An epidemiological study of child health and nutrition in a northern Swedish county. VI. Relationship between general and oral health, food habits, and socioeconomic conditions

Gösta Samuelson, M.D., Hans Grahnén, Dr.Odont., and Erik Arvidsson, Fil. kand.

During the fall of 1967, a general study of the health and nutritional status of 1,401 children, aged 4, 8, and 13 years, was carried out in three areas of a northern Swedish county that differ from one another both geographically and, in part, socioeconomically. The study was conducted in an urban area, the city of Umeå, and in two rural areas, the inland area and the mountain foreland. In all three areas, 8- and 13-year-olds were examined, and in the city of Umeå, 4-year-olds were also included. The epidemiological study consisted of a food consumption survey (1), a methodological study of the recall technique (2), medical and anthropometric examinations (3), hematological examinations (4), and oral health studies (5).

The food consumption survey (1) showed that according to 24-hr recall, the children in all age groups consumed a diet that complied with or exceeded the 1968 recommendations of the U. S. Food and Nutrition Board, except those for iron and vitamin D. For 13-year-olds, iron intake was below the recommended level, and for all age groups, the intake of vitamin D from the diet was lower than that recommended. The fat content was generally high, 40 to 42% of the energy intake. The boys consumed more food than the girls and thus had a higher intake of several nutrients. Similarly, children in the rural areas consumed more than children in the city of Umeå.

The children’s consumption patterns were investigated by means of a food habit history. In personal interviews it was ascertained how often the children consumed different types of food and dishes. The consumption frequencies were graded from “never,” “every other month,” et cetera, up to “more than four times per day.” The food habit history revealed significant differences between the eating patterns in the three areas. In general, vegetables and fruit were consumed more frequently in the city of Umeå than in the two rural areas, where berries and, to a large extent, potatoes were eaten more often. Products such as blood pudding and liver were consumed more frequently in the city of Umeå than in the mountain foreland, whereas cereal products such as porridge, gruel, pancakes, et cetera, were eaten more often in the rural areas than in the city of Umeå.

With respect to between-meal eating, it was found that children in the city of Umeå frequently ate fruit between meals, whereas buns and cakes and/or soft drinks were consumed more often in the two rural areas. The total frequency of consumption of sweets, on the other hand, was lowest in the mountain foreland.

The medical and anthropometrical investigation (3) revealed that the children’s general health and nutritional status was good. The girls in all age groups had greater skin-fold thickness than the boys. The skinfold measurements of urban children were greater than those of children in the mountain foreland. Eight- and thirteen-year-olds in the city of Umeå were also somewhat taller, on the average, than children in the

1 Supported by the Bank of Sweden Tercentenary Fund and the Swedish Patent Revenue Research Fund.
2 Assistant Professor, Department of Paediatrics, University Hospital, Umeå, Sweden. 3 Professor and Chairman, Department of Pedodontics, University of Umeå. 4 Institute of Mathematics and Statistics, University of Umeå.
mountain foreland. In all age groups, there was a rather high frequency of subclinical upper respiratory infection.

The hematological investigation (4) revealed iron-deficiency anemia in only two 13-year-old girls. Children in the city of Umeå had lower mean hemoglobin and packed red cell volume values than children in the mountain foreland.

The oral health studies (5) included determinations of caries indices and indices of plaque and gingivitis. In addition, certain anamnestic data were registered, including the frequency of tooth brushing and of pre- and postreuptive exposure to fluoride. In the 13-year-old group, the girls had higher caries indices than the boys, as well as a larger number of erupted permanent teeth; the plaque and gingival indices, on the other hand, were higher in boys. There were differences between the three areas with respect to caries indices. Regional differences in plaque and gingival indices were found only for 13-year-olds, with the lowest values in the city of Umeå, where the frequency of tooth brushing was highest.

