A cluster-analytical approach towards physical activity and eating habits among 10-year-old children

Dieter Sabbe1*, I. De Bourdeaudhuij2, E. Legiest1,2 and L. Maes1

Abstract

The purpose was to investigate whether clusters—based on physical activity (PA) and eating habits—can be found among children, and to explore subgroups’ characteristics. A total of 1725 10-year olds completed a self-administered questionnaire. K-means cluster analysis was based on the weekly quantity of vigorous and moderate PA, the excess index (weekly consumption of sugar and/or fat) and the daily diversity index. Chi-squares tested gender differences in clusters and associations with socio-economic status (SES), overweight, controlling for gender. Following distribution was reliable: Sporty Healthy Eaters (n = 242; high vigorous PA, average moderate PA, low excess, higher diversity), Sporty Mixed Eaters (n = 288; high overall PA, very high excess, high diversity), Moderate Active Healthy Eaters (n = 221; average vigorous PA, highest moderate PA, lower excess, higher diversity), Unsporting Unhealthy Eaters (n = 276; below average on all indexes, diversity extremely low) and Sedentary Healthy Eaters (n = 318; lowest overall PA, higher excess, highest diversity). The Sporty Healthy Eaters and Sporty Mixed Eaters comprised more males, Sedentary Healthy Eaters more females. No associations with SES or overweight were found for the clusters. Co-occurrence of healthy and unhealthy behaviour exists. Only Sporty Healthy Eaters combine high levels of PA with low excess index and higher dietary diversity index. Effective ways of directing children to selective, individual relevant recommendations should be developed.

Introduction

Inappropriate dietary and physical activity (PA) habits are two modifiable risk factors associated with obesity and several chronic diseases [1, 2]. Consequently, promoting healthy dietary habits and sufficient PA among children has the potential to improve the immediate health of young people, as well as contributing to their health and emotional well-being during adulthood [3].

Dietary recommendations promote increased dietary diversity along with reducing the intake of selected nutrients such as fat, refined sugar and salt [4]. Dietary diversity might most appropriately be encouraged by promoting the daily intake of foods from each of the major groups of the Food Guide Pyramid [5, 6]. In addition, the Food Guide Pyramid recommends limiting the amounts of fat, saturated fat, cholesterol, sodium and refined sugars in order to prevent so-called ‘well-being syndromes’, both for adults and children. Several studies showed that tracking nutrient intake and dietary habits was stronger among young children than among adolescents [7–10]. Bertheke Post et al. [11] reported a moderate stability of nutrient intake over 20 years, from adolescence to adulthood. These results highlight the importance of developing healthy eating practices during early childhood but especially...
during adolescence in order to maintain an optimal position in rank.

Children and adolescents are recommended to accumulate at least 60 min of moderate to vigorous PA on most days of the week, or preferably on each day [12–14]. While young people are more active than adults, previous research has reported a decrease in the levels of PA with ageing, in adolescents as well as in pre-adolescent boys and girls [15–20]. Despite this decrease in absolute time of PA, past exercise behaviour or exercise habits emerged as a consistent predictor of adult activity status [21]. In their prospective study, Trudeau et al. [22] reported a small but significant positive impact of childhood PA on adult PA. These reflections justify the need for systematic promotion of (at least moderate) PA among younger children already in order to pursue firmly established habits.

Until now, eating habits and PA among children have mostly been examined independently. Nevertheless, identifying group disparities (or clusters) in health-enhancing or health-declining behaviour remains important in order to target health promotion interventions. Only a few studies addressed the relationship among multiple lifestyle-related health factors, and their demographic correlates. Pioneers Nutbeam et al. [23] examined adolescent health behaviour’s multi-dimensionality applying other statistical techniques. Pronk et al. [24] emphasized the importance of exploring such relationships for a variety of reasons. First, health improvement planning efforts may benefit from the prevalence, distribution and frequencies of these clusters. Second, there might be a potential synergistic effect of multiple healthy lifestyle factors on the risk of chronic conditions and health outcomes. A better understanding of the clusters and demographic correlates may give additional insights to improve specific recommendations in order to prevent chronic diseases.

