Adolescents’ proxy reports of parents’ socioeconomic status: How valid are they?

N Lien, C Friestad, K-I Klepp

Abstract

Study objective—Children’s proxy reports on indicators of their parents’ socioeconomic status (SES) have either been used uncritically or dismissed as invalid. This paper examines the validity of young adolescents’ reports of parental SES by comparing adolescent reports with parents’ own reports of SES.

Design, setting and participants—In 1990, 924 13 year olds, along with 648 of their fathers and 735 of their mothers, participated in the baseline survey of The Norwegian Longitudinal Health Behaviour Study. Data on parental occupation were collected from both adolescents and parents at baseline, while data on parental education were collected at follow up from the adolescents at age 15 and from their parents when the adolescents were 19 years old.

Main results—Three different ways of grouping the SES categories based on occupational data were investigated, and the strength of agreement was good for all three groups, with κ statistics ranging from 0.65 to 0.86. There were no significant improvements of agreement when comparing adolescent data from age 15 to adolescent data from age 13. The strength of agreements between the adolescents’ and parents’ reports of parental education were fair; κ statistics were 0.30 and 0.38 for fathers’ and mothers’ education, respectively. The proportions of unclassified answers or no responses from the adolescents were similar for questions on occupation and education, and ranged from 11% to 16%.

Conclusions—The agreement between adolescents’ and parents’ reports of SES based on occupation was judged to be good, but adding a few specific questions may cue the adolescent to provide more detailed information, thereby reducing the numbers of unclassified answers or non-responders.

Over the past two decades there has been a marked increase in research on the relation between socioeconomic status (SES) and health, and the focus has shifted from investigating the relation between health outcomes and SES to investigating the mechanisms behind such relations. Whether inquiring into the presence of a gradient or the mechanisms behind a SES-health gradient, the strength of the findings depends on the reliability and validity of the measures of both SES and health. In this article we will investigate the validity of measures of parental SES (based on occupation) and education obtained from adolescents through comparing the data to parents’ own reports.

The most commonly used measures of social class in epidemiological research are occupation, education and income. These three measures may be considered indicators of independent dimensions of social class and thus provide different pieces in the puzzle of describing the mechanisms behind the observed variation of health between social classes. At least two areas of research have acknowledged the difficulties in obtaining reliable and valid measures of the traditional indicators of social class, and have attempted to include others measures of social class. Researchers working in the developing world have proposed to include data on household possessions and other living conditions in the definition of SES, in addition to the more traditional measures of social class. Similarly, Currie et al argue that a “family affluence scale” would be useful when attempting to overcome the problem of a large proportion of missing data on parental occupation obtained from 11 year olds, 13 year olds and 15 year olds in the WHO-Survey of Health Behaviour of School-Aged Children (HBSC). The measures investigated by Currie et al included number of phones and number of cars owned by the family, the child’s amount of spending money and whether the child had her/his own bedroom. Although, these measures seem to be easily obtainable and may be useful for the above mentioned study, the problem of comparing results across time and across cultures may be even greater than the problem of comparing results based on the more traditional measures of social class. Thus, both Durkin et al. and Currie et al advise that measures of wealth should be included in addition to, not in place of, traditional measures of social class. This points to the importance of validating data obtained from children and adolescents regarding parental occupation and/or education.

The problem of non-response of children on questions concerning their parents’ education or occupation has been well documented in the early literature on proxy reports. However, no absolute level of non-response has been identified in order to determine when the proxy reports by children should be considered invalid. Furthermore, Looker suggests that if data could show that non-response by children and parents confines to the same families, it may be that the measure itself is problematic.
for those families, rather than that the children do not know the answer or do not understand the question. Looker8 concludes her extensive review of the literature pertaining to the accuracy of children’s proxy reports of parents’ occupation and education by stating that “it seems that proxy reports are most accurate if the respondents are high school seniors living at home and reporting on their parents’ current status characteristics.” This indicates that both age of the proxy respondents as well as closeness in space and time to the parental characteristic in question is important for the validity of the measure on parental occupation and education obtained from children. However, a recent study from Scotland found that by using a “mini-interview” when questioning 11 year olds about their parents’ economic and occupational activity, the proportion of missing data was negligible and the strength of agreement with parents own reports were very good. Thus, given the right methods and enough resources to collect data, obtaining valid data on parental SES from children younger than high school seniors may be possible.