In connection with the food consumption survey (1), information about the families' socioeconomic situation was obtained in personal interviews with the mothers. The following anamnestic data were registered: number of children, mother's age, mother's formal education, father's formal education, dwelling space expressed as the number of persons per room, and the family's total yearly income, including all types of supplementary economic aid received. There proved to be more children per family in the rural areas than in the city of Umeå. The educational level of the parents was higher in the city of Umeå than in the two rural areas, between which there were no significant differences in parent's education. The parents' median combined total income was higher in the city of Umeå for all age groups investigated. The two rural areas did not differ from one another in this respect. The dwelling space per person was larger in the city of Umeå than in either of the two rural areas, between which there was no significant difference.

It was found that there were sex and regional differences with respect to findings in the medical, odontological, and anthropometrical examinations. Regional differences might be due to different food habits, oral hygiene and/or socioeconomic conditions. The observed regional differences in the frequency of caries do not seem to be explainable by differences in exposure to fluoride (5).

The aim of the present study was to investigate whether there was a connection between the children's medical and oral health status, their frequency of consumption of different foods, and their socioeconomic situation.

Material and methods

The original material is shown in Table 1. All the children were selected from the population register for Västerbotten county. In the city of Umeå, every second child was selected; in the two rural areas, all the children in the two age groups studied were included. Those with chronic diseases were excluded (1).

For analysis of the relationship between general and oral health, food habits, and socioeconomic conditions, stepwise multiple regression analyses were performed. The following medical variables were used: 1) hemoglobin values (Hb), 2) packed red cell volume (PCV), 3) erythrocyte microsedimentation rate (ESR), 4) skin-fold measurements (triceps and subscapular skin folds), 5) body weight, 6) body height, and 7) the number of colds per year.

The odontological variables used were: 1) caries index in the primary dentition — decayed, extracted, filled teeth (def) for 4-year-olds; 2) caries index in the primary dentition for 8-year-olds according to Gerdin (6); 3) caries index in the permanent dentition — decayed, missing, filled teeth (DMFT) for 8- and 13-year-olds; 4) gingival index according to Löe and Silness (7); 5) plaque index according to Silness and Löe (8); 6) number of primary teeth; 7) number of permanent teeth; 8) frequency of tooth brushing (times per day); and 9) frequency of mouth rinsing with water after meals (times per day).

Socioeconomic and demographic variables that were included were: 1) mother's education, 2) father's education, 3) family's total per capita income, 4) number of persons per room, 5) number of children, 6) mother's age, 7) sex, and 8) area (city of Umeå, inland area, mountain foreland).

Food frequency consumption variables were: vegetables, fruit (total), berries, meat, fish, eggs, sausage, porridge, sweets (total), coffee (except 4-year-olds), buns and cakes (total); and for between-meal eating, the variables were: fruit, sweets, buns and cakes, and soft drinks.

The consumption frequency variables are derived from the food habit history. Variables for which the frequency of consumption were approximately the same for most of the children; for example, potatoes,
milk, milk products, and blood pudding, have not
been included.
For the age variable, the child's age is given in
months.

Comments on the definition of variables

The definitions of most variables are given in
previous papers (1, 3–5). However, the definitions of edu-
cation and food frequency consumption need further
elucidation.

Education. The parents' educational levels were
graded according to the following:
1 = elementary school only (≤ 6 years),
2 = education beyond elementary school up to a
total of 9 years formal education,
a) more than 6 but less than 9 years formal edu-
cation,
b) complementary schooling after some working
experience,
c) vocational school,
3 = 9 to 11 years of formal education,
a) began, but did not complete gymnasium,
b) degree from technical or business school,
c) other education, at least 2 years.
4 = 12 or more years of formal education. Degree
from gymnasium, academic studies, etcetera.

Because the educational level depends not only
on the number of years of study, but also on the type and
scope of education received, the above classification
was used. That is, the educational level was coded on
an interval scale instead of on an ordinal scale for the
purpose of analyzing the connection between the edu-
cational level of the parents and different parameters
applying to the children.