A previous study in adults showed that simultaneous health interventions, addressing both PA and nutrition, were recommended, as they were more efficient in increasing PA and reducing fat intake [25]. It is important to note that a cluster-analytic approach differs from the general practice of investigating inter-correlations between health behaviours. In their investigation of the co-occurrence of several energy balance-related behaviours, Kremers et al. [26] explored inter-correlations. The small associations they found show that active individuals do not always have healthier diets. These lower associations, however, do not exclude co-occurrence of behaviours within certain groups.

In our manuscript, cluster analysis is used to detect possible similarities between health behaviours in specific subgroups. The purpose of the study is to investigate whether two dietary habits (particularly dietary diversity and dietary excess), vigorous and moderate PA among children cluster in a reliable way among 10-year-old children. Additionally, we want to investigate whether the identified groups could be characterized by the demographic variables gender, socio-economic status (SES) and body mass index (BMI). Specific relationships between these parameters could stimulate targeted health promotion in order to maximize the relevance of the communication concerning appropriate dietary habits and PA.

---

**Material and methods**

**Participants and procedure**

Data were gathered between October and December 2002. One hundred elementary schools from two Flemish regions were randomly selected from the list of schools subsidized by the Flemish government. The principals were sent a letter and afterwards contacted by phone. A total of 69 principals agreed to co-operate in the study. All children of the fifth grade were invited to participate in the study (n = 1957). In the classroom, the children completed a self-administered questionnaire on eating habits and on PA, demographic variables and possible psychosocial determinants, under the supervision of one researcher and their classroom teacher. Each of the children was given a letter of consent and a questionnaire to be completed by the parents. Parents were able to contact the researchers if they had any questions. Documents were collected by the teacher in a closed envelope which the research
team gathered 1 week later. Response percentage of the parents was 88.1, which resulted into 1725 child–parent couples. Questionnaires of pupils, whose parents did not provide written consent, were destroyed. The average age of the children was 10 years, 895 of the children were boys (51.9%), and 1686 had the Belgian nationality (98%).

### Measures

**Dietary habits and PA**

Based on a validated 19-item Food Frequency Questionnaire (FFQ) [27], two dietary indexes were calculated, a ‘dietary diversity index’ and an ‘excess index’. FFQ categories ranged from ‘never’ to ‘more than once a day’. Similar to the dietary diversity score of Drewnowski et al. [28], the dietary diversity index measures whether the overall daily diet covers the Food Guide Pyramid. The five major groups are dairy, meat, grain, fruit and vegetables (water consumption was not included). Children’s scores ranged from 0 to 5. Score 0 suggests that none of the essential products is consumed daily during an average week, while score 5 indicates daily consumption of products belonging to all Food Guide Pyramid groups. The excess index is a measure for the weekly intake of sugar- and/or fat-containing products. This measure is a summation of the frequency of the consumption of unhealthy products such as sugary sodas, French fries, cookies, candies and chips. Children indicating never to eat these products got score 7 while score 35 was allocated to children consuming all considered items daily. Strictly taken, the excess index cannot be interpreted as one single behaviour but rather as a composite measure for the frequency of the consumption of different food products. For reasons of readability in the text, the index will sometimes be referred to as ‘health-related behavioural indicators’ when all indexes are mentioned.

In their study, Vereecken et al. (in preparation) examined both indexes’ validity and reliability. For all but few food products, they found an over-estimation of the mean consumption frequency measured by FFQ compared with the frequency assessed with a food diary. Correlations between the two instruments varied from 0.17 (light soft drinks, which were not included in any of our indexes) to 0.67 for white bread. These results are fair and comparable to these found in the literature. Furthermore, a satisfactory test–retest reliability for the dietary diversity score (0.71) and for the excess index (0.85) was found.

PA in diverse domains was assessed making use of the ‘Flemish Physical Activity Questionnaire’ (FPAQ). Based on the FPAQ, the level of moderate and vigorous PA was calculated and expressed in minutes per week. Vigorous PA is the summation of sports participation at school and during leisure time. Moderate PA is the addition of moderate activity at school, and the minutes of active transportation to and from school.

