Sampling procedure, participation rates and representativeness in the Swedish part of the European Youth Heart Study (EYHS)

Anita Hurtig Wennlöf1,2,*, Agneta Yngve1 and Michael Sjöström1

1Unit for Preventive Nutrition, Biosciences at NOVUM, Karolinska Institute, Huddinge, Sweden: 2Division of Biomedicine, Department of Caring Sciences, Örebro University SE-701 82 Örebro, Sweden

Submitted 4 July 2002: Accepted 23 September 2002

Abstract

Objective: The European Youth Heart Study (EYHS) is a cross-sectional, school-based population study on risk factors for future cardiovascular disease in children, with an overall participation rate in Sweden of about 50%. To study the representativeness of the participants in the Swedish part of EYHS, a comprehensive non-participant follow-up study was carried out.

Design: A structured multilevel analysis model was developed, addressing each level in the sampling procedure. The income, educational and occupational categories of the geographical regions of the study (level I), school catchment areas (level II) and parents (level III) were compared with official data. Participating and non-participating pupils (level IV) were compared through a questionnaire.

Setting: Thirty-seven state schools in two regions of Central Sweden (Örebro and southern Stockholm) were visited during the school year 1998/1999.

Subjects: Boys and girls aged 9 and 15 years were randomly sampled through a multiphase sampling procedure.

Results: Data for socio-economic status for levels I and II corresponded well to national and regional official data. At level III, non-manually working parents were slightly over-represented among parents of participating children. At level IV, non-participating subjects corresponded in most respects to participants with a few exceptions – mainly more interest in physical exercise among participants.

Conclusions: Based on the knowledge from the non-participant study, we do not foresee problems regarding interpretation of the outcomes in the EYHS, despite the low participation rate.
Material and methods

Material, EYHS main study and the non-participant study

Data were collected between the beginning of September 1998 and the end of May 1999. The sampling frame was all state schools for children aged 9–10 years (grade 3) and 15–16 years (grade 9) in Örebro and the southern region of Stockholm. The municipality of Örebro has 160,000 inhabitants, of whom there are about 1500 pupils in each of these age groups. The southern region of Stockholm, defined here as the municipalities of Huddinge, Haninge, Botkyrka, Nynäshamn, Salem, Södertälje and Tyresö, has about 370,000 inhabitants, of whom there are about 5000 pupils in each of the two age groups. The non-participant study was carried out in the schools that collaborated in the EYHS.

Methods

Sampling in EYHS main study

The two geographical regions were chosen, and within those a multiphase sampling procedure was used. All state schools with more than 20 children in the age groups in question were initially included. The schools were stratified with regard to school grade and the mean income level in their catchment areas (below or above the mean in their municipality). A random sampling procedure from each stratum was carried out, and the sampled schools were invited to participate in the study.

Non-participant study

The non-participant study followed when the main study was closed at every school. One visit was scheduled to each class unit, and a questionnaire was handed out to all pupils present on that single occasion – participants, non-participants and those not sampled. The questionnaire was a shortened hard copy version of the computerised EYHS main study questionnaire, with some additional questions about reasons for participating or not. The questionnaire included self-reported height and weight, habits and attitudes regarding physical activity, food, smoking, drugs and family life.

The analysis addressed each level in the sampling procedure as follows.

- Level I (regions): from official data, income, educational and occupational categories for the regions were compared with the same variables for Sweden.
- Level II (schools): the mean income level of each school’s catchment area was compared with the mean income level in the municipality. Non-collaborating schools were also contacted by phone to find out why the school had decided not to participate in the study.
- Level III (parents): to evaluate the representativeness of the parents of participating children, we compared data on self-reported income, occupational and educational categories from the parents’ questionnaire with official regional data. For comparison, data for the age

From the complete lists of pupils in the collaborating schools, groups of pupils proportional to the sizes of the respective schools were randomly selected with the aim of obtaining 250 participants in each age and gender group (totally 1000 participants). Fifty per cent over-sampling was performed in accordance with experiences from a pilot study. The sampling procedure was organised as follows.

1. Sampled pupils received a letter with information, a written informed consent form and a questionnaire to be completed by both of the parents or the legal guardians. This included questions about their income, education and occupation and the child’s health. All information and questionnaires were distributed through the class teachers. The pupils were asked to return the completed form and the questionnaire to the class teacher in an enclosed envelope within a week.
2. A reminder was distributed in the same way to non-responding families, 10 days after the first letter.
3. For those who agreed to participate, a third letter with practical information was distributed one week before the day of most tests in the test battery (see Introduction).
4. Lists of participants were distributed to the teachers in order to ensure that the pupils would be allowed, and reminded, to undergo the tests during school hours.
group 30 to 55 years was used, which best corresponded to the age of the parents (mean 43 years, range 26–72 years). Data from the parents’ questionnaire concerning own or close relatives’ CVD were compared with national data9–11.