In the Norwegian Longitudinal Health Behaviour (NLHB) Study, data on parental occupation and education have been collected from both parents and adolescents by means of questionnaires. In the present study, we investigate to what extent there is agreement between responses given by adolescents and their parents, whether the agreement improved as the adolescents grew older, and whether the measure of agreement was stronger for SES (based on occupation) than for education. In addition, three possible ways of grouping the occupational data as a measure of SES were explored.

**Methods**

In 1990, a sample of 1190 13 year olds in 22 schools in Hordaland county and their parents were invited to participate in the NLHB Study—a 10 year cohort study. The study focuses on the health behaviour, life style and self reported health of adolescents. Information was collected by means of questionnaires containing core questions repeated at every survey, as well as questions on special topics varying from survey to survey. During the first three years, adolescents completed (age 13–15) the questionnaires at school. Thereafter, participants received the questionnaires by mail. Parents were asked to complete surveys when their children were 13, 16 and 19 years old. This study was approved by the Norwegian Data Inspectorate, and it has been conducted in full accordance with ethical principles, including the provisions of the World Medical Association Declaration of Helsinki.

**Subjects**

Written consent from parents/guardians and the adolescents were collected before the first survey. At baseline, 924 of the adolescents participated (55% boys), a response rate of 77.6%. Of the 263 adolescents not participating, 222 could not be surveyed because their parents/guardians did not provide written consent, and 41 adolescents refused. Furthermore, three adolescents were excluded because of incomplete surveys at baseline. No differences in either gender or rural/urban residence were found between participants and non-participants. Thus, the baseline sample may be considered representative of the 1977 birth cohort attending ordinary school in the region. At baseline, 648 fathers and 735 mothers participated, representing 70% and 80% of the baseline sample of adolescents, respectively.

Data from the adolescent surveys in 1990 and 1992 (age 13 and 15) and data from the parent surveys in 1990 and 1996 (adolescents’ age 13 and 19) were used in the analyses for this article. Four cohorts were constructed to enable separate analyses for SES (based on occupation) and education for each parent (father/mother), and to account for the different types of attrition caused by non-participation or non-response/unclassifiable answers. The first and second cohort included pairs of adolescent-father (or mother) with adolescents who participated in 1990 and 1992, and fathers (cohort 1) or mothers (cohort 2) who participated in 1990. The third and fourth cohorts were constructed in a similar way for adolescent-parent pairs with adolescents who participated in 1992 and fathers (cohort 3) or mothers (cohort 4) who participated in 1996.

**MEASURES**

**Occupational status (1990 and 1992)**

In 1990, adolescents and their parents were asked the same open ended question on parental occupation; “What kind of work does your father/mother/you have? Try to describe as closely as you can what he/she/you is working as for example industrial worker, doctor, farmer, teacher, clerk.” The adolescents could leave the question blank if they did not know the answer. The same question was included in the 1992 adolescent survey. The answers were coded according to the “Nordic Occupational Classification”, and thereafter collapsed into nine SES groups and a group of unclassifiable answers according to the system developed by Statistics Norway. This system is based on three criteria; type of work (manual/non-manual), level of competence (mainly length of education) and amount of authority at work. In addition, a division is made between employees and self employed. The complete system has 33 categories, but categories can be collapsed in four systematic steps, which results in broader categories. Because of the simplicity of the question used to collect occupational data in this study, it was not possible to incorporate either the criteria of authority or the division between employees and self employed with the exception of farmers/fishermen who are considered self employed. Thus, the final 10 categories (see table 4) are a modified version of one of the intermediary steps in this system. During analyses, the 10 categories were further collapsed to overcome the problem of small groups, as well as to obtain a classification that
might be useful in further research on the relation between SES and health based on these data. The final six groups used in the analyses were manual (unskilled and skilled manual), lower non-manual, higher non-manual (middle and higher non-manual), farmers/fishermen, those not involved in economic activity (students, retired/on welfare, houseworkers), and those who could not be classified.

