviours continued to increase until age 18. However, longitudinal studies with more frequent data collections than those in Young in Norway are needed to enlighten this question in greater detail, and to study whether there are gender differences in this respect.

Whereas some researchers report that the increased involvement in mid-adolescence reflects a temporary increase in participation rather than frequency (Farrington, 1986; Moffitt, 1993), others suggest that this finding may be culture-specific (Farrington & Wikström, 1994). The present paper indicates that the increase in involvement from early to mid-adolescence for both genders reflects participation as well as frequency. This is in line with the findings from a recent Norwegian study of young adolescents (Bendixen & Olweus, 2001). For boys the reduction from mid-adolescence to early adulthood reflected a decrease in both participation and frequency. For girls, however, this reduction reflected only a reduction in frequency. Consequently, for girls the increased number of participants seems to be maintained into early adulthood. Thus, whether changes in involvement in conduct problems over time reflect participation, frequency, or both, does not only seem to be culture-specific but also to be gender-specific. Further studies are however needed to draw firm conclusions about such gender differences. Moreover, further studies should address the factors that contribute to young people’s changing involvement in conduct problems over time, and whether the processes involved are gender specific or not.

4.2. Individual stability of conduct problems

In line with earlier studies (Brook et al., 1996; Ferdinand & Verhulst, 1995) the present study found the relative stability of conduct problems from early to mid-adolescence and from mid-adolescence to early adulthood to be similar for boys and girls. However, boys more often than girls seemed to preserve their rank ordering over a long time, i.e. from early adolescence to early adulthood. This was found in both the bivariate analyses and in the model that estimated the stability between all three assessments. In the latter analyses boys but not girls showed a direct effect from conduct problems in early adolescence to conduct problems in early adulthood. One possible explanation of this finding may be that some boys with a relatively short period of high involvement in conduct problems in mid-adolescence, when the involvement peaked, probably had a higher level of conduct problems than the boys who were involved in such problems over a longer period. That girls did not show a direct effect from early adolescence to early adulthood probably reflects that the girls who showed increased involvement in mid-adolescence were still
less involved than the girls who had such problems over a longer period. Thus, whether there are gender differences in the stability of relative involvement in conduct problems or not seems to depend on the time period considered. Moreover, the conclusions drawn from such studies may be influenced by the instruments applied. Our findings are in accordance with the results from other studies applying general definitions of conduct problems (Brook et al., 1996; Ferdinand & Verhulst, 1995). However, some publications considering more aggressive behaviour conclude that the stability of such behaviour is stronger for boys than for girls (Huesmann et al., 1984; Olweus, 1981). Further studies are needed to enlighten whether there are more pronounced gender differences in the stability of behaviour when considering overt and aggressive aspects of conduct problems rather than more covert and less aggressive aspects of such behaviour.

Some researchers have argued that it is more useful to study the stability of absolute involvement in conduct problems (Loeber & Stouthamer-Loeber, 1998). When applying such a strategy we found no male/female differences in the probability of having a high level of conduct problems in mid-adolescence and early adulthood for those who had a high level of involvement rather than a low level or no involvement at earlier assessments. Thus, although boys have a far greater probability than girls of being involved in serious conduct problems at all times, girls who have such problems at any time during adolescence have just as high a probability as boys of being involved in serious conduct problems later. As far as we know no other studies have considered whether there are gender differences in this respect. Such studies are needed to conclude whether there are gender differences in the absolute stability of conduct problems or not.

Finally, we considered gender differences in the proportion that followed different patterns of involvement in conduct problems over time. The respondents who had a high level of conduct problems at all three assessments probably represent the LCP group presented by Moffitt (1993), whereas the respondents who showed a high level of involvement only in adolescence probably represent the AL group. However, Moffitt argues that information about pre-adolescent behaviour is needed in order to differentiate between the sub-groups presented in her taxonomy. On the other hand, the fact that the first group still had a high level of conduct problems in early adulthood indicates that their conduct problems are not limited to adolescence. If, in spite of the lack of information about the adolescents’ pre-adolescent behaviour, we compare these groups with those presented in Moffitt’s taxonomy, we find the male-to-female ratio to be higher for the LCP group (3.7) than for the AL group (1.4). This is both in accordance with Moffitt’s taxonomy (Moffitt, 1994; Moffitt and Caspi, 2001) and earlier empirical findings (Kratzer & Hodgins, 1999; Moffitt & Caspi, 2001).