- Level IV (individuals): for comparisons between participating and non-participating pupils, an objective measure (written school marks) and a subjective measure (questionnaire) were used. Non-participants are defined as sampled pupils not participating in the study (i.e. sampled pupils who gave a negative answer to the question of participation, or who did not respond at all, or who gave a positive answer regarding participation but did not attend the tests).

The collaborating schools were asked to report the individual written marks from the ordinary schoolwork. No written marks are given in grade 3. In grade 9 written marks are usually given in 17 subjects, and there are four categories of credit points in each subject (0, 10, 15 and 20 credit points, respectively); thus a pupil can receive a maximum of 340 credit points. The total individual credit points, respectively. In grade 3 (9–10 years, born 1989) and among 15–16-year-olds these rates were 36% and 26%, respectively (Table 1). Demographic data showed differences between participants and non-participants. Split into grade groups (i.e. gender and grade). The large number of tests increases the risk of false significances. For this reason it is preferable, instead of using the common local significances, to use the overall significance, or family error rate (FER). FER is defined as the probability that one or more false significances out of $k$ tests is less than or equal to $\alpha$.

Bonferroni’s method is the most common approach when handling multiple comparisons, but it is very restrictive and means quite simply a change of the local significance level, $\alpha$, to $\alpha$ divided by $k$, where $k$ is the number of tests. The Bonferroni procedure guarantees that FER is less than or equal to $\alpha$. Holm12 introduced another, less restrictive method to keep FER less than $\alpha$. The method of Holm proceeds as follows. Sort the $P$-values of the $k$ tests in increasing order, $P_1, P_2, \ldots, P_k$. If $P_1 > \alpha/k$, none of the $k$ tests are significant, and the test procedure is finished. If $P_1 \leq \alpha/k$, test 1 is significant, and now $P_2$ is examined. If $P_2 > \alpha/(k - 1)$, none of the $(k - 1)$ remaining tests are significant, but if $P_2 \leq \alpha/(k - 1)$, test 2 is significant and $P_3$ is examined. This procedure goes on until $P_i > \alpha/(k - i + 1)$ and the procedure is interrupted.

The level of significance used was $\alpha = 5\%$. Holm’s multiple test procedure for correction of mass significance effects12 was applied to the questionnaire data. As an illustration of the method of Holm, the local $P$-values from the chi-square tests are reported together with modified $P$-values according to Holm, for one subset of data.

### Results

#### Participation in the main study

**Schools and individuals**

The school boards of 42 out of 87 sampled schools gave a positive answer to participation (47% in Örebro and 51% in Stockholm).

In total, 2313 pupils were sampled. Of these, 1154 (50%) agreed to participate and gave written informed consent, 566 (24%) declined to participate and 593 (26%) did not respond at all. Of those who agreed to participate, 17 (1%) failed to keep their appointment and thus a total of 1137 (49%) of the sampled pupils participated. Split into grade and gender groups, the participation rates among 9–10-year-old boys and girls were 59% and 65%, respectively, and among 15–16-year-olds these rates were 36% and 46%, respectively (Table 1). Demographic data showed

<table>
<thead>
<tr>
<th>School</th>
<th>Grade 3 (9–10 years, born 1989)</th>
<th>Grade 9 (15–16 years, born 1983)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Girls</td>
<td>Boys</td>
</tr>
<tr>
<td>Örebro</td>
<td>62%</td>
<td>61%</td>
</tr>
<tr>
<td>Southern</td>
<td>68%</td>
<td>57%</td>
</tr>
<tr>
<td>Stockholm*</td>
<td>Total</td>
<td>65% (290)</td>
</tr>
</tbody>
</table>

* Southern Stockholm – Botkyrka, Haninge, Huddinge, Nynäshamn, Salem, Södertälje and Tyresö municipalities.
that approximately 0.5% of the Swedish children in the actual age groups was assessed. Non-response items, such as registrations of low quality or missed venepuncture, will be reported separately.