Dahl used the first six categories (see table 4) and a category of self employed, thus excluding those not involved in economic activity in his study of social inequalities in ill health. As we had no data on self employment, we applied the first six categories only in an attempt to copy this application of the SES system. In a study of dietary habits among Norwegian adults, Johansson et al applied a dichotomous version of the first five categories (see table 4); blue collar workers (the first three categories) and white collar workers (the next two categories), thus excluding farmers/fishermen in addition to those not involved in economic activity.

### Table 1 Differences of participants and non-participants of the educational cohort at age 13

<table>
<thead>
<tr>
<th>Socioeconomic status</th>
<th>Cohort 1 Fathers</th>
<th>Cohort 2 Mothers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Included in the cohort</td>
<td>67% (617)</td>
<td>76% (701)</td>
</tr>
<tr>
<td>Parent missing 1990</td>
<td>27% (246)</td>
<td>18% (162)</td>
</tr>
<tr>
<td>Adolescent missing 1992</td>
<td>3% (31)</td>
<td>4% (34)</td>
</tr>
<tr>
<td>Missing parent and adolescent 1992</td>
<td>3% (30)</td>
<td>3% (27)</td>
</tr>
<tr>
<td>Baseline number of adolescents</td>
<td>100% (924)</td>
<td>100% (924)</td>
</tr>
</tbody>
</table>

### Table 2 Differences of participants and non-participants of the cohort for socioeconomic status (SES based on occupation) according to adolescents' reports at age 13

<table>
<thead>
<tr>
<th>SES Status</th>
<th>Cohort 1 (n=617)</th>
<th>Cohort 2 (n=701)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual</td>
<td>29</td>
<td>8</td>
</tr>
<tr>
<td>Non-manual, low</td>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td>Non-manual, middle/high</td>
<td>43</td>
<td>48</td>
</tr>
<tr>
<td>Farmer/fisherman</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Non-workers</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>Unclassified</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

### Table 3 Differences of participants and non-participants of the educational cohorts according to adolescents' reports at age 15

<table>
<thead>
<tr>
<th>Education</th>
<th>Cohort 1 (n=466)</th>
<th>Cohort 2 (n=516)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elementary</td>
<td>16</td>
<td>22</td>
</tr>
<tr>
<td>Upper secondary</td>
<td>43</td>
<td>44</td>
</tr>
<tr>
<td>College/university</td>
<td>23</td>
<td>16</td>
</tr>
<tr>
<td>Other</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

### Education (1992 and 1996)

In 1992, the adolescents were asked about their parents' education; “How much education do you think your father/mother has?” To facilitate comparison with answers to a similar question given to the parents in 1996, the pre-coded answers were collapsed into four categories in the following manner; elementary school (7 and 9 years of school), upper secondary school (vocational and trade and general education), college/university and other. The question on education included in the 1996 parent survey was phrased: “How many years of education do you have?” The collapsed categories were: elementary school (no education beyond 9 years of mandatory school), upper secondary school (1–2 years and 3 years of upper secondary school), college/university (less than 4 years and 4 or more years of college/university), and other.