Food frequency consumption. The frequency of con-
m summption of foods was originally expressed as the num-
ber of occasions of consumption per unit of time. An
index was formed by taking the square root of the
consumption frequency. This was done to make the
differences in effect between different levels of con-
m summption comparable and to reduce the effect of
wrong estimates, that is, to reduce the skew-
ness of the variables. If the square root had not been
used, an increase in frequency from once to twice a
day would have been regarded as half as great as an
increase in frequency from twice to four times per
day. Using the square root index, the relation between

TABLE 1
Age, sex, and geographical distribution of the 1,401 children studied

<table>
<thead>
<tr>
<th>Area</th>
<th>Age (years) and sex of children</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Boys</td>
<td>Girls</td>
</tr>
<tr>
<td>City of Umeå</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>Inland area</td>
<td>98</td>
<td>90</td>
</tr>
<tr>
<td>Mountain foreland</td>
<td>99</td>
<td>96</td>
</tr>
<tr>
<td>Total</td>
<td>99</td>
<td>99</td>
</tr>
</tbody>
</table>

TABLE 2
Number of omissions for different variables; age distribution

<table>
<thead>
<tr>
<th>Variables</th>
<th>Age groups, years</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Microsedimentation rate (ESR)</td>
<td>6</td>
<td>19</td>
</tr>
<tr>
<td>Triceps and subscapular skin fold</td>
<td>47</td>
<td>0</td>
</tr>
<tr>
<td>Deft (4-year-olds only)</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>DMFT</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Gerdin's index (8-year-olds only)</td>
<td>15</td>
<td>181</td>
</tr>
<tr>
<td>Gingival index</td>
<td>15</td>
<td>181</td>
</tr>
</tbody>
</table>

the two increases becomes 1.1.5. In an interview study
like the present one, the degree of precision is such that
in some cases a response of twice a day actually
could have been once or three times per day. To re-
duce the influence of such errors, the index above was
formed, and in the following text is denoted as "fre-
cquency of consumption."

Omissions of variables

The number of omitted variables are shown in
Table 2. They were fairly evenly distributed between
the sexes and between the different areas. The excep-
tion was skin-fold measurements, for which all the
omissions for methodological reasons were in 4-year-
olds in the city of Umeå (3). The many omissions for
plaque and gingival indexs were due to the technical
and methodological factors and were unsystematic
in nature. As the omissions were fairly evenly distributed
among the areas investigated and the remaining ma-
terial was sufficiently large, these omissions should not
have any decisive effect on the results (5).

Regression analyses

Multiple regression analyses were used to analyze the
correlations between different groups of variables.
Eight analyses were performed, with each age group treated separately, as shown in Table 3.

Multiple regression analysis is a method of studying the relationship between the regressand (dependent variable) $y$ and a set of regressors (independent variables) $x_1, x_2, \ldots, x_k$ assuming a linear model

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_k x_{ik} \quad (i = 1, 2, \ldots, n),$$

where $n$ is the number of individuals in the sample. The unknown parameters $b_0, b_1, \ldots, b_k$ are estimated in a way that minimizes the sum of squares

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \left(y_i - \hat{b}_0 - \hat{b}_1 x_{i1} - \cdots - \hat{b}_k x_{ik}\right)^2$$

This minimizing problem is solved by the method of least squares, giving unique estimates $\hat{b}_0, \hat{b}_1, \ldots, \hat{b}_k$ of the parameters $b_0, b_1, \ldots, b_k$. The strength of the dependence between the regressand and the set of regressors is usually measured in terms of $R^2$, the square of the multiple correlation coefficient (also called the coefficient of determination) defined as

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

$R^2$ shows the amount of the variation of the regressand explained by the regressors. If $R^2 = 0$, no variation is explained, and if $R^2 = 1$, no further variation is left to explain. The problem of deciding whether every single regressor in the model is explanatory enough to be accepted as an important contribution to the regression equation has been handled as follows. It is obvious that if $b_i = 0$, the variable $x_i$ is superfluous. The decision problem is therefore solved by testing the null hypothesis $H_0: b_i = 0$, against the alternative hypothesis $H_1: b_i \neq 0$. Rejecting $H_0$ at some significant level specified in advance implies acceptance of the hypothesis that the regressor $x_i$ makes a significant