Non-participant study

Level I, regions

No essential differences between regional and national population levels were found in the studied variables (Table 2). The largest differences were observed for occupation, where the Stockholm region had 4 percentage units more non-manual workers than Sweden as a whole. For all other categories of occupation and education the differences between regional and national data were smaller. The mean income level differed by 8000 SEK between the regional population of Stockholm and the national population (206 000 SEK and 198 000 SEK for Stockholm and Sweden, respectively).

Level II, schools

No special pattern in income levels of catchment areas was found in non-collaborating schools; high- and low-income areas declined collaboration to a similar extent (data not shown). The non-collaborating schools frequently referred to a shortage of facilities for the implementation of the project as the main reason for not participating in the EYHS; three to four small rooms were required during the days of tests.

Level III, parents

Data on parental income, occupation and education, in comparison with official regional and national data, are also given in Table 2. Results from the parents of participating pupils are presented for the two study locations (Örebro and Stockholm) separately, for comparison with regional data, and the weighted averages for the two study locations (Sweden) are given for comparison with the national population. Parents of EYHS participants had a median income of 201 000 SEK, compared with a national mean income level of 198 000 SEK (Table 2). At a regional level, the parents of participants in Örebro had a lower income than the regional population (median value 191 000 SEK and mean value 198 000 SEK, respectively), while the income (median) of the parents in Stockholm equalled the regional population mean income (206 000 SEK). The distribution of the occupational and educational levels showed similar patterns in the two regions. Manual workers were under-represented by 13 and 7 percentage units in Örebro and Stockholm, respectively, as compared with official data for the regions. Non-manual workers were over-represented by 9 and 4 percentage units in Örebro and Stockholm, respectively. A low educational level was under-represented by 12 and 10 percentage units in Örebro and Stockholm, respectively, and a university level was over-represented by at most 6 percentage units compared with official regional data.

Forty-five per cent of the parents reported to have at least one close relative with diagnosed CVD, and 2% of the parents reported own CVD.

Level IV, individuals

Written marks. Information on all written marks was obtained for 1930 pupils, covering about 90% of the pupils in grade 9 in the collaborating schools. There was a significant difference between the participants and the rest of the pupils in average credit points, all 17 school subjects included (217 ± 58 vs. 183 ± 71, respectively; \( P < 0.001 \)).

Table 2 Income level and occupational and educational categories in Swedish families participating in the European Youth Heart Study (EYHS), compared with regional and national populations

<table>
<thead>
<tr>
<th></th>
<th>Örebro</th>
<th>Stockholm</th>
<th>Sweden</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parents</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income (SEK)</td>
<td>191 000*</td>
<td>198 000†</td>
<td>201 000*</td>
</tr>
<tr>
<td>Missing (%)</td>
<td>6</td>
<td>–</td>
<td>5</td>
</tr>
<tr>
<td>Occupation (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manual worker</td>
<td>30</td>
<td>43</td>
<td>35</td>
</tr>
<tr>
<td>Non-manual worker</td>
<td>53</td>
<td>44</td>
<td>49</td>
</tr>
<tr>
<td>Other‡</td>
<td>13</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>Missing</td>
<td>4</td>
<td>–</td>
<td>4</td>
</tr>
<tr>
<td>Educational level (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 9 years</td>
<td>15</td>
<td>27</td>
<td>20</td>
</tr>
<tr>
<td>≤ 12 years</td>
<td>42</td>
<td>44</td>
<td>44</td>
</tr>
<tr>
<td>University level</td>
<td>32</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>Other education</td>
<td>8</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Missing</td>
<td>3</td>
<td>–</td>
<td>4</td>
</tr>
</tbody>
</table>

* Median value.
† Arithmetic mean value.
‡ Own business or not gainfully employed (e.g. student, unemployed, retired).
Participants had significantly higher marks in physical education than their non-participating peers (13.8 ± 4.4 vs. 11.5 ± 5.5, respectively; P < 0.001). The distribution of the credit point categories is shown in Fig. 1.

Questionnaire. Questionnaire data were collected for a total of 2360 pupils. Thirty-two questionnaires were excluded because of a very bad quality, and 559 came from pupils not sampled in EYHS and were not further analysed. The remaining 1769 questionnaires were from 810 non-participants and 959 participants. About 80% of the pupils were present in the classroom at the time of distribution of the questionnaire.