### STATISTICAL ANALYSES

Descriptive statistics were obtained by frequency distributions. Strength of agreement between adolescent and parental reports, as well as stability in reports within the adolescent group were assessed by κ correlations with 95% confidence intervals (CI) and the percentage of agreement for each cross tabulation. The κ statistic has the advantage over percentage agreement in that it is adjusted for the proportion of agreement expected by chance, but was developed as a measure of reproducibility of dichotomous nominal variables. Altman has proposed the following guidelines to assess the strength of agreement according to the value of κ; very good (0.81, 1.00), good (0.61, 0.80), moderate (0.41, 0.60), fair (0.21, 0.40) and poor (<0.20).

As attrition analyses showed that the main reason for exclusion from the cohorts was parents who did not participate (table 1), χ² and one way analysis of variance (ANOVA) tests were carried out to determine if there were statistically significant differences in the responses given by adolescents both included and excluded from the four cohorts.

We used ANOVA with Bonferonni’s test to investigate whether differences existed between adolescents within the cohorts who gave no responses, unclassifiable or classifiable answers. Variables used in these analyses were physical activity (times per week), proportions of smokers, proportion of adolescents living with their father/mother and proportion of girls.

In addition, we performed cross tabulation with χ² testing to investigate whether there were systematic differences between the three groups determined by the parents’ answer to the question on occupation, and to investigate if the non-responders or those providing unclassified responses at age 13 provided classified responses at age 15.

All analyses were performed using SPSS 9.0.

### Results

Response rates and reasons for exclusions from the four cohorts are presented in table 1. Mothers and fathers who did not participate...
caused exclusion of 18% and 27% of the adolescents from the SES analyses and 38% and 43% from the education analyses, respectively. Among those adolescents excluded because of lack of parental data, significantly fewer were living with the parent (tables 2 and 3). In the SES cohorts, there were also significantly higher proportions of girls among those excluded because of non-participation of the parent (table 2). Judged by the data provided by the adolescents, the cohorts contained higher proportions of parents classified as having middle/higher non-manual occupations or college/university education, compared with those excluded because of lack of parental data (tables 2 and 3). Those excluded from cohort 1 (fathers’ occupation) also had a higher proportion of non-responses than expected by chance (table 2).

Within the cohorts, the adolescents not responding or giving incomplete answers did not differ from their peers with complete answers with respect to physical activity, smoking, gender, or living with the parent in question (results not shown). Among the non-responding 13 year olds approximately one third of their parents also did not respond, and there were lower proportions of adolescents with parents in the middle/higher non-manual category than expected by chance (see appendix).

Descriptive statistics (tables 4 and 5) show the classification into SES groups based on occupation and educational level according to the reports given by the adolescents and their parents. Table 4 shows a high resemblance in the proportion categorised into the different SES groups based on responses to the occupational question given by adolescents and their parents. The main difference was observed in the proportions of unclassified and no response answers, which were higher among the adolescents than among the fathers. However, the proportion of unclassified answers relating to fathers’ SES decreased from age 13 to age 15 for the adolescents. Table 5 shows the results from the questions on parental education. All parents replied to this question, while 13%–14% of the 15 year old adolescents did not. This non-response rate is of about the same magnitude as the sum of unclassifiable and no response answers to the occupational question in table 4. The distribution of parents’ educational level based on the adolescents’ answers appeared to be systematically lower than the distribution based on the parents’ own reports.

Adolescents’ reports at ages 13 and 15 on their parents’ SES group showed high inter-rater stability, with k statistics above 0.7 (results not shown). Similarly, there were only small or no changes depending on whether adolescent reports at age 13 or 15 were used for comparison with parents’ reports on SES groups (table 6). This finding was consistent across gender of the parent and different combinations of SES categories.

More than 50% of those not responding or giving unclassifiable answers at age 13 were in the classified group at age 15. Still, some 40% of those who did not respond at age 13 also did not respond at age 15 (see appendix).