<table>
<thead>
<tr>
<th>Analyses</th>
<th>Regressands</th>
<th>Regressors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis 1</td>
<td>Medical variables except number of colds</td>
<td>Food variables</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Age (Sex; area)</td>
</tr>
<tr>
<td>Analysis 2</td>
<td>Medical variables except number of colds</td>
<td>Socioeconomic variables</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Sex)</td>
</tr>
<tr>
<td>Analysis 3</td>
<td>Caries indices</td>
<td>Medical variables</td>
</tr>
<tr>
<td></td>
<td>Plaque index</td>
<td>Age (Sex; area)</td>
</tr>
<tr>
<td></td>
<td>Gingival index</td>
<td></td>
</tr>
<tr>
<td>Analysis 4</td>
<td>Caries indices</td>
<td>Food variables</td>
</tr>
<tr>
<td></td>
<td>Plaque index</td>
<td>Age</td>
</tr>
<tr>
<td></td>
<td>Gingival index</td>
<td>Number of teeth</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Frequency of tooth brushing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Frequency of rinsing the mouth</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Sex; area)</td>
</tr>
<tr>
<td>Analysis 5</td>
<td>Caries indices</td>
<td>Socioeconomic variables</td>
</tr>
<tr>
<td></td>
<td>Plaque index</td>
<td>(Sex)</td>
</tr>
<tr>
<td></td>
<td>Gingival index</td>
<td></td>
</tr>
<tr>
<td>Analysis 6</td>
<td>Food variables</td>
<td>Socioeconomic variables</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Sex)</td>
</tr>
<tr>
<td>Analysis 7</td>
<td>Medical variables except number of colds</td>
<td>Food variables</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Socioeconomic variables</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Age (Sex)</td>
</tr>
<tr>
<td>Analysis 8</td>
<td>Caries indices</td>
<td>Medical variables</td>
</tr>
<tr>
<td></td>
<td>Plaque index</td>
<td>Food variables</td>
</tr>
<tr>
<td></td>
<td>Gingival index</td>
<td>Socioeconomic variables</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Age</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Number of teeth</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Frequency of tooth brushing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Frequency of rinsing the mouth</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Sex)</td>
</tr>
</tbody>
</table>
contribution to the explanation of the regressand \( y \). In testing \( H_0: \beta_j = 0 \), the method described by Draper and Smith (9) was used.

In the present work we selected the stepwise multiple regression technique for investigating the relations between regressand and regressors. A standard program, Dixon Edition (BMD02R, 1968, Health Sciences Computing Facility, UCLA) was used. The calculations were performed on a computer CD 3200. The regressors and regressands are listed in Table 3. The linear relationships between the regressands and regressors were calculated. In analyses 1, 3, and 4, in which no socioeconomic variables were used, both sex and area were forced. The three areas were included in the same regression analysis, because we assume the correlations between the variables to be equal in the three areas. An example from analysis 4 is given in the appendix.

In analyses 2, 5, 6, 7, and 8 the variable sex was forced and separate analyses were performed for the city of Umeå and the two rural areas combined. We believe the combining of the rural areas to be justified because we found no significant differences between these two areas with respect to the socioeconomic variables used. Only regressors entering at a significance level of \( \leq 1\% \) have been considered. Depending on the great number tested hypothesis (about 1,400), approximately 14 of the significant correlations at the 1\% level are produced at random. On the 0.1\% level, the corresponding figure is 1 to 2.

Results and discussion

The results of the stepwise regression analyses are given in Figs. 1 to 5. It is important to point out that the food frequency consumption variables contribute little to the explanation of general and oral health as reflected by the variables used. In almost all analyses we obtained small \( R^2 \) values (0 to 10\%). The reasons may be: a) we had not been able to collect some important variables that would have explained much of the variation in the different regressands, e.g., for odontological variables trace elements such as fluorides, saliva factors, dental care, tooth morphology, and malocclusions; b) the food frequency consumption variables, both as regressors and regressands, probably apply to only a short period, whereas most of the medical variables and caries indices reflect a relatively long period; c) some variables are derived from interview data, which is a factor of uncertainty; d) values of some variables are based on subjective judgments. It is to be observed that as the material is relatively large, only a small influence of one variable on another may cause statistical significance.