Philippaerts et al. [29] investigated the FPAQ’s validity by calculating Pearson correlation coefficients between the questionnaire and the Computer Science and Applications (CSA) uniaxial accelerometer among young adolescents. Intra-class coefficients generally exceeded 0.70 and all Kappa values but one varied from 0.44 to 1.00. Although transport variables showed smaller relationships with CSA, sport participation during leisure time, sport participation summed with total transport and the frequencies of moderate and hard activity were significantly correlated with CSA ($r$ between 0.48 and 0.78). These data suggest that the questionnaire and its composite measures provide valid information.

**Socio-demographics**

Taking into account the advantages described by Winkleby et al. [30], we conceptualized SES as the highest parental education. Based on parental answers, three educational categories were calculated: low (no degree, primary school, vocational training), medium (technical or general secondary education) or high (higher education, university). We found that 15.5% of the families were situated in the lowest category, 24.5% in medium class and 50.3% of the parents were of a higher SES (9.6% of the answers on this variable were missing).

Parent-reported weight and height were used to calculate the BMI ($\text{kg/m}^2$) and overweight. Cut-off scores were adjusted by means of Cole’s curves.
The distribution of overweight was 77.2% not overweight, 7.8% overweight and 14.8% missing values.

Statistical analysis
Analyses were carried out using the SPSS 12.0 for Windows package. K-means cluster analysis was used to identify subgroups (clusters) of respondents with similar dietary and PA habits. Chi-squares were used to confirm associations between cluster membership and gender, SES and overweight. For SES and overweight, these tests were controlled for gender.

The K-means partition method aims at minimizing the distances between observed scores and the cluster centres. In the final cluster solution, each case is allocated to one cluster. Before clustering, all variables were transformed into standardized z-scores. This prevents variables measured in larger ranges from contributing to the distance to a larger extent than variables with smaller ranges do. Moreover, in order to obtain similar categories and to approach a normal distribution of scores on the excess index, vigorous and moderate physical activities were recoded into six categories—based on their cumulative frequencies.

A possible instability of the results could be one of the cluster analysis’ drawbacks. The reliability of the created cluster solution was examined by randomly dividing the sample into two subsamples in which the same clustering procedure was repeated. Additionally, a Kappa degree of concordance in cluster membership was calculated by comparing memberships of both subsamples separately with those of the total sample.

Following these procedures, the five-cluster solution was identified as the most adequate and reliable representation based on the four health behaviour variables. So-called membership concordance ranged from good to excellent: \( k = 0.969 \) \((P < 0.0001)\) for one subgroup and \( k = 0.537 \) \((P < 0.0001)\) for the complementary group. By means of analysis of variance (ANOVA) and a post hoc Bonferroni test, we investigated cluster differences on each index.

Results

Children’s characteristics
The total sample can be characterized by an average of 3 hours and 20 min of vigorous PA and an additional hour and 11 min of moderate PA per week. Thirty-five percentage of the respondents reached the health-related activity guideline \([32]\) by accumulating 60 min of (at least) moderate PA a day, 57% came short of the norm (8% missing values).

Of the participants, 7.4% indicated not to eat any of the necessary products on a daily basis, while 11% of the children had a diversified daily diet (10% missing values). After recoding the scores on the excess index and the scores on the PA levels, all index scores approached normal distribution.

Inter-correlations
Spearman’s correlations between the dietary diversity index, excess index, moderate and vigorous PA were low (all \( r \)'s <0.20—see Table I). The highest correlations were found between vigorous and moderate PA.