The most common reason given for non-participation was ‘I didn’t want to have a blood sample taken’; 38% of the 9-year-olds and 33% of the 15-year-olds gave this answer. The second most common answer (22%) given by the 9-year-old children was ‘My parents didn’t want me to participate’, with similar responses from boys and girls. This answer was given as an explanation by only 4% of the non-participating 15-year-olds. In the older group, the second and third most common answers given by the girls were ‘My parents didn’t want me to participate’ (5%) and ‘I didn’t want to perform a bike test’ (5%), and by the boys, ‘I didn’t want to leave home without breakfast in the morning’ (12%) and ‘Because my friends were not participating’ (3%).

The analysis of the self-reported height and weight showed no significant differences between participants and non-participants between genders or grade groups. However, when body mass index (BMI) was calculated from the self-reported values, significant differences in BMI was found among the 9-year-old girls between participants and non-participants, with values of 16.7 ± 2.2 and 17.1 ± 2.3 kg m⁻², respectively (P < 0.05).

Among the girls in grade 3, no significant difference between participants and non-participants was found regarding answers in the questionnaire. In the girls in grade 9, the answers to one question differed significantly, namely the question/statement: I really like doing physical education at school (Alternatives: yes/no). The participants showed a significantly more positive attitude to physical education.

In boys, significant differences between participants and non-participants were observed for three questions in grade 3, and for two questions in grade 9 (one of the questions was the same in both age groups); all concerning physical exercise and eating habits.

Among the boys of both school grades, the participants were more positive to the statement If I were to exercise most days, it would get or keep me in shape (Alternatives: yes/perhaps/no). The participating boys in grade 3 took more part in clubs, etc., as seen by the answers to the question: How often do you take part in exercise at clubs, such as sport clubs, youth clubs, etc? (Alternatives: hardly ever or never/once or twice a week/nearly every day/every day).

There were also differences in eating habits among the boys. According to the questionnaire, the participants in grade 3 ate fewer hamburgers, and the participants in grade 9 ate fewer pizzas, compared with the non-participants. The questions were: How often do you eat burgers? (Alternatives: every day/nearly every day/once or...
The correction for FER using the method of Holm decreased the number of significant differences in the chi-square test for all four age and gender groups, from 34 to six questions. An example for third grade boys is given in Table 3.

Discussion

The individual decision to volunteer and participate in a study depends on a number of factors and is not a random process. In studies including children, the situation is complicated by the fact that other persons (e.g. parents, teachers and friends) influence the final decision to take part or not. For this reason, the present study of the representativeness of the Swedish sample in the EYHS addressed adult characteristics (i.e. characteristics of the parents) as well as the habits and attitudes of the children.

The primary finding was that participants and non-participants reported similar attitudes and behaviours in most areas covered by the questionnaire, except for interest in physical exercise and to some extent for food habits among the boys. This selection bias in voluntary amount of physical exercise may decrease the range of exposure to physical activity, but will not necessarily influence the outcome (i.e. the risk for future CVD), as the risk factors are supposed to be multi-factorial. Nevertheless it is of importance, since physical activity is one of the variables of interest. We will make use of this information in our further interpretations of the data, taking into consideration that we may have lost part of the group least interested in physical exercise, which is one component of total physical activity.

Moreover, we found the socio-economic background of the participants to be slightly above average, compared with that of the general Swedish population, and the data will be interpreted with this taken into account.

The reason for studying the SES of the participants as one aspect of representativeness is based on the known risk in epidemiological studies of under-representation of the most disadvantaged groups. Groups with lower SES are less likely to participate in risk factor screening programmes or to receive preventive services. Socio-economic factors have been associated with non-participation in previous Swedish health surveys, but the findings are conflicting and vary from underestimation of differences between participants and non-participants to overreporting of such differences. The impact of socio-economic conditions in childhood on future CVD is not fully understood. Health behaviour is found to be related to SES in children also, although the adult socio-economic destination seems to be the best predictor of mortality, not the childhood origin per se. Despite a major political goal to decrease social inequalities in Sweden, it has recently been shown that the differences are still maintained or even increasing in some groups. Furthermore, the most deprived neighbourhood environments have been shown to have their own independent increased risks for a poor cardiovascular health profile. These findings correspond to those in international studies.

On the school level, or the level of school catchment areas, only income data were available, and these were used only as a discriminator to dichotomise schools into those with high and low income. The reason for analysing the school level was a concern that all kinds of schools should be represented, as further characteristics may follow with income, such as immigrant rates and unemployment rates, for example – data we could not analyse for every individual school.