It is important to realize that several of the regressors are related. In interpreting the results as shown in the figures, attention should be directed not only to the individual variables but also to groups of variables. The variables used in the analyses represent only a few of the numerous elements that are included in the concepts of diet and socioeconomic conditions.

Figure 1 shows that the medical and odontological regressands of 4-year-olds were affected by socioeconomic conditions and diet. For example, the total frequency of consumption of sweets was positively correlated to deft values and plaque indices. The medical variables did not affect the odontological variables. Socioeconomic conditions, in turn, influence the frequency of consumption of different foods; for example, a higher educational level of the mother or the father, or both, decreased the consumption of sweets by the children and a higher income per capita decreased the between-meal eating of buns and cakes. It also appears from the figure that children of parents with a higher level of education had, on average, lower deft values and lower gingival indices.

For 8- and 13-year-olds in the city of Umeå (Figs. 2 and 4), as well as in the two rural areas taken together (Figs. 3 and 5), both medical and odontological regressands were affected by socioeconomic conditions and diet. For example, for the 13-year-olds, frequent consumption of porridge and buns and cakes in between-meal eating negatively influenced the Hb and PCV values and the body weight, respectively. A high frequency of consumption of sweets and buns and cakes increased Gerdin's index, DMFT, and body weight.

For both 8- and 13-year-olds, a greater number of erupted permanent teeth and higher body weight or height, or both, increased the DMFT values. Our results are consistent with Filipsson's finding (R. Filipsson, personal communication) that there is a positive correlation between body height and the number of erupted permanent teeth in children. For the 8-year-olds, frequent tooth brushing decreased the plaque index, and
for the 13-year-olds, frequent tooth brushing gave both a lower plaque index and a lower gingival index. The caries indices were not correlated to the frequency of tooth brushing.

Concerning socioeconomic conditions, higher educational levels of the parents decreased the caries indices, except for the 8-year-olds in the rural areas. In all age groups, one can also see the connection between socioeconomic conditions, food habits, and general and oral health status; for example, a higher educational level of the parents decreased the children's frequency of eating sweets and/or buns and cakes. Socioeconomic conditions also affected the frequency of fruit and porridge consumption. It is striking that the frequency of coffee drinking was influenced by these conditions; higher education of the parents decreased coffee drinking, except for 13-year-olds in the city of Umeå. In childhood, the habit of drinking coffee is probably usually connected with consumption of buns and cakes. Improved living standards, especially higher educational levels of the parents, affected the urban children's frequency of consumption of fish and berries in a positive direction. This effect could not be demonstrated in the rural areas, where the mean consumption of fish and berries, available di-
rectly from nature, was higher than in the city of Umeå.

The age variable has not been given in Figs. 1 to 5 because the relationship found in the analyses between age and body height and weight ought to be self-evident. The differences found between the sexes and between geographic areas have been reported earlier (1, 3–5).

As has been pointed out, the parents’ education was coded on an interval scale. To check the results given by regression analysis with parents’ education as regressor, the material was divided into four groups according to the parents’ educational level, graded in the manner described. For each of these groups, the mean values of the variables entering the regression analysis were calculated for children in the city of Umeå and in the rural areas, respectively, each age group being treated separately. The results showed throughout a rise or fall of the mean values for the variables that, according to regression analysis, were affected by the parents’
educational level. This supports the validity of the educational scale used in the regression analyses.

When all the variables previously used were used as regressors in analyses of the medical variables, caries, plaque, and gingival indices (Analyses 7 and 8, Table 3), it was found that both food and socioeconomic variables exerted an influence. Neither type of variable had a predominant effect, as far as could be discerned.

The effect of socioeconomic conditions on general health status could be discerned in the 8- and 13-year-olds. In the city of Umeå, social conditions (fewer persons per room and higher educational level of the parents) were reflected in a higher hemoglobin level in both age groups. In Odin's study of 1931 (10), which was carried out in approximately the same parts of Västerbotten county as the present investigation, a connection was also found between crowded living conditions and lower hemoglobin values. This is probably due to underlying dietary factors. Crowded living conditions and the presence of many children in the family, along with other factors, often reflect a poor socioeconomic situation (11, 12). That nutritional status and food intake are influenced by socioeconomic factors has
been shown, for instance, by Devadas and Easwaran in investigations from India (13). They found in a study of preschool children that the family income appeared to have a decisive influence; i.e., low income groups had a predominantly cereal diet, the upper income families had a more varied diet. It is to be observed that the educational level of the parents in their study did not have much influence on dietary practices. This is probably due to the generally low educational level in the population studied. Also in studies in Sweden during the last decade (14, 15), the importance of socioeconomic factors in the development of children has been emphasized.