Description of the clusters
Figure 1 presents the five clusters obtained from the cluster analyses based on the four health behaviour indexes. Distinguishing characteristics of each cluster are indicated by high or low \( z \)-scores. Final cluster centres, expressed in \( z \)-scores, are presented in Table II.

| Table I. Inter-correlations of indexes indicating health behaviours |
|----------------------|----------------|----------------|
|                      | Moderate physical activity | Excess index | Dietary diversity index |
| Vigorous physical activity | 0.18** | 0.03 | 0.07** |
| Moderately intensive physical activity | 0.03 | 0.07* |
| Excess index | 0.12** |

\*P < 0.005, **P < 0.001.
Cluster 1 (Sporty Healthy Eaters, \( n = 242 \)) comprises subjects who practice a lot of vigorous PA (an average of \(~8\) hours week\(^{-1}\)). The level of moderate PA is a bit lower than the total sample’s average. Scores on the excess index are low, while subjects in this cluster score high on the dietary diversity index. Subjects in Cluster 2 (Sporty Mixed Eaters, \( n = 288 \)) are also characterized by a high level of vigorous PA (mean time: 7 hours and 21 min), and an average amount of moderate PA (mean time: 1 hour and 17 min). Children within this cluster have the highest scores on the excess index but they also score relatively high on the dietary diversity index. Children in Cluster 3 (Moderate Active Healthy Eaters, \( n = 221 \)) show an average level of vigorous PA (almost 5 hours week\(^{-1}\)) but present the highest level of moderate PA (1 hour and 36 min week\(^{-1}\)). The score on the
excess index is rather low, while their dietary pattern is diversified. Cluster 4 children (Unsporting Unhealthy Eaters, \( n = 276 \)) score below the sample’s average on all PA and nutrition indexes. Their scores are not extremely low, except for dietary diversity. Members of Cluster 5 (Sedentary Healthy Eaters, \( n = 318 \)) are typified because of their low level of PA (vigorous: <3 hours, moderate: 53 min), an average score on the excess index and a higher score on the dietary diversity index.

ANOVA en post hoc Bonferroni results are marked in Table II.

**Cluster group characteristics**

Table III presents the relationship between the clusters and gender, SES and overweight. The significant chi-square for gender shows that the Sporty Healthy Eaters and Sporty Mixed Eaters comprise more males, while the Sedentary Healthy Eaters comprise more females. A relatively equal distribution of males and females was found in the Moderate Active Healthy Eaters and Unsporting Unhealthy Eaters. After controlling upon gender distribution, the associations between the clusters and SES and overweight were not significant. Thus, apart from gender, the clusters cannot be characterized by other factors included in this study, such as SES or overweight.

**Discussion**

This study aimed to investigate the cluster distribution based on four health indexes among 10-year olds.

Five reliable and meaningful clusters were identified. A drawback of cluster analysis is the possible dependency of the results on the selected sample [32]. The stability of our cluster solution was guaranteed by repeating the procedure in two randomly selected subsamples. Very similar results were found within each subgroup. The cluster names reflect the member’s health-related behavioural characteristics: Sporty Healthy Eaters, Sporty Mixed Eaters, Moderate Active Healthy Eaters, Unsporting Unhealthy Eaters and Sedentary Healthy Eaters. Cluster prevalence was more or less equal for each subgroup. The Sporty Healthy Eaters and Sporty Mixed Eaters include more boys. The Sedentary

---

**Table III. Percentages and chi-squares showing differences between clusters for related variables**

<table>
<thead>
<tr>
<th></th>
<th>Sporty healthy eaters (%)</th>
<th>Sporty mixed eaters (%)</th>
<th>Moderate active mixed eaters (%)</th>
<th>Unsporting unhealthy eaters (%)</th>
<th>Sedentary healthy eaters (%)</th>
<th>Chi-square</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>40</td>
<td>33</td>
<td>55</td>
<td>48</td>
<td>67</td>
<td>82.9***</td>
</tr>
<tr>
<td>Male</td>
<td>60</td>
<td>67</td>
<td>45</td>
<td>52</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td><strong>SES males</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>13</td>
<td>21</td>
<td>18</td>
<td>15</td>
<td>21</td>
<td>14.6</td>
</tr>
<tr>
<td>Medium</td>
<td>20</td>
<td>32</td>
<td>23</td>
<td>29</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>67</td>
<td>47</td>
<td>59</td>
<td>56</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td><strong>SES females</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>14</td>
<td>28</td>
<td>16</td>
<td>14</td>
<td>15</td>
<td>14.4</td>
</tr>
<tr>
<td>Medium</td>
<td>25</td>
<td>26</td>
<td>28</td>
<td>36</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>61</td>
<td>46</td>
<td>56</td>
<td>50</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td><strong>Overweight males</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>13</td>
<td>6</td>
<td>11</td>
<td>8</td>
<td>5</td>
<td>7.4</td>
</tr>
<tr>
<td>No</td>
<td>87</td>
<td>94</td>
<td>89</td>
<td>92</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td><strong>Overweight females</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>15</td>
<td>5</td>
<td>15</td>
<td>9</td>
<td>11</td>
<td>7.3</td>
</tr>
<tr>
<td>No</td>
<td>85</td>
<td>95</td>
<td>85</td>
<td>91</td>
<td>89</td>
<td></td>
</tr>
</tbody>
</table>