Other researchers, in addition to Moffitt (1993), have argued that it is necessary to differentiate antisocial behaviour according to the age of onset (e.g. Stattin & Magnusson, 1995; Patterson & Yoerger, 1997; Thornberry & Krohn, 2001). Another promising way of sub-dividing adolescents with conduct problems has been presented by Loeber and his colleagues (e.g. Loeber, 1996; Loeber, Keenan, & Zhang, 1997). They consider progression along three different pathways, from more frequent milder antisocial behaviour to less frequent but more severe criminal behaviour. Both these methods of sub-dividing antisocial adolescents are dominated by studies of boys. An important question for further research is whether the different patterns described fit as well for girls as they do for boys.
4.3. Limitations

Although the present study has several strengths—the study followed a general population sample from early adolescence to early adulthood with limited attrition, the sample was large enough to compare girls and boys, and the measure of conduct problems that was used has adequate reliability—some limitations should be mentioned. First, the rate of conduct problems is probably higher for the adolescents who did not initially participate in the study than for those who did. Moreover, our analyses indicate that the level of conduct problems is somewhat higher for those who dropped out during the study (altogether 32%) than for those who participated at all three assessments. However, the effect of attrition on the level of involvement seemed to be small and, importantly, similar for girls and boys. Although the base rate of conduct problems may be higher in both these groups, it is difficult to see any reason why the pattern of gender differences in changes in the aggregate level and individual stability of conduct problems should be very different from that of the present sample.

Second, information about the adolescents’ involvement in conduct problems was based on self-reports. Information from reports by parents, teachers, or peers would probably have increased the validity. However, earlier studies have found a high correlation between youths’ self-reports and parents’ reports of conduct problems (Boyle et al., 1993). On the other hand, low correlations have been detected for parent–teacher and youth–teacher assessments (Boyle et al., 1993). Again, we cannot see any reason why this should have an impact on our findings concerning gender differences.

Third, the changeable manifestations of antisocial behaviour make studies that apply a narrow definition of such behaviour likely to artificially reduce the stability (Loeber, 1991). Our measure of conduct problems contained few items concerning overt and aggressive conduct problems. Furthermore, two of the three items measuring such behaviour concerned school opposition. This may have resulted in an underestimation of the stability of conduct problems. Moreover, as argued earlier, the overt aspects of conduct problems are probably seen as more inappropriate for girls than the covert aspects. It is difficult to know whether the lack of such items has influenced our findings about gender differences, and if so, in which direction.

4.4. Conclusion

The findings in the present paper indicate that both genders show a peak of involvement in conduct problems in mid-adolescence. The extent to which this peak reflects participation, frequency, or both varies for girls and boys. In most cases, the individual stability of conduct problems was similar for girls and boys. However, boys showed a greater stability of relative involvement than girls did in the long term, i.e. from early adolescence to early adulthood. Moreover, a greater proportion of boys than girls showed a high level of involvement in conduct problems both during adolescence and early adulthood.

Acknowledgements

This research was funded by the Research Council of Norway, Medicine and Health.
Appendix A. Conduct problem items

1. Played truant from school.2
2. Travelled by bus or train etc. or got into a cinema without paying.
3. Taken objects worth less than NOK 500 (approximately US$ 60) from a shop or kiosk without paying.3
4. Stolen something worth more than NOK 1000 (approximately US$ 110).
5. Stolen a car or a motorcycle.
6. Broken in order to steal.
7. Deliberately destroyed or broken windowpanes, bus seats, street lamps, post boxes, or similar.
8. Vandalized or caused damage to the value of more than NOK 1000.
9. Sworn in front of a teacher (see footnote 2).
10. Had a severe quarrel with a teacher (see footnote 2).
11. Fought with a weapon (e.g. knife).

References


---

2 At t3 the three questions concerning school-related conduct problems also included work-related conduct problems (e.g. played truant from school or work).

3 The question concerning theft of objects worth less than NOK 500 was changed after t1. At t2 and t3 the subjects were asked about theft of objects worth between NOK 100 and NOK 500.