The effect of socioeconomic factors on caries status (16–18), standards of oral hygiene (19), and periodontal health (20, 21) in children has previously been noted. Koch and Martinsson (22) studied 14-year-old schoolchildren in Malmö, Sweden, with high and low caries frequencies. They found that the caries situation of the children was correlated to several social factors, especially to the parents' educational level and the family's social class.

The evidence for the value of oral hygiene practices in preventing caries is ambiguous. However, a few investigations (23, 24) suggest that good oral hygiene reduces caries activity. The Vipeholm study (25)
Fig. 5. The 13-year-old group. The relationship between medical, odontological, and food frequency consumption variables in the city of Umeå and the rural areas taken together. Influence of socioeconomic variables in the rural areas.

demonstrated that between-meal consumption of sweets is a strong caries-promoting factor. The importance of eating between meals, particularly of sugar and sweets, is emphasized in several studies (23, 26–30). It has also been found that standards of oral hygiene and the frequency of tooth brushing in children are related (19, 23).

Although the materials and methods in the literature cited differed considerably from those in our study, these and other previous studies of children and adults nonetheless point to a similar connection between socioeconomic conditions, food habits, and the state of general and oral health.

Summary

An epidemiological survey of the general and oral health and food habits of 1,401 children aged 4, 8, and 13 years has been carried out in one urban area and two rural areas in a northern Swedish county.

The purpose of this study was to investigate, by means of multiple regression analysis, whether there was a connection between the children's general and oral health, their frequency of consumption of various foods, and the socioeconomic conditions in the family.

The investigation showed that socioeco-
nomic conditions, especially the educational level of the parents, were correlated to the frequency of the children's consumption of various foods. Among the 4-year-old urban children, better socioeconomic conditions were positively correlated to the frequency of fruit and vegetable consumption, and negatively correlated to the consumption of sweets, buns and cakes between meals. For the 8-year-olds, better socioeconomic conditions were negatively correlated to, among other things, the consumption of porridge, sweets, and buns and cakes.

For the 13-year-olds, better socioeconomic conditions were positively correlated to the consumption of vegetables, berries, and fish and negatively correlated to the consumption of sweets, buns and cakes.

A high frequency of consumption of certain foods, such as the latter, negatively affected the caries indices and the state of oral hygiene (plaque indices). The frequent eating of some foods was correlated to certain medical and anthropometrical data. Among the food variables, frequent consumption of porridge was associated with decrease in the Hb and PCV values for the 13-year-olds. Body height or weight, as well as number of permanent teeth, affected the DMFT value. Frequent tooth brushing decreased the plaque index, except for the 4-year-olds.

Our thanks are due to Associate Professor Gunnar Eklund, Uppsala, for advice and critical discussions, and to Mr. Hans Wallberg, Fil. kand., for help with the computer analysis at the computer center, University of Umeå.

Appendix

An example from the present work

In order to show, in detail, the behavior and the advantages of the stepwise multiple regression procedure, we have chosen analysis 4, which concerns 13-year-old children (Fig. 4 or 5). As regressand, we chose the DMFT index. The regressors were the food frequency consumption variables, specified age in months, number of teeth (permanent), frequency of tooth brushing, and frequency of rinsing the mouth. Values of these variables were recorded for 614 children.