Despite the under-representation of lower SES families in this study, we found a parental income level (median value) in agreement with the national mean value. This might be due to part-time working in well-paid families, but we have no information on part-time work. The population data are based on all inhabitants, with or without children. The restriction to a certain age group, 30–55 years, limits the error in comparing these different groups, but we do not know the magnitude or direction of that possible error.

There is no selection bias concerning the inherited risk for CVD according to the parental questionnaire data. This conclusion is based on the concordance between the parental results and national data on CVD prevalence. An over-representation of higher SES families may result in an underestimation of risk factors, if we assume that CVD risk factors correlate inversely to SES. Griffin et al. point out another possible cause of non-participation that may also result in an underestimation of the risk, namely the fear of finding one’s child out of the expected normal ranges. On the other hand, the response rate is likely to be high when the subjects have a special motivation to participate, e.g. experience of a high CVD risk, and this would give an overestimation of risk factors. The decision to participate for the youngest children is probably highly influenced by the parents’ attitude towards the study, which is reflected by the fact that, in the questionnaire in the non-participant study, 22% of the children stated that their parents didn’t want them to participate. Families with close relatives with CVD can be defined as a high-risk subgroup. A recent Swedish, population-based case–control study of myocardial infarction, the Stockholm Heart Epidemiology Program (SHEEP), reported 58% of the referents to have at least one close relative with CVD. However, as the SHEEP referents were older than the parents in the EYHS (mean age 59 and 43 years, respectively), the difference between 58% in SHEEP and 45% in EYHS is expected. As the proportion of participating children’s families with close relatives with
this diagnosis correspond very well to official data on CVD, we do not have to consider the possibility of an overestimation of risks due to over-representation of children from high-risk families, or of a recall bias phenomenon.

The reasons for non-participation on an individual level are not easily detected. The reported concerns about the blood sampling might partly be the easiest way to conceal lack of interest, but we did not discover any underlying explanation with the method used. The statistical analyses of the reported written marks are consistent with our impression during the study period that the most successful pupils participated to a higher degree in the EYHS. The explanation for this cannot be related exclusively to SES, as we found only minor discrepancies in the SES distribution compared with the general population. More likely, it mirrors a generally positive attitude towards all kinds of school-based activities among the participants.

Previous studies in this area have revealed a similar pattern of participation as in our study. It has been noted before that adolescent girls are more interested than boys in participating, that there is a skewed SES distribution between participants and non-participants and that adolescents are less interested in participating than younger children. Non-participant studies designed to find a few subject characteristics (e.g. interest in physical activities) could be expected to indicate that participants are generally younger, more educated and more likely to report healthier patterns2,3,27.

In conclusion, we found no main objections against the choice of regions, or against the sampling results on a school level, and therefore conclude that the sampling frame was representative of the source population. As the random sampling procedure is considered to have achieved a representative sample of families, the slight selection bias found regarding SES is most probably due to the previously described socio-economic effect on propensity to respond1–4.

As the participation was quite demanding on the subjects (e.g. blood sampling and maximal ergometer bicycle test), the self-selection bias observed on an individual level was not unexpected. Based on the very few significant differences found between participants and non-participants in the questionnaire result in the non-participant study, we do not foresee problems regarding interpretation of the results in the EYHS main study, despite the low participation rate.

References


Table 3: Test procedure and comparisons between local significant values, i.e. for each question and the overall significances, FER adjusted as described by Holm12. The example shows test results for boys of grade 3, and includes the 13 locally significant results from the 48 questions

<table>
<thead>
<tr>
<th>Local P-values, sorted in increasing order</th>
<th>Modified a-level</th>
<th>Local significance</th>
<th>Modified significance</th>
<th>FER significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>How often do you take part in exercise at clubs, such as sport clubs, youth clubs, etc?</td>
<td>0.000</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>I feel that I can easily keep up with other children of my age when we are playing games and sports.</td>
<td>0.004</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>I feel that I can easily keep up with other children of my age when we are playing games and sports.</td>
<td>0.005</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>How often do you eat burgers?</td>
<td>0.002</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>How often do you eat chips?</td>
<td>0.009</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>How often do you eat pizza?</td>
<td>0.013</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>I feel that I can easily keep up with other children of my age when we are playing games and sports.</td>
<td>0.014</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>How often do you drink beer?</td>
<td>0.015</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>How often do you have fizzy drinks?</td>
<td>0.017</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>How often do you eat chips?</td>
<td>0.019</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>How often do you eat pizza?</td>
<td>0.020</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Number of significant tests</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Number of significant tests</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
</tbody>
</table>

FER – family error rate.