Discussion

The results of this study support those of West et al by providing evidence that adolescents as young as age 13 or 15 are able to produce valid
answers to questions about their parents’ SES based on an open ended question about parent occupation. The strength of agreement between adolescent and parent reports on SES groups was good according to the rule of thumb for evaluating $\kappa$ values.\textsuperscript{15} The strength of agreement was fairly similar whether pertaining to the SES of the fathers or SES of the mothers, indicating that for validity reasons there is no reason to prefer one to the other. In contrast with the review by Looker,\textsuperscript{8} in which she found increasing accuracy of proxy reports with increasing age of the child, we did not find significant differences in strengths of agreement when using data obtained from the adolescents at age 15 compared with data obtained at age 13. However, the lower value of the $\kappa$ statistic found for education does support Looker’s\textsuperscript{8} notion that children’s proxy reports of parental characteristics remote in time (for example, education) may be less valid than their reports on present characteristics (for example, occupation). As the $\kappa$ statistic is known to be sensitive to both prevalence in each category and the number of categories,\textsuperscript{17} it was surprising to find that the strength of agreement did not vary much between the three different ways of grouping the SES categories.

There were especially two types of attrition in this study that could potentially cause a threat to the generalisability of the results—parents who did not participate and non-response or unclassifiable answers from the adolescents. Based on the adolescents’ responses the groups of parents who did not participate seemed to contain lower proportions of parents in middle/high non-manual occupations and parents with college/university education than the groups of parents who participated. This type of response bias towards high SES/education is not unusual in health surveys,\textsuperscript{15} 16 17 and could bias the results of analytical studies based on such data. However, we do not think children of parents not participating in health surveys would provide less valid answers on measures of SES than those whose parents do participate in health surveys. Difficulties in reporting could occur in low SES families when parents have multiple jobs or frequently change jobs, but they may in general have an easier time giving a single label description of the job in question. High SES families on the other hand could have difficulties in giving a short and comprehensive description because of the complexity of the jobs.

In agreement with results from other studies,\textsuperscript{7} 8 the proportion of non-responses or unclassifiable answers among the adolescents whose parents did participate was substantial (11\%–16\%). The $\chi^2$ analyses of adolescent non-responders and those providing unclassifiable answers indicated that a higher proportion of the parents of these children were also non-responders, which supports Looker’s\textsuperscript{8} theory that some families may have more problems reporting on this measure. As outlined above both low and high SES families could have difficulties providing a simple answer to a question on type of work. However, based on the data presented here we cannot conclude that one of these explanations is more probable than the other.

Among the parents who did not participate in the study there were lower proportions of the adolescents who reported that they were living with the parent in question. This could partly explain the lack of response from the parent as well as the tendency towards higher proportions of non-response/unclassifiable answers by the adolescents. However, when the parent did participate whether the adolescent was living with the parent in question did not seem to differ between responders, non-responders, or those giving answers that were not classifiable. These apparently contradicting results may be explained by the type of contact the adolescents have with the parent they are not living with. Frequent contact or good relations may ensure that the parent participates and that the adolescent knows the work of the parent. Thus, the adolescent’s ability to provide a valid answer may not be a simple question of living with the parent or not, as suggested by the studies reviewed by Looker.\textsuperscript{8}

The values of the $\kappa$ statistic for education were much lower than those for SES (based on occupation), but within the interval of fair values according to Altman.\textsuperscript{15} There are at least two obvious and plausible explanations for these low agreements between adolescents and parents on this question. Firstly, low agreement could be caused by lack of relevance and closeness in time of parental education in the daily life of an adolescent. The data were, however, collected half a year before the adolescents had to make a major choice about their own future plans, as mandatory school in Norway ends at age 16. The issue of work versus education, and for those opting for education the issue of vocational versus general studies, could therefore be expected to be relevant to the adolescents.\textsuperscript{18} The second explanation is methodological, and pertains to the different answer
categories provided in the adolescent and parent questionnaires. The options in the adolescent questionnaire reflected former versions of the school system. The answer categories thus differentiated between the amount of years of mandatory school (7 or 9 years) and also provided three different options at the voluntary upper secondary level (vocational, trade and general), while there was only one option for higher education. In addition to being confusing in itself, the number of options for lower level education could have caused a systematic reporting bias relating to parents with college degrees in business or engineering, for instance. As the adolescents themselves were about to make a choice about their upper secondary education, they may have been less aware of the possibilities beyond this level.