***P < 0.001.
Healthy Eaters comprise a significantly larger number of girls.

While previous literature used cluster analysis of multiple health-related behaviours and their determinants [26, 32–37], few studies have explored in particular the co-occurrence of these specific dietary and PA components. As far as we know, these are the first analyses using cluster analysis to examine these health-behavioural scores among children.

Meaningful clusters were found while correlations between the health indexes per se were low. This confirms the hypothesis that low correlations do not exclude co-occurrence of health-related behavioural indicators within certain groups.

Kremer et al. [26] draw attention to the contradictory findings concerning the associations between PA and eating habits. Some studies failed to show a relationship between diet and PA, others have shown significant correlations between PA and dietary fat, fibre and sucrose intake, with active individuals having healthier diets. The use of cluster analysis might be a good and meaningful alternative here.

In accordance with the current state of the literature, no SES differences were found in total PA among the children [38]. The observed gender differences in dietary habits and PA are compatible with previous studies. Vereecken et al. [39] hypothesized several explanations for gender differences in dietary habits: differences in the required energy, differences in health consciousness, differences regarding appearance and different socialization into adult-role patterns. Also unbalanced PA patterns between boys and girls have been generally observed before: boys and male adolescents are more physically active than girls [40, 41].

One of our most essential theoretical findings is that only one cluster represents children that have ‘healthy’ scores on each of the health-related behavioural indicators. Although they indicate lower levels of moderate PA, Sporty Healthy Eaters participate in sports activities a lot and show healthy dietary patterns. This finding implies that health-related behaviour indexes do not always discriminate in the same direction. Children practicing particular health-related habits are not predisposed to be involved into other specific healthy behaviour. De Bourdeaudhuij and van Oost [32] and Vingerhoets et al. [42] came to the same conclusion in their investigation of health-behaviour clusters among adults. De Bourdeaudhuij and Van Oost [32] assumed that people adhere to one health behaviour in order to consciously compensate for the lack of another. Based on the indexes’ character and the young age of the children, this hypothesis seems less plausible in our study. Nevertheless, we observe a high PA level in combination with high levels of high-energy food consumption (measured by the excess index) among the Sporty Mixed Eaters. This co-occurrence is related to Prentice’s and Jebb’s [43] description of inherent feedback systems capable of regulating energy homeostasis with great precision. Energy intake and energy expenditure are conceptualized as inter-related and complementary entities in order to keep the ‘energy balance’ stable and to approach the ‘set point’ of body weight. They state that this mechanism can occur without any apparent cognitive or physiologic recognition. Generally, Sedentary Healthy Eaters neither consume high amounts of energy-dense foods nor show high levels of PA. This behavioural pattern could also be interpreted in terms of this energy balance hypothesis.

Not much extra characterization of the clusters—based on gender, SES and overweight—was possible in the present study. Gender differences in cluster allocation mainly reflected the well-known higher sport participation in boys. The lack of significant differences in overweight among clusters can be expected, based on the young age of these children. However, it could be assumed that a continuous pattern of low PA and unhealthy eating might result in a higher overweight prevalence in the most ‘unhealthy’ cluster. Longitudinal follow-up will solve this question.