The first steps of the analysis divided the children into six cells, representing every possible combination of sex and geographical area. This was done by estimating the parameters $b_1, j = 0, 1, 2, 3$ in the model

$$Y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{3i} (i = 1, 2, \cdots, 614)$$

where

- $x_{1i} = 1$ and $x_{4i} = 0$ if the $i$:th individual lived in the city of Umeå
- $x_{1i} = 0$ and $x_{4i} = 1$ if the $i$:th individual lived in the inland area
- $x_{1i} = x_{4i} = -1$ if the $i$:th individual lived in the mountain foreland
- $x_{1i} = -1$ if the $i$:th individual was a girl, otherwise $x_{1i} = 1$

$Y_i$ is the DMFT value recorded for the $i$:th individual.

By the method of least squares, the fitted equation was:

$$\hat{y}_i = 6.18 - 0.51 x_{1i} - 0.47 x_{3i} - 0.41 x_{4i}$$

which gave $R^2 = 0.087$, i.e., 8.7% of the variation of the regressand was explained by differences in DMFT values between the three geographical areas and between the sexes. The negative value of $b_2$ and the way variable $x_3$ was coded told us that girls, on an average, had higher DMFT index than boys and that the difference between the means was $2 \times 0.41 = 0.82$. In the same way, we were able to locate differences between the geographical areas.

By using the model above, we assume that the effect of the regressors on the regressand is equal in all cells. In those analyses in which we had the socioeconomic variables as regressors, we could not assume this to be true, which is why we performed separate analyses for the city of Umeå and for the two rural areas combined.

The first regressor to enter the equation was the one for which the partial correlation with the regressand was highest when variables $x_1$, $x_2$, and $x_3$ were held constant. In other words, the regressor should have high correlation to the DMFT index and relatively low correlation to the variables already in the equation (in this step, variables $x_1$, $x_2$, and $x_3$). The variable with these qualifications turned out to be number of permanent teeth (variable $x_{11}$). The fitted equation was:

$$\hat{y}_i = -3.13 - 0.44 x_{11} - 0.52 x_{21} - 0.20 x_{31} + 2.8 x_{111}$$

Downloaded from https://academic.oup.com/ajcn/article-abstract/24/11/1361/4733155
and $H_0: b_{11} = 0$ was rejected at the 0.1% significance level. A positive value of $b_{11}$ implies that, on an average, children with a large number of permanent teeth have a higher DMFT index than those with fewer permanent teeth. We also found that $b_9$ had become smaller in absolute value, indicating that a great deal of the difference between the sexes was due to the fact that, on the average, girls in this age group had a larger number of permanent teeth than boys. This is in accordance with Samuelson et al. (5).

The next variable to enter was the one in which partial correlation to DMFT was highest when variables $x_{11}, x_2, x_3, x_{111}$, and $x_{111}$ were kept constant. This variable turned out to be the total frequency consumption of buns and cakes (variable $x_{32}$).

$H_0: b_{32} = 0$ was rejected at the 1% significance level and the fitted equation was

$$y_i = -3.76 - 0.39 x_{11} - 0.57 x_{11} - 0.23 x_3 + 0.71 x_{111} + 0.13 x_{111}$$

which tells us that children with a frequent (total) consumption of buns and cakes and with a large number of permanent teeth had, on an average, a higher DMFT index than those who did not belong to this category.

The next variable to enter was the one most correlated to the regressand when variables $x_{11}, x_2, x_3, x_{111}$, and $x_{32}$ were held constant. This variable was frequency of consumption of sweets between meals (variable $x_{34}$). $H_0: b_{34} = 0$ was rejected at the 1% level and the fitted equation was:

$$y_i = -4.41 - 0.44 x_{11} - 0.55 x_1 - 0.21 x_{11} + 0.70 x_{111} + 0.13 x_{111} + 0.13 x_{111}$$

which tells us that frequent consumption of sweets between meals is also a factor that increases the DMFT index.

Of the remaining regressors, age in months (variable 6) proved to have the highest partial correlation to DMFT when the regressors already in the equation were held constant. $H_0: b_6 = 0$ was not rejected at the 1% significance level, so the stepwise procedure stopped.

The last fitted equation gave $R^2 = 0.26$. However, about 9% of the variation of the regressand was explained by differences between the sexes and between areas, and about 15% by number of permanent teeth after sex and area had been included, leaving only about 2% for the two food frequency consumption variables to explain.

REFERENCES