Although the $\kappa$ statistic is considered better than percentage agreement, its widespread use, or misuse, has been criticised by MacIntyre and Willett.\textsuperscript{14} Firstly, their criticism relates to the extension of the use of the $\kappa$ statistic beyond dichotomous data for which it was originally intended, and thus to its application to nominal as well as ordinal polychotomous data.\textsuperscript{14} Secondly, they point out that the $\kappa$ statistic was not intended for assessing validity, because in validity studies one of the instruments is expected to provide a truer value than the other, and thus other measures of agreement may be more informative.\textsuperscript{15} However, the measures of agreement suggested to assess validity are better for analysing continuous or true ordinal polychotomous data than quasi-ordinal data like the ones presented in our study. Furthermore, Nelson et al.\textsuperscript{16} point out that comparison of index and proxy respondent data often has been considered as an issue of reliability rather than validity. West et al.\textsuperscript{17} argue that on some issues like parental occupational status, the children’s reports may be more truthful than their parent’s, thus questioning the parental answers as the gold standard by which to measure validity.

Despite his demonstration of the $\kappa$ statistic’s sensitivity to both the prevalence in each category and the number of categories, Altman\textsuperscript{18} argues strongly for the use of $\kappa$ statistics, as other statistics often used (for example, Pearson's correlations coefficients and the $\chi^2$ test) measure association rather than agreement. However, the requirement for a perfect agreement between the answers given by the two responders does make the $\kappa$ statistic a conservative measure of agreement. The weighted $\kappa$ has been constructed to take into account the off diagonal distance between mismatches.\textsuperscript{19} Because of the characteristic of the data in this study, assigning weights to distances between the SES categories would have been arbitrary, and thus of little use. The high and similar values of the $\kappa$ statistic across the three models may be taken as an indication that grave mismatches between parents and adolescents' reports were not common and thus the influence of the number of categories less important. These results are encouraging for future analyses on differences in self reported health and health related behaviour of these adolescents.

Our study shows that an open ended question about parental occupation in a questionnaire answered by adolescents at the age of 13 and 15 provides valid data on parents' SES among the 85%–90% who gave classifiable answers. Researchers may be able to increase the proportion of classifiable answers regarding parental occupation by including more specific questions to cue adolescents as to what information to provide. More specific questions could consequently also increase the precision of the answers, thus making assignment of job codes less dependent on personal judgement of researchers. This approach would be in line with the intended use of the SES system, which requires data from a range of questions to assign a person to one of 33 categories.\textsuperscript{11} In addition, such an approach might be less sensitive to economic changes than developing additional questions on material wealth as suggested by Currie et al.,\textsuperscript{7} even though social status related to occupation clearly is not indifferent to economic and cultural changes. An alternative to including more questions on the survey about parent occupation is to have the researcher (or the person distributing the questionnaire) emphasise the importance of answering the question carefully and encourage students to ask for help. This would be a similar, but less costly approach to the “mini-interview” performed by West et al.\textsuperscript{19} However, the ethical aspects in relation to the anonymity of the adolescent would have to be considered carefully.

Because of the methodological problems related to the question on education in our study, further research should investigate the possibility of adolescents’ ability to provide valid proxy reports of their parents’ education. This may be an especially salient issue when considering health related behaviour in adolescents as a mediating factor between social status and inequalities in mortality and morbidity later in life,\textsuperscript{20} as research on health related behaviour among adults has shown a relation between those outcomes and level of education.\textsuperscript{21,22}

Funding: The Norwegian Research Council provided financial support for this work.

Conflicts of interest: none.