Unexpectedly, clusters could not be characterized by SES. Based on the relationship between low SES and unhealthy diet found in the literature in this age group, we had expected that the unhealthiest cluster (Unsporting Unhealthy Eaters) would include more children of lower SES. A possible explanation for this lack of difference could be that
the SES effect disappeared because of the combination with PA within the clusters, as a relationship between SES and PA was not found at this age in previous studies [32].

The rationale behind studying the clustering of health behaviours is double. While the cluster analysis gives insight on how health-related behaviour clusters in young people, the cluster analysis also identifies high-risk groups and thus provides a first step to primary prevention strategies [34]. The observation of children divided into distinct subgroups—characterized by health-related behavioural differences—could be a solid basis for targeted health promotion. Targeted communication is intended to reach a particular population subgroup based on characteristics presumed to be shared by the group’s members [44]. Messages are supposed to be specifically relevant for the intended audience and to contribute to an increased effectiveness. Our findings imply that children need selective advice on PA and eating habits in combination with clear reinforcements of healthy PA behaviours. The results make clear that we cannot assume that children who eat healthy are also sufficiently physically active and vice versa. There are specific subgroups with more or less healthy or unhealthy behavioural patterns and the latter ones form the group that is most at risk. As these groups at risk could not be identified by SES or BMI, screening of health behaviour is needed. If—by means of a simple screening instrument—children could be assessed in terms of the cluster they belong to, effective ways of directing them to relevant advice could be developed. The combination of nutritional behaviour and PA within this screening can save costs and money. Computer tailoring could appear to be a very useful technique concerning these topics. Future research also needs to address the specific determinants that distinguish between health-related behaviours within all clusters. Targeted health promotion messages and prevention strategies could focus on these determinants in order to optimize every cluster’s PA and/or eating habits.

Finally, it has to be kept in mind that this study is subject to some limitations. The data relied on children’s and parent’s self-reports of PA, eating behaviours and socio-demographic variables. But it is important to keep in mind that the strengths of the measurement instruments we used have been proven in previous and unpublished research [27, 29]. The cross-sectional design rather limits the study’s impact. The described sample selection method was not stratified and, as a consequence, representative results cannot be guaranteed. Nevertheless, the random sampling procedure and the extensive sample size justify generalization of the outcomes. Additionally, the study had to deal with a considerable amount of missing values on two variables. Results of independent samples t-test—comparing children with and without missing values on several variables—indicated no significant differences in terms of health index means. BMI calculations resulted in the largest number of missing data (14.8%). The amount of missing values is comparable with the quantity found in the study of Janssen et al. [45]. BMI in the latter study was calculated by means of children’s self-reported weight and height. The fact that we are dealing with a composite measure and parents’ lack of knowledge of children’s exact height and weight are two plausible explanations for the non-response. Due to the extensive sample size, it was not feasible to measure all children’s body dimensions. A final limitation of the present study was that only few factors were available for characterizing the clusters. Future studies should investigate other factors in order to better characterize these subgroups.

The strength of the present study was the investigation of PA, eating behaviour and socio-demographic correlates in a large school sample and the evaluation of clustering of the health behaviours. Even though the practical implications are rather preliminary, the conceptual and theoretical knowledge of the clusters’ existence and distribution is scientifically relevant. As mentioned earlier, the cluster-analytical distribution of people’s health behaviour (concerning smoking behaviour, sleeping patterns, PA, etc.) has been investigated earlier, but the relationship between each of the specific composite health behaviour scores within clusters has not been shown.
A prospective cohort study could provide additional information concerning the stability and/or evolution of the identified cluster solution and their socio-demographic correlates. Results of these analyses and specific determinant research could form a steady basis for targeted health promotion in order to prevent the decrease of specific health behaviours and reinforce the maintenance of other healthy behaviours during early adolescence.

Acknowledgements

Special thanks to the participants, their parents, teachers and principals for their voluntary contribution to this project. This project was made possible thanks to the Ghent University.

Conflicts of interest statement

None declared.

References


Received on September 9, 2005; accepted on September 14, 2006